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, 21 Dezember 2021

Abstract

In dieser Arbeit werden Werkzeuge und Maßnahmen untersucht, die zur Kostenkontrolle von *AWS-Diensten(Amazon Web Services)* beitragen. Darüber hinaus werden allgemeine Optimierungsmaßnahmen aufgezeigt, die bereits über die Jahre hinweg von anderen *Cloud-Nutzern* getestet wurden. Die Optimierungsmaßnahmen werden von AWS als *Best Practices* empfohlen. Die Grundlage dieser Arbeit sind Empfehlungen von Cloud-Anbietern bezüglich Kostenüberwachung und -optimierung, Erfahrungen von Experten dieses Fachgebiets und Anregungen aktueller Fachliteratur.

Diese Arbeit ist besonders bedeutsam für Teams, die AWS-Cloud-Dienste in aktuellen Projekten anwenden und die Kosten in der Cloud besser verstehen und optimieren möchten. Wenn die Kosten für Cloud-Dienste wie alle anderen Kosten betrachtet werden, ist es konsequent, über ihre Überwachung, Kontrolle und Optimierung nachzudenken. Ein häufiges Problem in Unternehmen ist das fehlende Verständnis der in der Cloud anfallenden Kosten.[1] Aus diesem Grund stehen Unternehmen, die noch eine *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 für Unternehmen relevant, die von einem Wechsel von klassischen Modellen (bekannt als On-Premise) zu cloudbasierten Modellen profitieren möchten.

[1] Vgl. Stern, Adam, 2018, The Truth About Cloud Pricing, o.S. [66]

Inhaltsverzeichnis

Abbildungsverzeichnis

Glossar

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.[2]

Cloud-Computing

Das NIST definiert Cloud Computing als das Modell zur Ermöglichung eines allgegenwärtigen, bequemen und bedarfsgerechten Netzzugangs zu einem gemeinsamen Pool konfigurierbarer Rechenressourcen (z. B. Netze, Server, Speicher, Anwendungen und Dienste), die mit minimalem Verwaltungsaufwand oder minimaler Interaktion mit dem Dienstanbieter schnell bereitgestellt und freigegeben werden können.[3]

Cloud-Dienst

Bei Cloud-Dienste geht es um sämtliche Infrastruktur-Komponenten wie die Server, Rechenleistung, Netzkapazitäten, Kommunikationsgeräte, Speicher, Archivierungs- und Backup-Systeme und andere Komponenten der Rechenzentrum- und Netzinfrastruktur, die von dem Cloud-Service-Provider zur Verfügung gestellt werden. Der Anwender kann über das Netzwerk (i. d. R. Internet) auf die virtuellen Services zugreifen. Beispiele für Cloud-Dienste stellen die Elastic Compute Cloud (EC2) von Amazon, die MicrosoftWindows Azure virtuelle Maschinen und die Google Compute Engine.[4]

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.[5]

Instagram-Story

Bei Instagram Stories handelt es sich um kurzen visuellen Content in der Regel Bilder

[2]Vgl. Mark Wilkins, 2021, AWS Certified Solutions Architect - Associate (SAA-C02), S.42.[1]

[3]Vgl. The NIST Definition of Cloud Computing. S.6 [47]

[4]Vgl. Helmut Krcmar, 2015, Einsatzfelder und Herausforderungen des Informationsmanagements. Informationsmanagement. 6. Auflage S.724[5]

[5]Vgl. Mark Wilkins, 2021, AWS Certified Solutions Architect - Associate (SAA-C02). S.95[1]

oder kurze Videos, die nach 24 Stunden automatisch aus der Applikation Instagram verschwinden(Stand November 2021).[6]

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.[7]

Tag

Ein *Tag* (Markierung) ist eine Markierung, die einer AWS-Ressource zuordnet. Jeder Tag (Markierung) besteht aus einem Schlüssel und einem optionalen Wert.[8]

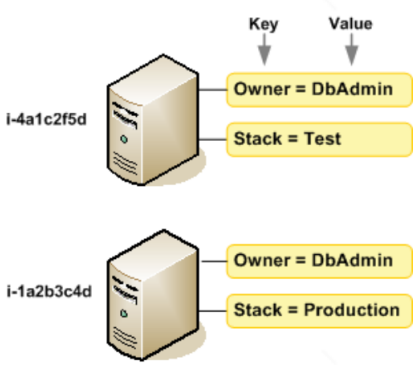


Abbildung 1
Beispiel für ein Tag[28], S.1570.

Metadaten

Metadaten liefern Informationen über den Inhalt eines bestimmten Objekts. Ein Bild kann beispielsweise Metadaten enthalten, die beschreiben, wie groß das Bild ist, die Farbtiefe, die Bildauflösung, wann das Bild erstellt wurde und andere Daten. Die Metadaten eines Textdokuments können Informationen darüber enthalten, wie lang das Dokument ist, wer der Autor ist, wann das Dokument geschrieben wurde und eine kurze Zusammenfassung des Dokuments.[9]

[6]Vgl. Online Marketing: Definition von Instagram Story? o.S.[50]
[7]Vgl. Mark Wilkins, 2021, AWS Certified Solutions Architect - Associate (SAA-C02)[1], S.42
[8]Vgl. AWS, 2019, Amazon Elastic Compute Cloud - Benutzerhandbuch für Linux-Instances, S.1570[28]
[9]Vgl. Techterms Definition Metadata[60]

Startkonfiguration
Eine Startkonfiguration ist eine Instance-Konfigurationsvorlage, die eine Auto-Scaling-Gruppe zum Starten von EC2-Instances verwendet.[10]

Maschinelles Lernen
Maschinelles Lernen ist ein Teilbereich der künstlichen Intelligenz (KI). Beim maschinellen Lernen werden Algorithmen darauf trainiert, Muster und Korrelationen in großen Datensätzen zu finden und auf Basis dieser Analyse die besten Entscheidungen und Vorhersagen zu treffen. Auf diese Weise wird die Rechenkapazität von EC2-Instanzen auf der Grundlage früherer Muster vorhergesagt.[11]

YAML
YAML ist eine benutzerfreundliche Daten-Serialisierungs Sprache für alle Programmiersprachen.[12]

[10] Vgl. Amazon EC2 Auto Scaling - Benutzerhandbuch. S.54 [33]
[11] Vgl. SAP, o.J., Definition von machinellen Lernen, o.S.[58]
[12] Vgl. YAML Org, , o.J., Definition von YAML, o.S.[76]

Abkürzungsverzeichnis

AWS Amazon Web Services

API Application Programming Interface

ASG Auto Scaling Group

CI/CD Continuous Integration / Continuous Deployment

EC Elastic Compute

GK Anzahl der gewonnenen Kunden

GB Gigabyte

JSON JavaScript Object Notation

TCO Total Cost of Ownership

MK Anfallende Marketingkosten

PAYG Pay-as-you-go

SSO Single Sign-On

TB Terabyte

VK Vertriebskosten

KPI Key Performance Indicators

1 Einleitung

1.1 Motivation

Die zunehmende Digitalisierung von Geschäftsmodellen, die auch durch die Corona-Pandemie vorangetrieben wird, lässt Cloud-basierte Applikationen an Bedeutung gewinnen.[13] Als direkte Folge davon ist die Nachfrage nach Server- und Speicherkapazität konsequent gestiegen. Die Relevanz von *Amazon Web Services*, kurz AWS, im Bereich der *Cloud-Computing*, ergibt sich aus einer vor kurzem veröffentlichte Studie von Raj Bala et al.. Diese wies eindrücklich daraufhin, dass AWS der aktuell weltweit führende Cloud-Anbieter anhand ihrer Klassifikation (*Magic Quadrant*[14]) für Cloud-Infrastruktur und Plattform-Services sei. [15] So erscheint AWS nicht nur aus diesem Grund als Fallbeispiel für diese Arbeit passend, weitere bedeutsame Faktoren sind seine frühe Präsenz (2006) als Cloud-Anbieter und seines großen Angebotes an Cloud-Diensten, welche für zahlreiche Anwendungsfälle geeignet sind.[16].

1.2 Problemstellung

Adam Stern wies in dem *Forbes*-Magazin daraufhin, dass ungefähr die Hälfte der US-amerikanischen Unternehmen Schwierigkeiten hätten ihre Kosten zu begründen.[17]
So erklärt er:

> 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.[18]

Darüber hinaus weist Tobias Regenfuß und Jochen Malinowski in einer Untersuchung daraufhin, dass es den Unternehmen an fachlichem *Know-How* im Bereich der Cloud-Computing mangele. Diese stelle eine der größten Hindernisse dar, um einen Wechsel von

[13] Es sei an diser Stelle darauf hingewiesen, dass in diesem Kontext Ahrens die Bedeutung Cloud-basierter Anwendungen im Bereich von deutschen Handelsunternehmen untersuchte (Vgl. Ahrens 2021)[71], sowie das *ifo Institut* anschaulich die strukturellen Veränderungen von der Corona-Pandemie auf den Arbeitsalltag in Deutschland nachzeichnete (Vgl. ifo Institut 2020)[70].

[14] Vgl. Laut Gartner stellt der Magic Quadrant eine zweidimensionale Matrix mit vier Quadranten dar. Jeder Quadrant steht für einen Unternehmenstypus im Markt. Im Uhrzeigersinn von links unten beginnend sind dies: *Nichenanbieter, Herausforderer, Marktführer* und *Visionäre*

[15] Bala et al, 2021, o.S.,[52]

[16] Die aktuellen Marktführer im Bereich der *Cloud-Computing* weltweit sind AWS, Google, Telekom und Microsoft (Synergy Reseach Group, 2019, o.S.[74])

[17] Stern Adam, 2018, The Truth About Cloud Pricing, o.J.[66].

[18] Stern Adam, 2018, The Truth About Cloud Pricing, o.J.[66].

On-Premise- zu Cloud-basierten Systemen gewährleisten zu können.[19].

Laut Anders Lisdorf die Kostenoptimierung für Cloud-Dienste ist ein entscheidender Punkt, da man ohne Optimierungsmaßnahmen mit höheren Kosten rechnen müsse als bei On-Premise Systemen.[20]

> 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.[21]

Diese Bachelorarbeit wird sich ausführlich mit ebendieser Problematik beschäftigen, um herauszustellen, wie Unternehmen mit den passenden Werkzeugen die Kosten ihrer Cloud-Dienste überwachen und optimieren können. Außerdem wird untersucht, wie mit der richtigen Auswahl an Diensten, Kosten optimiert werden. Zudem wird aufgezeigt, welche Maßnahmen nötig sind, um unerwartet hohe Kosten bei Cloud-Diensten zu vermeiden. In diesem Sinne wird untersucht wie die Kosten von Cloud-Diensten minimiert beziehungsweise optimiert werden können. Diese Arbeit wird sich hierbei spezifisch auf die Kostenoptimierung von *Amazon S3* und *EC2-Server-Instanzen* auch mithilfe von folgenden Überwachungswerkzeuge: *Cost-Explorer, CloudWatch* und *Trusted Advisor fokussieren.*

1.3 These

Wenn Unternehmen die Nutzung der Cloud-Dienste und die damit entstehenden Kosten überwachen, sind sie besser in der Lage, Optimierungsmaßnahmen zu ergreifen. Durch die Wahl der korrekten Cloud-Dienste und die Reduktion derer Nutzung, werden dementsprechend auch deren Kosten reduziert.

[19] Vgl. Regenfuß und MalinowskiStern, 2020, Hohe Erwartungen an die Cloud: Hürden meistern, Mehrwert maximieren. o.S.(Webversion) oder S.11 in der PDF-Version auf Englisch[1]
[20] Vgl. Anders Lisdorf, 2021, Cloud Computing Basics: a Non.-Technical Introduction, S.152 [4]
[21] Anders Lisdorf, 2021, Cloud Computing Basics: a Non.-Technical Introduction, S.152 [4]

1.4 Struktur der Arbeit

Diese Bachelorarbeit ist in folgende Kapitel unterteilt:

Kapitel 2 befasst sich mit dem Begriff der *Cloud-Economy* und erläutert das Potenzial der Cloud-Diensten im wirtschaftlichen Sinne. In diesem Kontext sollen die Cloud-Dienste EC2-Instanzen und Amazon S3 kurz dargestellt werden.

Kapitel 3 zeigt die verschiedenen Zahlungsmodelle für EC2-Instanzen. Es werden Kriterien vorgestellt, die dazu beitragen sollen, 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.

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

2 Grundlagen

In diesem Grundlagenkapitel werden Erfolgschancen für Unternehmen aufgelistet, die Cloud-Dienste in ihre Geschäftsprozesse integrieren. Mit Cloud-Diensten sind die Dienste eines beliebigen Cloud-Anbieters im Allgemeinen gemeint und nicht ausschließlich AWS-Dienste. 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 Kosten verursachen 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, den sogenannten *TCO*.[22] [23]

2.1 Cloud Economics

Cloud Economics befasst sich mit den Kosten und den Vorteilen von Cloud Computing und die dahinterstehenden wirtschaftlichen Grundsätzen. Anhand des *Pay-as-you-go-Modell (PAYG)* können zum Beispiel nur die Cloud-Dienste in Anspruch genommen werden, die in dem Moment für das Unternehmen verbraucht werden. Damit entfällt die Notwendigkeit hohe Investitionen in Hardware zu tätigen, wie bei On-Premise-Systemen, wo Hardware im Voraus für den künftigen Bedarf.[24] Durch den Verzicht auf Hardware entfallen die Kosten für Reparatur und Wartung. Die Cloud-Anbieter übernehmen dabei viele Verwaltungsaufgaben. Laut Larry Carvalho und Matthew Marden führe dies zu einer Abnahme der nötigen Fachkräften.[25] So ist die die Nutzung von Cloud-Diensten in unabhängiger Weise möglich; in Selbstbedienung und mit der Freiheit Dienste ohne Einschränkungen zu gebrauchen. Das bedeutet jedoch gleichzeitig, dass die Nutzerin oder der Nutzer von Cloud-Diensten Verantwortung für die anfallenden Kosten übernehmen.

[22] TCO steht für *Total Cost of Ownership*(Vgl. Gartner, o.J., o.S.[64].)
[23] Vgl. Ubuntu, delivered by Canonical: A business guide to hybrid/multi-cloud, S.2.[46]
[24] Vgl. Anders Lisdorf, 2021, Cloud Computing Basics: a Non.-Technical Introduction, S.23[4]
[25] Larry Carvalho and Matthew Marden, 2015, Quantifying the Business Value of Amazon Web Services, S.1[48]

2.1.1 Skalierbarkeit

Hierbei bezieht sich diese Arbeit auf die Möglichkeit, die Kapazität von Cloud-Diensten zu skalieren. Um die Leistung der IT-Infrastruktur aufrecht zu halten, ist es zum Beispiel möglich, das Serversystem so zu konfigurieren, dass es auf wechselnde Lastanforderungen reagiert. Auf diese Weise kann Zeit mit der Verwaltung von IT-Infrastruktur eingespart werden, welche dann genutzt werden kann, um sich auf die wesentlichen Geschäftsaktivitäten zu konzentrieren. [26] Dies war der Fall bei *Walgreens* im Jahre 2020 in den Vereinigten Staaten. Sie haben unter anderem 750 virtuelle Maschinen und *SAP HANA* auf *Azure Instanzen* migriert. Diesbezüglich kommentierte Dan Regalado:

> 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.[27]

So erklärte Dan Regalado, dass *Walgreens* mit dem Einsatz von kleinen Instanzen und Auto-Scaling eine Serverinfrastruktur erreicht hat, die sich dem Bedarf an Rechenkapazität anpasst.

2.1.2 Flexibilität

Hiermit ist die Möglichkeit gemeint, wenn nötig und unter den Bedingungen des Cloud-Anbieters, Cloud-Dienste in Auftrag zu geben und kündigen zu können. Für Cloud-Dienste gibt es im Allgemeinen eine Vielzahl von Optionen, von denen einige Beispiele unten aufgeführt werden:

- Verschiedene Betriebssysteme, ohne oder mit Lizenzierung.
- Die meistverbreiteten Programmiersprachen, unter anderem *Java, C++, Go, JavaScript und Python.*[5]

[26] Mark Wilkins, 2021, AWS Certified Solutions Architect - Associate (SAA-C02), S.29.[1]

[27] Microsoft, 2020, Customer Story-Walgreens Boots Alliance delivers superior customer service with SAP solutions on Azure, o.S. [41]

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

Durch die Vielzahl der verfügbaren Diensten ist es möglich, Prototypen und Experimente in kurzer Zeit durchzuführen.[28] Softwareprojekte können schnell auf den Markt gebracht werden. Je nach ihrem Erfolg ist es möglich, sinnvolle und kosteneffizientere 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.

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 oder die Vertriebsabteilung des Cloud-Anbieters benötigt werden.[29] Andererseits besteht ebenso die Gefahr, dass hohe ungewollte Kosten entstehen, wenn jemand versehentlich oder in unverantwortlicher Weise Dienstleistungen in Anspruch nimmt.

2.1.4 Keine Vorabkosten

Das Pay-as-you-go-Modell (PAYG) wird von einer Reihe von Cloud-Anbietern angeboten.[30] Dieses erfordert keine Vorauszahlungen für die Nutzung den verschiedenen Cloud-Diensten. Wenn nur für die monatlich verbrauchten Diensten bezahlt wird, verringert sich zudem die Anfangsinvestition in die IT-Infrastruktur oder fällt ganz weg. Dies ist besonders für kleine Unternehmen bedeutsam, die nicht über die finanziellen Mittel verfügen, um in eine IT-Infrastruktur zu investieren. Es besteht jedoch die Möglichkeit, bestimmte Beträge für die zu konsumierende Dienste im Voraus zu bezahlen.[31]

2.1.5 Technische Fachkompetenz

Bei einem Einsatz von Cloud-Diensten ist zu bedenken, dass weitere Investitionen wie technische Schulungen für das Personal erforderlich werden. Der *TÜV Rheinland* bietet

[28] Vgl. IDC, 2015, Business Value of AWS S.7[48]

[29] Vgl. Anders Lisdorf, 2021, Cloud Computing Basics: a Non.-Technical Introduction, S.28[4]

[30] Die aktuellen Marktführer im Bereich der *Cloud-Computing* weltweit sind AWS, Google, Telekom und Microsoft (Vgl. Synergy Reseach Group 2019, o.S.[74]).

[31] Im Unterkapitel 3.2 wird eine Berechnung der Einsparungen durch die teilweise oder vollständige Vorauszahlung der Kosten für die Nutzung von Serverinstanzen gezeigt.

zum Beispiel Kurse zur Ausbildung von *Cloud Architekten* an. Die Kurse dauern i.d.R. drei Tage und kosten 2.136,05 € pro Teilnehmer. Maßnahmen wie die genannten Kurse wirken so einem der Hauptprobleme entgegen, mit denen Unternehmen bei der Migration in die Cloud konfrontiert werden. Regenfuß und Malinowski verdeutlichen das in einer von *Accenture* im Jahr 2020 durchgeführten Umfrage. 38% der Befragten gaben an, dass fehlende Kompetenzen im Unternehmen in Bezug auf die Cloud ein Hindernis für eine Cloud-Migration seien.[32]

2.2 Amazon Cloud-Dienste

Im Folgenden liegt der Fokus auf *AWS-Diensten*. Einer der am häufigsten genutzten AWS-Dienste ist *Amazon Elastic Computing Instances EC2*, mit dem virtuelle Maschinen erstellt werden können.[33] Ein weiterer wichtiger AWS-Dienst ist *Amazon Simple Storage Service (S3)*, der zum Speichern von Objekten verwendet wird.[34] [35]

Wie Lynn Langit, eine erfahrene Cloud-Architektin, feststellte, könne bis zu 80% der AWS-Rechnung aus Gebühren für EC2-Instanzen bestehen.[36] Laut des *AWS Solutions Architekten* Daniel Peña Silva ist Amazon S3 einer der am häufigsten genutzten AWS-Dienste.[37] Deshalb fokussiert sich diese Arbeit auf die Überwachungs- und Optimierungsmaßnahmen, hauptsächlich für EC2-Instanzen und Amazon S3.

Wie aus der Abbildung 2 hervorgeht ist, werden darüber hinaus seit 2020 weltweit mehr Daten in Serverfarmen als auf lokalen Geräten gespeichert.[38] Dies bietet einerseits Vorteile im Bezug auf die Geschwindigkeit der Arbeitsabläufe, andererseits birgt aber auch Risiken wie Datendiebstahl. Das Thema Datendiebstahl wird in dieser Arbeit nicht behandelt.

[32] Regenfuß und Malinowski (Accenture), 2020, Hohe Erwartungen an die Cloud: Hürden meistern, Mehrwert maximieren, S.11[1]

[33] Kimberly Mlitz, 2021, Cloud infrastructure services vendor market share worldwide from 4th quarter 2017 to 3rd quarter 2021, o.S.[72]

[34] Objekte sind in AWS die Grundeinheit in welchen Dateien in den Amazon S3-Speichereinheiten gespeichert werden. Neben den Objekten werden Metadaten, wie das Datum der Objekterstellung und das Datum der letzten Aktualisierung gespeichert.(Vgl. Amazon Simple Storage Service User Guide, S.4[20])

[35] Amazon Elastic Computing Instances EC2 werden im Folgenden als *EC2-Instanzen* und Amazon Simple Storage Service als *Amazon S3* oder *Amazon S3-Speichereinheiten* bezeichnet.

[36] Lynn Langit, 2021, LinkedIn Learning: AWS Controlling Cost. Minute 0:20-0:45[57]

[37] Vgl. Daniel Peña Silva, 2021, LinkedIn: Listado de todos los Servicios de AWS.[56]

[38] Vgl. Statista: 2020 überholt die Cloud lokale Speichermedien.[69]

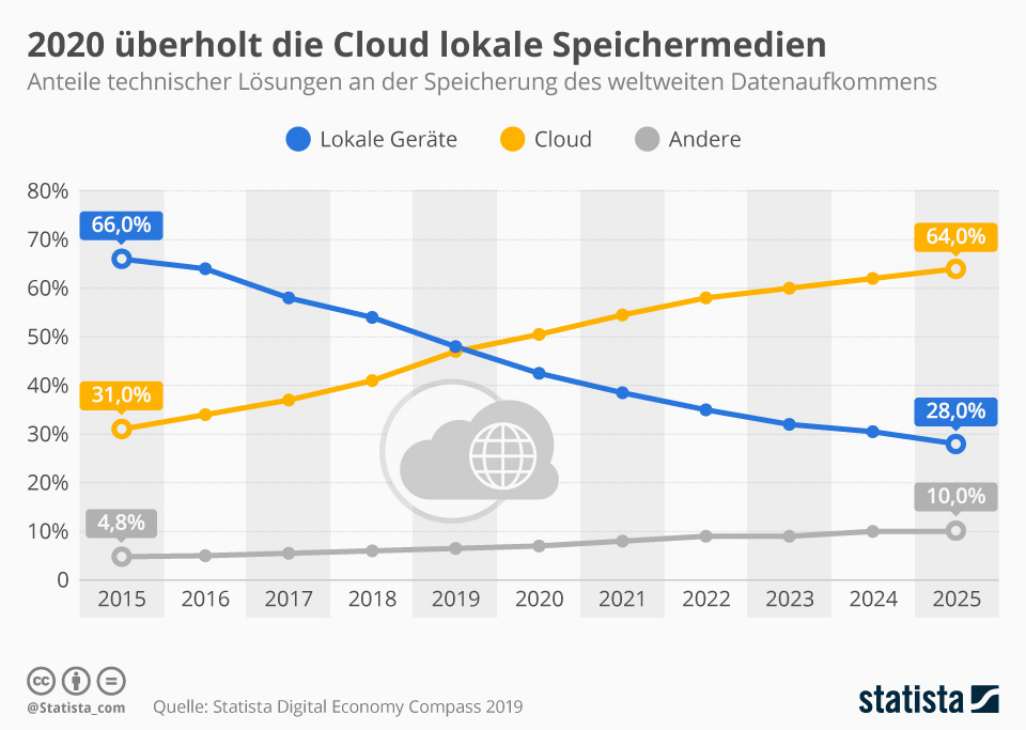


Abbildung 2
2020 überholt die Cloud lokale Speichermedien, Statista, 2019 [69]

Dieses grundlegende Kapitel hat einige potenzielle Vorteile der Nutzung von Cloud-Diensten für Unternehmen aufgezeigt. Darüber hinaus geht der Trend in den letzten Jahren zur Nutzung von Cloud-basierten Diensten. Das nächste Kapitel befasst sich mit den Zahlungsmodellen für EC2-Instanzen und den damit einhergehenden Abwägungen, die bei der Wahl dieser Modelle in verschiedenen Szenarien zu berücksichtigen sind.

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, es ist daher die teuerste Alternative, die auf Stundenbasis berechnet wird. Die Modelle *Saving Plans* und *reservierte Instanzen(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. Somit weist jedes Zahlungsmodell seine Vor- und Nachteile auf und eignet sich für unterschiedliche Anwendungsfälle.[39] Gute Ergebnisse können auch durch die Kombination mehrerer Zahlungsmodelle erzielt werden.[40]

Die beste Leistung wird außerdem angestrebt, indem sich diese Instanzen in räumlicher Nähe zur Mehrzahl der Endnutzer, befinden.

3.1 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.[41]

Die Preise beim dem On-Demand Zahlungsmodell variiert je nach Instanz Typ, Region und der übertragenen Datenmenge.[42] In der Abbildung 3 werden Preisbeispiele für die Region Ohio verfügbaren Linux-Instanzen gezeigt.

[39] 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 AWS-Diensten, mit denen sie kommunizieren werden, zu platzieren.

[40] Dieser Aspekt wird in Unterkapitel 3.4 ausführlicher dargestellt.

[41] Vgl. AWS, 2019, Amazon Elastic Compute Cloud - Benutzerhandbuch für Linux-Instances, S.344[28].

[42] Die aktuellen Preise für die verschiedenen Regionen sind auf der AWS-Website in der Sektion EC2 - On-Demand-Preise zu finden. Vgl. AWS On-Demand Instances Pricing.[4]

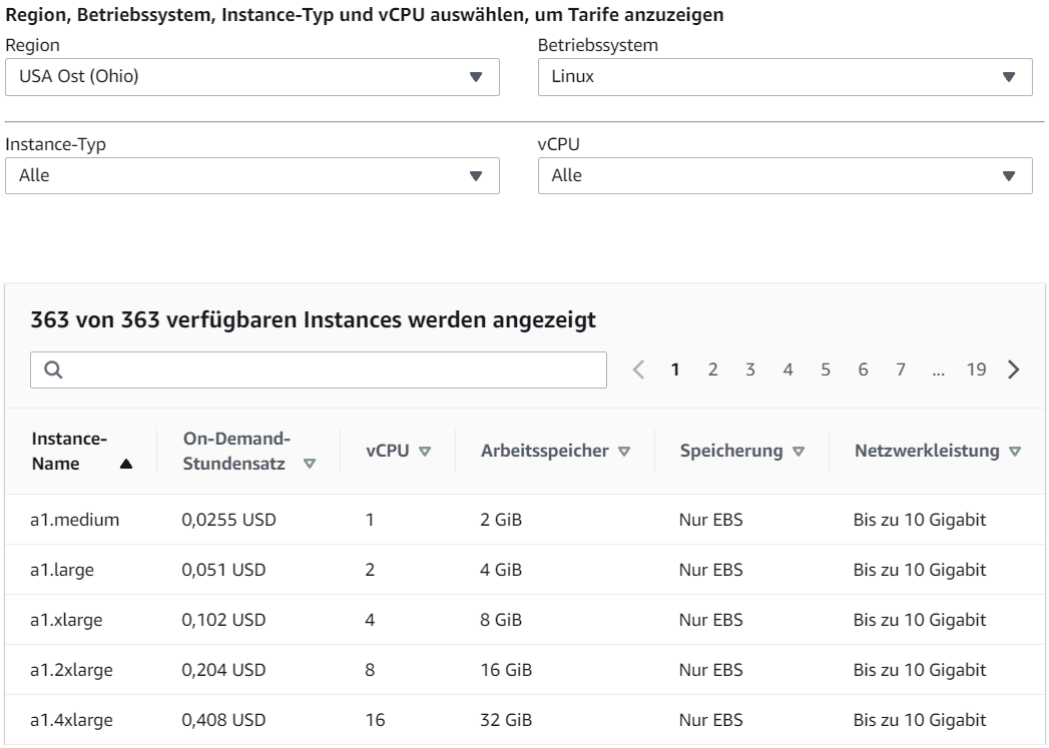


Abbildung 3
On-Demand Preisbeispiele von EC2-Instanzen.
Quelle: AWS Pricing Calculator for EC2, 2021, o.S, [4].

Es ist zu beachten, dass Instanzen mit denselben Eigenschaften (Instanz-Familie, Arbeitsspeicher, Netzwerkleistung usw.), 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 US-Dollar pro Stunde gemessen wird.[43] Um die reduzierten Preise zu bekommen, müssen Verträge über ein oder drei Jahre abgeschlossen werden.

Abbildung 4 zeigt die möglichen Einsparungen je nach Zahlungsmodell. Die Einsparungen hängen davon ab, ob man die Instanz-Familie und die Verfügbarkeitszone später verändern kann oder nicht. Je geringer die Flexibilität für spätere Änderungen, desto höher die Einsparungen.

[43] AWS-Dienste werden in US-Dollar abgerechnet. Zahlungen in anderen Währungen sind auch möglich. Quelle: AWS, AWS-Console in Kontoeinstellungen, 2021, o.S.

Mögliche Einsparungen laut AWS			
Saving Plans		Reserved Instances	
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%

Abbildung 4
Mögliche Einsparungen bei reservierten Instanzen and Saving Plans.
Quelle: AWS, 2021, o.S., [8, 12].

Die *Compute Saving Plans* bieten die Flexibilität *die Familie, die Größe, die Verfügbarkeitszone (AZ), das Betriebssystem* oder *der Mandant* von EC2-Instanzen zu wechseln.[44][45] Diese Option ist bei *EC2-Instance Saving* nicht möglich und daher bietet die zweite Option eine etwas höhere Einsparung.

> „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." [46]

Bei den EC2-Instance Saving Plans hingegen muss eine Instanz-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.

EC2 Reserved Instance Marketplace

Sollte sich herausstellen, dass die Kapazität der reservierten Instanzen viel zu wenig oder gar nicht genutzt wird, kann diese Rechenkapazität auf dem *RI Marketplace* (Marktplatz für den Kauf von reservierten Instanzen) zur Verfügung gestellt werden. Somit kann ein Teil der Investition zurückgeholt werden. Dies ist für Standard reservierten Instanzen möglich. Diese werden in Spot-Instanzen umgewandelt, damit andere Nutzer sie beantragen können. Dafür sollte der Instanz-Anbieter eine Servicegebühr in Betracht ziehen.[47]

[44] Vgl. AWS, 2021, AWS Saving Plans Pricing, o.S.[12].
[45] Vgl. Mark Wilkins, 2021, AWS Certified Solutions Architect - Associate (SAA-C02), S.95.[1].
[46] Vgl. AWS, 2021, AWS Saving Plans Pricing, o.S.[12].
[47] Stand November 2021 beträgt diese Gebühr 12% (Vgl. AWS, 2021, Amazon EC2 Reserved Instance Marketplace, o.S.[25]).

Optionale Vorauszahlung

Zusätzlich ist es bei Saving Plans und reservierten Instanzen möglich im Voraus zu zahlen. Im Gegenzug wird ein niedrigerer Gesamtpreis angeboten. AWS bietet diesbezüglich drei verschiedene Optionen an: eine teilweise, keine oder eine vollständige Vorauszahlung.[48] Bei teilweiser Vorauszahlung ist eine Anzahlung von etwa 50% zu leisten.

Die Abbildung 5 zeigt den Vergleich zwischen den drei Vorauszahlungsoptionen für 20 Instanzen über drei Jahren im Zahlungsmodell Saving Plan. Hier wird deutlich, dass es kaum einen Unterschied zwischen einer teilweisen und keinen Vorauszahlung gibt. Eine erhebliche Einsparung ergibt sich jedoch, wenn man für den gesamten Zeitraum der gebuchten Instanzen im Voraus bezahlt.

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 5
Mögliche Einsparungen durch Vorauszahlungen für
EC2 Instanzen in Saving Plans Zahlungsmodell.
Eigene Darstellung.
Quelle: AWS, 2021, AWS Pricing Calculator, o.S.[18].

Die Berechnungen wurden anhand des *AWS Pricing Calculator* für Instanz-Familie *t4g.xlarge* und in der Region EU (Frankfurt) durchgeführt.[18]

[48] Vgl. AWS, 2021, AWS Pricing Calculator o.S.[18].

3.3 Spot-Instanzen

Wie in Unterkapitel 3.2 genannt, bieten EC2 Spot-Instanzen die Möglichkeit aus den ungenutzten EC2-Instanzen anderer Nutzer zu profitieren. Mit einem Preisvorteil von bis zu 90% gegenüber *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 werden, wird einen Maximalpreis festgelegt. Ist der Preis der Spot-Instanz höher als der angegebene Maximalpreis, ist die Spot-Instanz für die aktuelle Einstellung nicht mehr verfügbar. Ein anderes Szenario tritt ein, wenn der Instanz Anbieter die Spot-Instanz erneut anfordert. Falls eine Spot-Instanz unterbrochen wird, benachrichtigt AWS den aktuellen Nutzer darüber zwei Minuten im Voraus. Dieses Ereignis ist ebenfalls auf *CloudWatch* verfügbar, damit weitere Alarme eingestellt werden können.[49] Da Spot-Instanzen anfällig für Unterbrechungen sind, ist es nicht empfehlenswert, für Produktionsumgebungen nur Spot-Instanzen zu verwenden.

3.4 Amazon EC2 Fleet

Instanzen-Flotten oder auf Englisch *fleet of instances*, bieten bei AWS die Möglichkeit mehrere Spot-Instanzen anzufordern, um einen bestimmten Bedarf an Rechenleistung zu decken.[50] Spot-Instanzen können aber auch für produktive Umgebungen verwendet werden.[51] Darüber hinaus ist es empfehlenswert, Instanzen aus verschiedenen Zahlungsmodellen zu kombinieren, um von den Einsparungen von der verwendeten Spot-Instanzen, Saving Plans und reservierten Instanzen zu profitieren. Die Kombination von Instanzen aus verschiedenen Zahlungsmodellen auf Produktionsumgebungen wirkt dem Nachteil entgegen, der mit Spot-Instanzen verbunden ist. Da, bei Unterbrechung von Spot-Instanzen, bleiben noch Instanzen anderer Zahlungsmodelle weiterhin in Betrieb.

Folgende Punkte sind für die Nutzung von Spot Fleet Instanzen zu berücksichtigen:

[49] Diese und andere Funktionalitäten von CloudWatch werden in Kapitel 4 näher erläutert.
[50] Vgl. AWS, 2019, Amazon Elastic Compute Cloud - Benutzerhandbuch für Linux-Instances, S.708[28].
[51] Vgl. Isaac Vallhonrat, 2020, Running Web Applications on Amazon EC2 Spot Instances, o.S.[26].

Wahl der Spot-Instanzen

Die zu berücksichtigenden Instanzen für die Instanzen-Flotte, müssen den Anforderungen der Applikation entsprechen. Um die Wahrscheinlichkeit zu erhöhen, dass mehr Spot-Instanzen gefunden werden, ist es daher empfehlenswert, die Kriterien der Suche zu erweitern. Dies kann erreicht werden, indem Instanzen ähnlicher Typen einbezogen werden. Die Berücksichtigung von den Instanzen-Familien, welche eine höhere Leistung aufweisen als erfordert wird, ist ebenfalls eine gute Option. Da in diesem Falle der Preis für Spot-Instanzen trotz höherer Leistung geringer sein wird, als der von den On-Demand-Instanzen[26].

Maximaler Stundenpreis

Wie im Unterkapitel 3.3 erwähnt, muss für die Anforderung von Spot-Instanzen ein Maximalpreis festlegt werden. In diesem Fall ist die Festlegung dieses Maximalpreises auch für die gesamte Instanzen-Flotte eine Option. Es kann erwartet werden, dass die Spot-Preise im Laufe der Zeit stabil bleiben, da sie keinen starken Preisschwankungen unterliegen. Diese Informationen sind nur mit einem AWS-Konto zugänglich.[52]

Festlegung von On-Demand-Anteil

Wenn alle oder eine große Anzahl von Spot-Instanzen nicht mehr verfügbar sind, muss die benötigte Rechenkapazität von Instanzen anderer Zahlungsmodellen, wie On-Demand abgedeckt werden. Die Standardeinstellungen liegen bei 70% On-Demand-Instanzen und 30% Spot-Instanzen.[53] Im Fall von vorhandenen reservierten Instanzen oder Instanzen von Saving Plans werden On-Demand-Instanzen zum entsprechend reduzierten Preis berechnet.[54]

Auto Scaling Groups

Auch als *Auto-Scaling-Gruppe*(ASG) bezeichnet und ist für die Skalierung der zu startenden Instanzen verantwortlich. Hierfür wird eine Startkonfiguration benötigt, welche definiert, unter welchen Bedingungen Instanzen gestartet oder beendet werden sollen. In der Startkonfiguration werden unter anderem der *Instanztyp, Security-Groups,* und *Tags* festgelegt.[55] Für die Nutzung von EC2-Flotten und Auto Scaling-Gruppen fallen keine

[52]Die aktuellen Preis und der Preisverlauf von Spot-Instanzen können in auf der AWS-Konsole abgefragt werden. (Vgl. AWS, 2021, EC2 Spot Instanzen-Anfragen und Preisverlauf o.S. [27]).

[53]Vgl. Isaac Vallhonrat, 2020, Running Web Applications on Amazon EC2 Spot Instances, o.S.[26]

[54]Vgl. AWS, 2019, Amazon Elastic Compute Cloud - Benutzerhandbuch für Linux-Instances„ S.690[28].

[55]Mehr zu dem Thema Auto-Scaling und seine verschiedenen Konfigurationen findet sich in Kapitel 5.

zusätzlichen Kosten an. Lediglich müssen die von der EC2-Instanzen verursachten Kosten getragen werden.[56]

3.5 Anwendungsfall: TrueCar

Instanzen in Zahlungsmodellen, die zu zeitlichen Verpflichtungen führen, bergen die Gefahr, dass die benötigte Rechenkapazität mittel- bis langfristig falsch eingeschätzt wird. Einerseits kann die reservierte Rechnerkapazität zu gering eingeschätzt werden. Als Konsequenz daraus wird die Rechnerkapazität großenteils mit On-Demand-Instanzen abgedeckt. Diese konnte im Anteil der reservierten Instanzen berücksichtigt und mit reduzierten Preisen berechnet werden. Andererseits, wenn zu viel Rechnerkapazität mit reservierten Instanzen reserviert und diese zu wenig gebraucht wird. Besteht die Möglichkeit, dass es die reine Nutzung von On-Demand-Instanzen eine kostengünstigere Option darstellt.

Im Folgenden wird näher auf die Strategie von *TrueCar Inc.* eingegangen, durch dessen Anwendung beide der oben genannten Problematiken vermieden werden können. TrueCar Inc. ist eine Preis- und Informations-Website für Neu- und Gebrauchtwagenkäufer mit Sitz in Santa Monica, Kalifornien.Dank einer Optimierungsstrategie konnten sie ihre AWS-Kosten durch die Nutzung reservierter Instanzen um etwa 40% senken.[57]

Um eine solche Einsparungen zu erreichen, musste das Team zuerst verstehen, wie AWS-Dienste wie *reservierte Instanzen, Cost-Explorer, Auto-Scaling-Gruppen* und *Lambda Funktionen* funktionieren. Damit haben sie eines der häufigsten Hindernisse überwunden mit denen Unternehmen bei der Nutzung von Cloud-Diensten konfrontiert werden; den Mangel an technischem Wissen in Bezug auf Cloud-Dienste.[58] Nachdem das TrueCar-Team die notwendigen Informationen, insbesondere über die reservierten Instanzen, verstanden haben, wurde anschließend die benötigte Rechenkapazität ermittelt.[59] Die Kosten der Instanzen in On-Demand wurden mit dem von reservierten Instanzen gegenübergestellt, um den *Break-Even-Point* dazwischen zu finden.[60] Der Schnittpunkt zwischen dem Preis der

[56] Vgl. AWS, 2019, Amazon Elastic Compute Cloud - Benutzerhandbuch für Linux-Instances, S.709[28].

[57] Vgl. David Wang, 2019, How TrueCar Saves 40% on AWS with EC2 Reserved Instances. O.s. [59].

[58] Vgl. Accenture Dienstleistungen GmbH. Hohe Erwartungen an die Cloud: Hürden meistern, Mehrwert maximieren, S.11[1].

[59] In dem Artikel über TrueCar wurde nicht explizit erläutert, wie die benötigte Rechnerkapazität berechnet wurde. Diese Informationen werden jedoch von Cost-Explorer angezeigt. Cost-Explorer bietet die Möglichkeit, die Nutzung der AWS-Diensten für die letzten 12 Monate anzuzeigen. Ausführlicher wird diese Thematik in Unterkapitel 4.2 behandelt werden.

[60] Vgl. Marceil Schweitzer und Ernst Troßmann: Die Break-Even-Analyse dient als Entscheidungshilfe für das Management. Bei einer Break-even-Analyse geht es immer um eine Gegenüberstellung positiver

reservierten Instanzen und von der On-Demand Instanzen bildet den sogenannten Break-Even-Point. Nach diesem sinke der monatliche Preis für die reservierten Instanzen, bis die reservierte Kapazität verbraucht wird oder der Zeitraum für die reservierten Instanzen endet.

Wie in der Grafik der Abbildung 6 dargestellt wird, liegt der Break-Even-Point zwischen dem Monat acht und neun. Nach diesem Punkt ist der Preis für reservierte Instanzen niedriger als der für On-Demand-Instanzen. Die Berechnung wurde für den Zeitraum von einem Jahr durchgeführt.

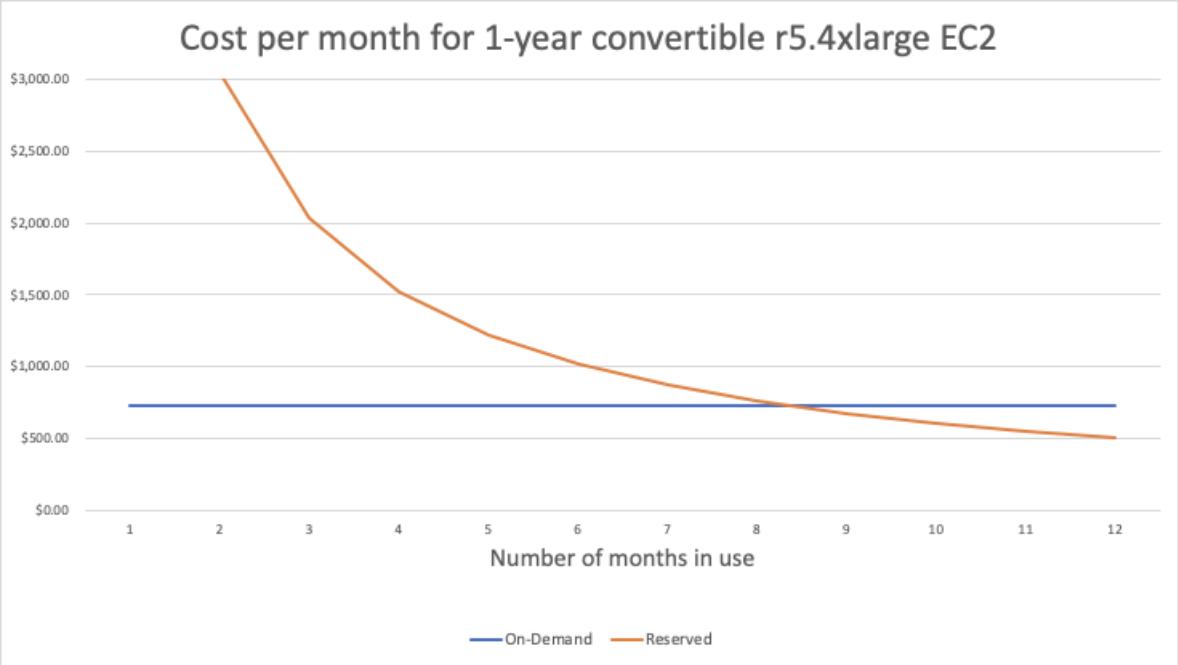


Abbildung 6
Monatliche Kosten für eine On-Demand-Instanz
im Vergleich zu einer reservierten Instanz.
Quelle: David Wang, 2019, How TrueCar Saves 40% on AWS with EC2 Reserved Instances. O.s.[59]

Nach der Buchung der reservierten Instanzen wurde deren Nutzung anhand von zwei Metriken mit Cost-Explorer überwacht.

1) Der **RI-Coverage** zeigt an, wie viel der On-Demand-Instanzen durch reservierte Instanzen abgedeckt wird. Ziel ist hierbei das *RI-Coverage* (Abdeckung der reservierten Instanzen) so nahe wie möglich an 100% zu halten.

und negativer Wirkungen von Maßnahmen, S.14.[2]

2) Die **RI-Utilization** zeigt an, wie viel Prozent der reservierten Instanzen verbraucht wurden. Hierbei wird versucht, die RI-Utilization nicht zu niedrig zu halten. Stattdessen sollte die Nutzung von On-Demand-Instanzen gering gehalten werden.

Der kürzlich vorgestellte Anwendungsfall hat gezeigt, wie die Verwendung reservierter Instanzen zu erheblichen Einsparungen bei TrueCar Inc. geführt hat. Ein wesentlicher Bestandteil dieser Optimierungsstrategie ist die Berechnung des Break-Even-Point, um zu wissen, wie viel Rechenkapazität reserviert werden sollte. Außerdem wurden die dafür nötige Metriken vorgestellt, mit denen die Ergebnisse der Optimierungsmaßnahme überwacht werden.[61]

Vergleich der Zahlungsmodelle

Die folgende Tabelle fasst die Eigenschaften der Zahlungsmodellen für die On-Demand-, reservierte, Saving-Plans- und Spot-Instanzen zusammen und listet typische Applikationen je nach Zahlungsmodell auf.

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 7
Vergleich der Zahlungsmodelle nach Eingenschaft und Anwendungsfall. Eigene Darstellung.
Quelle:Plusserver, 2021, Kostenoptimierung in AWS, S.9.[61]
Spot by NetApp, 2021, What are AWS spot instances?, o.S.[67]
AWS, 2021, On-Demand, Reserved Instances, Saving Plans and Spot-Instances, o.S.[4, 8, 12, 21]

[61] Um diese Metriken im Blick zu behalten und nicht jeden Tag den Cost-Explorer aufrufen zu müssen, wurde eine Benachrichtigung an *Slack* eingerichtet. Slack ist eine Messaging-App für Unternehmen. (Vgl. Slack, o.J., Was ist Slack?, o.S.[65].) Dies war über die Cost-Explorer API und eine Lambda-Funktion möglich.

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, um die Preisvorteile von den Zahlungsmodellen zu nutzen. Beginnend mit dem On-Demand-Zahlungsmodell, gefolgt von Reserved Instanzen und Saving Plans.[62] In dieser Reihenfolge sinkt der Preis und mit ihm steigt die Verpflichtung, sich langfristig zu binden. Schließlich wurden Spot-Instanzen vorgestellt, die die niedrigsten Preise bieten, aber keine volle Verfügbarkeit sicherstellen. Darüber hinaus wurde ein Anwendungsfall vorgestellt, der erhebliche Einsparungen bei der Verwendung reservierter Instanzen zeigt.

Im nächsten Kapitel werden CloudWatch, Cost-Explorer und Trusted Advisor vorgestellt. Diese Werkzeuge sollen ein besseres Verständnis über die Nutzung und Kosten von AWS-Diensten, die Analyse von Metriken ermöglichen und Empfehlungen zur Kostenoptimierung geben.

[62] Für das On-Demand-Zahlungsmodell gibt es keine Kostenreduzierung, aber wie in Unterkapitel 5.1.1 gezeigt gibt Maßnahmen, um die Nutzung von Instanzen zu reduzieren.

4 Kostenüberwachung

CloudWatch sammelt Metriken von AWS-Diensten und bietet die Möglichkeit, Alarme und Aktionen zu konfigurieren, die wiederum AWS-Diensten auf der Grundlage dieser Metriken betreffen. Für die Visualisierung von Metriken bietet CloudWatch die Erstellung von personalisierten Dashboards. Cost-Explorer konzentriert sich auf die Überwachung der Nutzung von AWS-Diensten und der dadurch verursachten Kosten. Diese bietet die Möglichkeit Kosten- und Nutzungsberichte der AWS-Diensten zu erstellen. Solche Informationen dienen für die Budgetierung, die Verfolgung von KPIs und Entscheidungsfindung in Bezug auf die operative Planung in Unternehmen.[63] Der *Trusted Advisor* bietet konkrete Empfehlungen auf der Grundlage von AWS *Best-Practices* und individuelle Prüfungen von AWS-Diensten.

Es existieren weitere Überwachungswerkzeuge bei AWS, auf die in dieser Arbeit nicht eingegangen wird, weil sie einen anderen Fokus als die Kostenüberwachung und -optimierung haben. Zum Beispiel CloudTrail, welches für die Überwachung von Governance, Compliance, Betrieb und Risiken im AWS-Konto ist. Mit CloudTrail können Benutzeraktivitäten über AWS-Dienste durch Ereignisse verfolgt werden.[64] Ein weiteres Werkzeug ist *AWS X-Ray*, welches zur Überwachung von Anwendungsleistung verwendet wird. Dies unterstützt Entwickler bei der Analyse und Fehlersuche in verteilten Produktionsanwendungen. Mit X-Ray kann man herausfinden, wie gut Anwendungen und ihnen zugrunde liegenden Dienste funktionieren. Auf diese Weise können Ursache von Leistungsproblemen und Fehlern ermittelt und behoben werden.[65]

[63] Die vorgenannten Konzepte werden in Unterkapitel 4.2 näher erläutert.
[64] Vgl. AWS, 2021, CloudTrail User Guide Version 1.0: What Is AWS CloudTrail?, S.1 [29]
[65] Vgl. AWS, 2021, X-Ray Developer Guide: What is AWS X-Ray?, S.1[29],

Tagging-Strategie

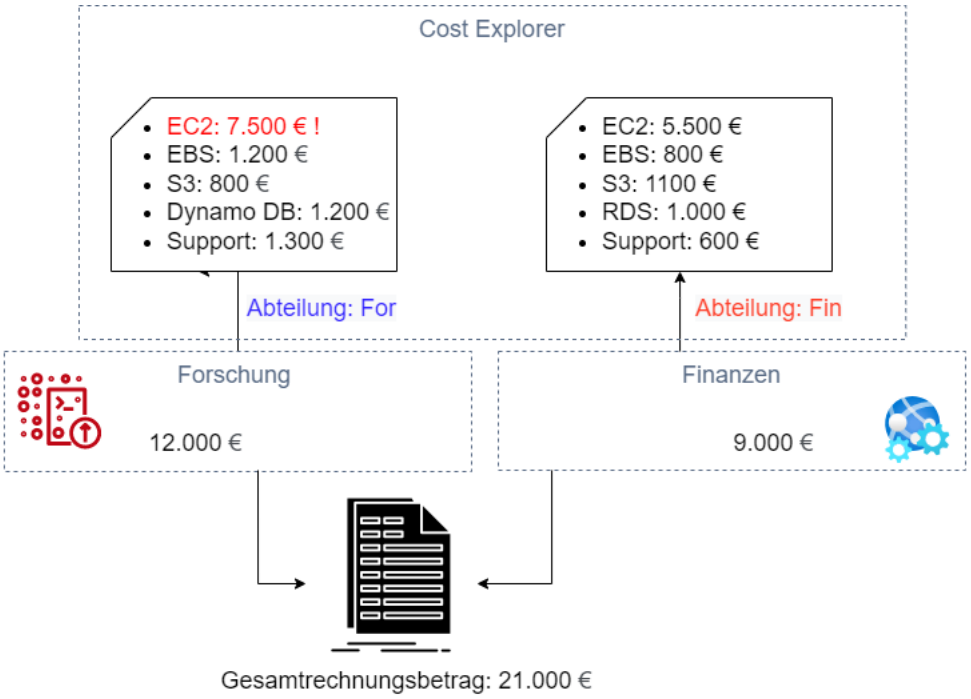
Die *Tags* sind bei AWS Informationen in Form von Metadaten, die an AWS-Dienste zugewiesen werden können.[66] Ein Tag besteht aus einem *Tag-Schlüssel* und einem *Tag-Wert*. Beispiele für Tag-Schlüssel sind *Abteilung, Projekt, Team, Region, Art des Dienstes* und *Umgebung*. Tag-Werte für den Tag-Schlüssel *Abteilung* könnten *Buchhaltung, Finanz, Entwicklung* oder *Marketing* sein. Sowohl bei Tag-Schlüssel als auch bei Tag-Werte wird zwischen Groß- und Kleinschreibung unterschieden.

Anwendungsbeispiel für Tags und Cost-Explorer

Durch die Verwendung von Tags ist es möglich, die Kosten auf den von der Organisation festgelegten Tags zu verfolgen. Es könnte zum Beispiel ein Szenario entstehen, in dem eine Abteilung innerhalb einer Organisation mehr Kosten verursacht als eine Andere. Dies ist nur durch den Anstieg der von AWS generierten Rechnung erkennbar, um den Grund für diesen Anstieg genauer zu verstehen, muss ihre genaue Ursache untersucht werden. Werkzeuge wie Cost-Explorer zusammen mit einer Tag-Strategie machen diese Art von Analyse möglich.

In der Abbildung 8 wird ein Szenario vorgestellt, wo die Kosten für EC2-Instanzen der Forschungsabteilung kontinuierlich angestiegen sind. Mitarbeiter der Forschungsabteilung waren nicht in der Lage, die Kostensteigerungen zu begründen. Um die von den einzelnen Abteilungen verursachten Kosten zu trennen, wurde ein Tag-Schlüssel mit dem Namen *Abteilung* angelegt. Um anschließend jedem AWS-Dienst einen entsprechenden Tag-Wert seiner Abteilung zuzuweisen. Mit Hilfe des Cost-Explorer konnte festgestellt werden, dass die Kosten für EC2 der Forschungsabteilung im Laufe der Zeit gestiegen sind.

[66] Vgl. AWS: AWS – Allgemeine Referenz - Referenzhandbuch. S.681[31]



	Monatliche Kosten pro Abteilung		
Monat	**Forschung**	**Finanzen**	**Gesamtkosten**
Mai 2021	€5,600.00	€8,900.00	€14,500.00
Juni 2021	€6,000.00	€8,300.00	€14,300.00
Juli 2021	€7,500.00	€8,000.00	€15,500.00
August 2021	€9,000.00	€9,200.00	€18,200.00
September 2021	€12,000.00	€9,000.00	€21,000.00

Abbildung 8
Trennung der Abteilungskosten durch Tags.
Quelle: eigene Darstellung.
Die Angaben dienen nur als Beispiel und entsprechen keiner realen IT-Infrastruktur.

Im vorliegenden Fall wurde festgestellt, dass Gastpraktikanten in der Forschungsabteilung Experimente durchführt hatten, in den EC2-Instanzen genutzt wurden. Die Instanzen wurden nach Beendigung des Aufenthalts nicht mehr abgeschaltet und haben kontinuierlich Kosten verursacht. So wurde für diesen hypothetischen Fall die Ursache für den Anstieg der Gesamtkosten einer einfachen Organisation mit zwei Abteilungen und wenigen Cloud-Diensten ermittelt. [67] Für

[67] Es gibt Unternehmen mit viel komplexeren Strukturen als diese, die weitaus mehr Cloud-Dienste in Anspruch nehmen.

Unternehmen ist eine Tagging-Strategie von Relevanz, um Kosten (Ausgaben[68]) für die Buchhaltungsabteilung genauere Daten zu ermitteln und um Budgets auf der Grundlage früherer Projekte erstellen zu können. Die Kostenüberwachung ist mit einer Tag-Strategie auf einer detaillierten Ebene möglich. Je nach festgelegten Tags können detaillierte Analysen der Cloud-Nutzung und -Kosten über Produkte, Einheiten, Umgebungen oder beliebige andere Bereiche hinweg erstellt werden.[69]

4.1 AWS CloudWatch

Amazon CloudWatch ermöglicht die Überwachung der Leistung von Diensten, auch bei Diensten, die über verschiedene Regionen verteilt sind. CloudWatch sammelt operative Daten, welche zur Verlaufsanalyse und für die Entscheidungsfindung in Bezug auf Optimierung und Fehlerbehebung hilfreich sind. 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 zur Überwachung von EC2-Instanzen in CloudWatch ist die *CPU-Auslastung*, *CPU-Utilization* auf Englisch. 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.[70] [71]

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

Metriken

Eine Metrik stellt eine Reihe von Daten über die Leistung eines Dienstes 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 einer Sekunde zulassen.

[68] Eine Ausgabe im Rechnungswesen liegt beim Abfluss von Zahlungsmitteln und/oder beim Eingehen von Zahlungsverpflichtungen in Form von Geldverbindlichkeiten, z.B. bei der Zahlung von Dienstleistungen, vor. (Vgl. Prof. Dr. Dr. h.c. Jürgen Weber,o.J., www.wirtschaftslexikon.gabler.de, o.S. [42])

[69] Vgl. Anders Lisdorf, 2021, Cloud Computing Basics: a Non.-Technical Introduction. S.152.[4]

[70] Vgl. Mark Wilkins, 2021, AWS Certified Solutions Architect - Associate (SAA-C02), S.185.[1]

[71] Diese Art von Aktionen werden im Kapitel 5 tiefer behandelt.

Ereignisse

Ein Ereignis ist in CloudWatch eine Änderung in einem AWS Dienst. AWS-Dienste können Ereignisse erzeugen, wenn sich ihr Status ändert. Beispielsweise wird ein Ereignis erzeugt, wenn *Amazon EC2 Auto-Scaling* Instanzen gestartet oder beendet werden oder wenn eine bestimmte Menge an Speicherplatz in einem *Bucket* erreicht wurde.[72] Ein Bucket ist ein Behälter, in dem Objekte bei Amazon S3 gespeichert werden.[73] Beispiele für Objekte sind Dateien wie Bilder und Videos.

Regel

Eine Regel ordnet eintreffende Ereignisse zu und leitet diese zur Verarbeitung an Ziele weiter. Eine einzelne Regel kann an mehrere Ziele weitergeleitet werden, damit diese alle Ereignisse parallel verarbeitet werden.[74]

Ziele

Ziele oder Targets sind AWS-Dienste, die aufgerufen werden, wenn eine Regel ausgelöst wird. *EC2 instances, AWS Lambda functions und Amazon SNS* sind unter anderem mögliche Ziele.[75] Die Ziele einer Regel müssen sich jedoch in derselben Region wie die Regel befinden. [76]

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 Gebühren von einem spezifischen AWS-Dienst ausgelöst werden. Benachrichtigungen können durch Amazon SNS oder zu einer E-Mail-Adresse geschickt werden.

Zu Testzwecken wurde ein Alarm erstellt, indem eine monatliche Ausgabengrenze von neun Euro für das AWS-Testkonto für diese Arbeit festgelegt wurde. [77]

[72] Vgl. AWS Cloud Watch Events: User Guide. S.1[14]
[73] Vgl. Amazon Simple Storage Service User Guide, S.4[20]
[74] Vgl. AWS Cloud Watch : User Guide. S.2[14]
[75] Vgl. Amazon SNS ist ein AWS-Dienst für die Benachrichtigung an Personen und an Applikationen.[32]
[76] Vgl. AWS Cloud Watch Events: User Guide. S.2[14]
[77] Eine Abbildung dieses Testalarms ist in Anhang 5.2.4 zu finden.

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.[78] Die in den Dashboards enthaltenen Informationen sind nicht nur für ihre Autoren von Relevanz. Weitere Personen innerhalb oder außerhalb einer Organisation können Zugriff auf Dashboards mit nützlichen Informationen bekommen, um Prozesse zu beschleunigen und Probleme schneller zu beheben. Um den Zugriff auf das Dashboard zu beschränken, ist es möglich, den Zugriff auf bestimmte Personen per E-Mail oder über SSO-Anmeldeinformationen zu beschränken.[79] *Single Sign-On (SSO)* ist ein Prozess der einmaligen Authentifizierung und Zugriff auf mehrere Ressourcen.[80] Außerdem hat die Einbindung von Dashboard-Informationen auf Infranet-Portale das Potenzial, Transparenz und eine schnelle Verbreitung von Informationen zu schaffen.[81]

Zu Testzwecken wurde ein Dashboard mit einigen *Widgets* erstellt.[82] Das erste Widget in der Abbildung 13 zeigt, die Anzahl der Aufrufe an der CloudWatch-API. Das zweite Widget zeigt die CPU-Auslastung und den eingehenden Netzwerkverkehr von einem Spot-Instanz. Diese Widgets verwenden Standardmetriken, deshalb verursachen sie keine Kosten.

Wie bereits erwähnt, ist es möglich den Zugriff von Dashboards freizugeben, ohne Zugang zu Ihrem eigenen AWS-Konto gewähren zu müssen. [83]

Fakturierungsalarme mit CloudWatch

AWS CloudWatch empfängt Abrechnungsmetriken von allen AWS-Diensten, auch von AWS-Rechnungen. Auf der Grundlage dieser Metriken ist es daher möglich, Regeln zu erstellen, die bei Überschreitung des geplanten Budgets, Alarme auslösen, wenn ein bestimmter Prozentsatz oder Betrag des festgelegten Budgets überschritten wurde. Die oben

[78] Vgl. AWS Cloud Watch : User Guide. S.28[15]

[79] Vgl. AWS Single Sign-On.[35]

[80] Vgl. Ziel von SSO ist es, die Anzahl von Login und Passwort in heterogenen Umgebungen zu reduzieren. Securing User Authentication using Single SignOn in Cloud Computing[75].

[81] Vgl. Business Knowledge Management: Wertschöpfung durch Wissensportale[3].

[82] Vgl. Ein Widget ist ein grafischer Weg, um Metriken in CloudWatch darzustellen. Unter anderem gibt es Widgets für Zahlen, Linien- und Balkendiagramme.

[83] Das hier erwähnte Dashboard wurde für den öffentlichen Zugriff temporär freigegeben. Über den Folgenden Link kann man auf das Dashboard zugreifen: t.ly/fNbyT

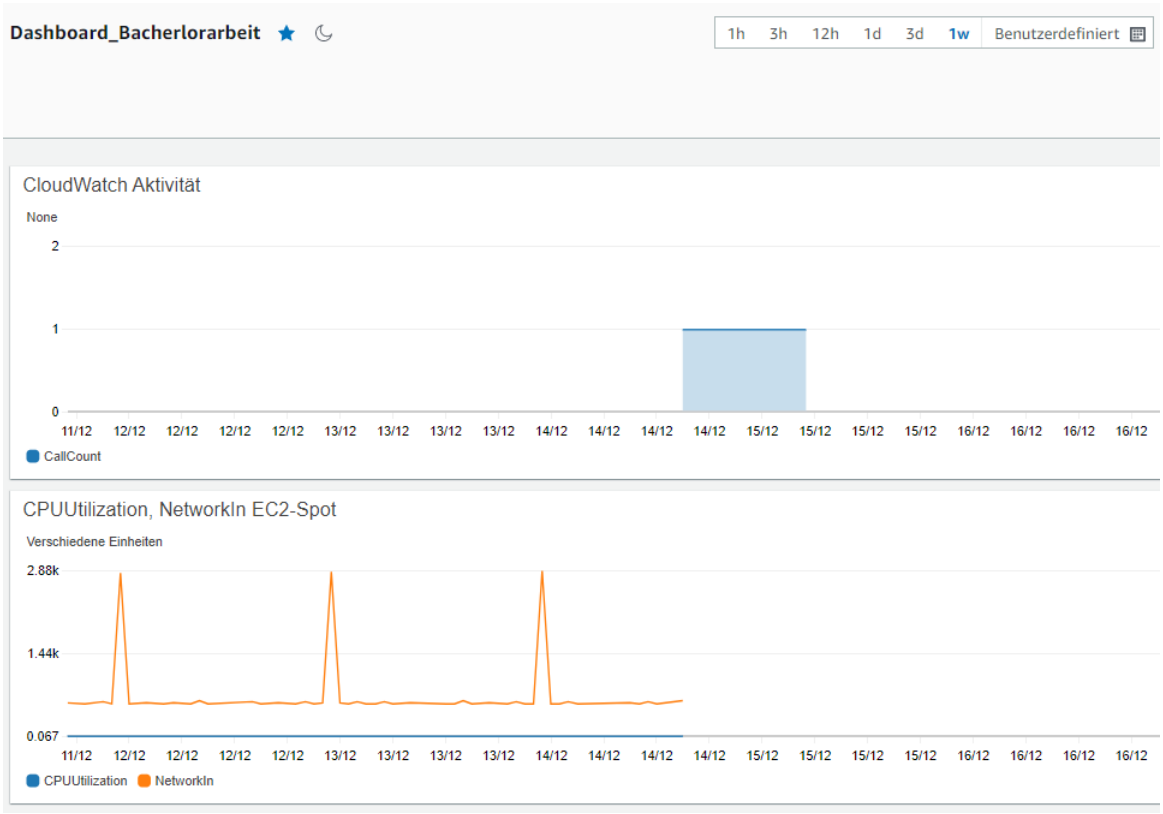


Abbildung 9
Dashboard-Test in CloudWatch.
Quelle: eigene Darstellung.

genannten Alarme finden ihre Anwendung unter anderem im Kostenverlaufsplan. Der Kostenverlaufsplan gehört zum Projektmanagement, welcher Kosten eines Projekts phasenweise oder kumuliert bereitstellt.[84] Im Anhang 5.2.4 befindet sich die Vorlage für die Erstellung eines Fakturierungsalarms in *JSON*- und *YAML*-Format.

Die Abbildung 10 zeigt ein in Phasen aufgeteiltes Projekt. In einem Fall wie diesem wäre es von Vorteil für jeden Prozentsatz oder jede Überschreitung der Budgets, einen spezifischen Alarm zu definieren.

[84] Vgl. Theo Peters und Nicole Schelter, 2021, Kompakte Einführung in das Projektmanagement. S.96[6].

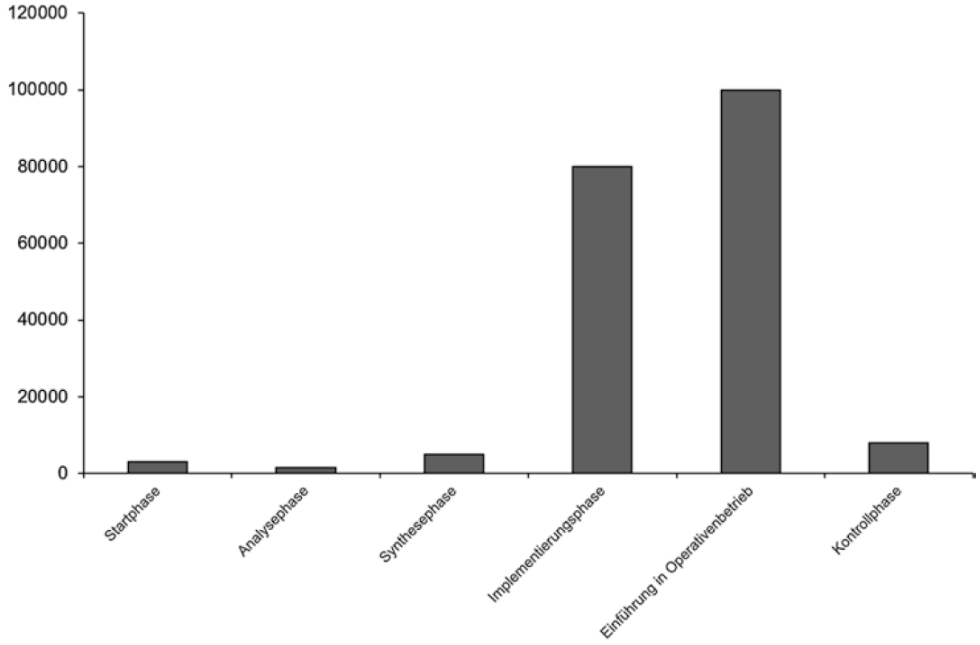


Abbildung 10
Kosten nach Projektphasen.
Quelle: Theo Peters und Nicole Schelter, 2021,
Kompakte Einführung in das Projektmanagement. S.97[6].

4.2 AWS Cost-Explorer

Der *Cost-Explorer* erstellt Berichte über die Kosten und die Nutzung von AWS-Diensten. Darüber hinaus wird eine Kostenprognose für die nächsten Monate erstellt, welche auf die Kosten der vergangenen Monaten basiert. Die Nutzung des Cost-Explorers ist kostenlos, nur API-Aufrufe sind kostenpflichtig. [85]

Standardberichte

Standardberichte sind vorgefertigte Berichte, die die Nutzung oder die Kosten nach einem definierten Zeitraum anzeigen. Diese zeigen eine grafische Darstellung der stündlichen, täglichen oder monatlichen Kosten nach Dienst. Ebenso wird die Abdeckung und die Auslastung von reservierten Instanzen oder die in Saving Plans Zahlungsmodell.[86] Zuletzt auch die Ausgaben auf dem AWS Marketplace.[87]

[85] Vgl. AWS Cost Management Pricing[24].
[86] Die Standardberichte über die Abdeckung und Auslastung der reservierten Instanzen wurde im TrueCar-Anwendungsfall verwendet. Dies befindet sich in Unterkapitel 3.5.
[87] Vgl. AWS Marketplace ist ein Einkaufskatalog für Software von Drittanbietern[36].

Anwendungsbeispiel für Standardberichte

Eine weitere Verwendung dieser Informationen findet sich im Bereich des Marketings. Als Beispiel soll nun ein Unternehmen gelten, dass ein Freemium-Dienst.[88] anbietet. Dessen Marketingabteilung möchte nun eine Werbekampagne durchführen Durch eine Werbekampagne werden in der Regel neue Nutzer generiert und zwar sowohl zahlende als auch nicht zahlende Nutzer. Normalerweise gilt: Je mehr Nutzer, desto größer die Belastung für die IT-Infrastruktur. Um die im Zusammenhang mit der Werbekampagne durch neue Nutzer entstehenden Kosten zu messen, werden die tatsächlichen Kundenakquisitionskosten (CAC) berechnet, wobei nur die Kosten der nicht zahlenden Nutzer berücksichtigt werden.[89] Zur Unterscheidung zwischen alten(vor der Werbekampagne) und neuen Nutzern wird das Erstellungsdatum des Nutzerkontos verwendet. Kunden, die aufgrund der Werbekampagne von der kostenlosen zur kostenpflichtigen Version des Dienstes gewechselt haben, werden in einer anderen Kategorie ausgeschlossen.[90]

Die Formel für die Berechnung der Kundenakquisitionskosten lautet wie folgt:
Anfallende Marketingkosten (MK) addiert mit den Vertriebskosten (VK) durch die Anzahl der gewonnenen Kunden (GK). Kosten von Nutzern, die den Dienst kostenlos in Anspruch nehmen, würden in diesem Fall in den Vertriebskosten enthalten sein. Auf diese Weise ist die Marketingabteilung in der Lage, die tatsächlichen Kosten pro zahlenden Neukunden zu berechnen, die durch die Werbekampagne generiert wurden.

Leistungskennzahlen (KPI)

Cost-Explorer-Berichte enthalten Daten, welche die Merkmale guter Leistungskennzahlen (Key Performance Indicators; KPI) erfüllen.[91] Sie sind aktuell, spezifisch und in Bezug auf die Zeit messbar.

In der Abbildung 11 werden die durchschnittlichen Kosten für EC2-Instanzen pro Stunde, der Prozentsatz der Instanzen einer bestimmten Generation und der Vorgängerversionen, die Abdeckung nach Zahlungsmodell und die Verteilung des S3-Speichers nach

[88] Vgl. o.V.o.J. Ein Freemium-Dienst bietet in der Regel zwei Versionen an, eine kostenlose und eine kostenpflichtige[54].
[89] Vgl. Kundenakquisitionskosten sind alle anfallenden Kosten in der Customer Acquisition-Phase für ein Unternehmen[53].
[90] Vgl. Cost-per-Action (CPA)[55].
[91] Vgl. Marc Optiz, 2019, Anforderungen an Kennzahlen. Prozessorientiertes Reporting. (S.130-132).[7]

Speicherklassen berechnet. In diesem Dashboard werden Metriken aus CloudWatch und Cost-Explorer-Berichten zusammengestellt.

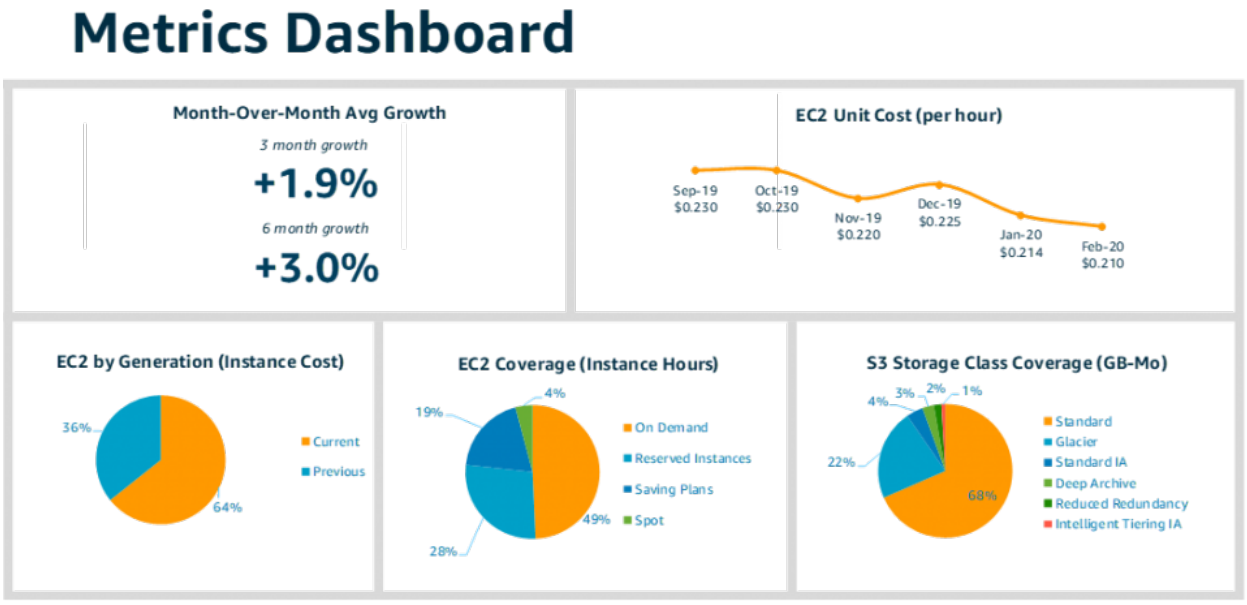


Abbildung 11
Dashboard mit Kennzahlen über EC2-Instanzen und S3-Speichereinheiten. [92]

Wie in der Abbildung 12 zu sehen, ist es möglich, mathematische Operationen mit den Metriken durchzuführen, um auf dem Dashboard nur die aussagekräftigsten Kennzahlen anzuzeigen.

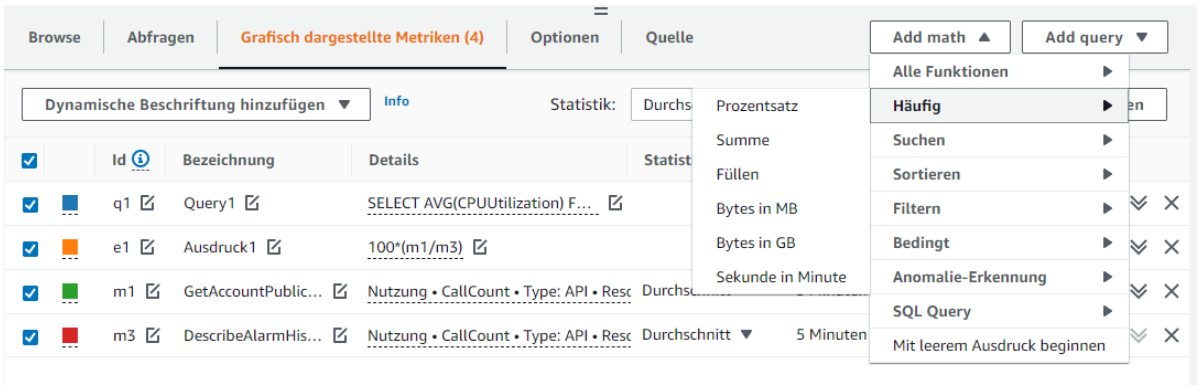


Abbildung 12
Mathematische Operationen an Cloud-Diensten in CloudWatch.
Quelle: CloudWatch AWS-Console

Budgetplanung

Die Budgetplanung ist eine Methode der Kostenkontrolle, die beim Start eines neuen Projekts eingesetzt wird.[93] Der Cost-Explorer liefert Berichte über die in den letzten zwölf Monaten entstandenen Kosten, inklusive der Kostenprognose der kommenden zwölf Monaten, welche zu einem guten Budgetplanung beitragen. Durch die Möglichkeit, die in den letzten Monaten angefallenen Kosten nach bestimmten AWS-Diensten, Projekt oder Abteilung zu trennen, 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"[44]. In dem Fall dieser Arbeit sind die obengenannten Ressourcen die AWS-Dienste.

Somit liefert der Cost-Explorer einerseits Informationen zur Rechtfertigung von Ausgaben aus im Voraus festgelegten Budgets, andererseits hilft bei der Planung künftiger Budgets und unterstützt gleichermaßen die Verfolgung von KPIs.

4.3 AWS Trusted Advisor

Der *AWS Trusted Advisor* ist ein Werkzeug, welcher Empfehlungen zur Kostenreduzierung, Verbesserung der Systemverfügbarkeit und Erhöhung der Systemsicherheit angibt.[94]Die Empfehlungen basieren auf Best-Practices, die im Laufe der Jahre durch die Betreuung von AWS-Kunden gesammelt wurden und Prüfungen, die auf dem bestehenden AWS-Konto durchgeführt wurden. In dieser Arbeit werden Empfehlungen in Bezug auf Servicekontingente und Kostenoptimierung insbesondere betrachtet, weil es sich um Empfehlungen handelt, die mit Kostenüberwachung und -optimierung zusammenhängen. Der Status von Trusted-Advisor-Prüfungen sind über CloudWatch Events zugänglich.

[93] Vgl. Indeed Editorial Team: Planning the budget properly. Cost Control Methods: Definitions and Examples, 2021, o.S. [45].

[94] Es ist zu berücksichtigen, dass nur limitierte Sicherheitsprüfungen (Standart sind sechs Prüfungen; Stand November 2021) im Plan Developer und Basic Support kostenfrei sind. Prüfungen für die Kategorie Servicekontingente sind komplett kostenlos. Detaillierte Informationen und Empfehlungen von der Kategorien Kostenoptimierung, Performance und Fehlertoleranz sind nur zugänglich, wenn ein Business- oder Enterprise-Konto vorliegt. Vgl. AWS, o.J. Trusted Advisor, o.S.[22]

Abbildung 13
AWS Trusted Advisor Kategorien[22]

Die Abbildung 13 zeigt die fünf Kategorien von Trusted Advisor mit jeweils drei verschiedenen Indikatoren. Diese zeigen an, welche Prüfungen durchgeführt wurden. Grün bedeutet, dass keine Fehler oder zu prüfende Empfehlungen vorhanden sind. Warnungen werden durch orangefarbene Indikatoren und Fehler durch rote Indikatoren angezeigt.

Diese Empfehlungen scheinen ein angemessener Startpunkt für die Untersuchung von AWS-Diensten zu sein. Eine genauere Untersuchung erfolgt mithilfe anderer Werkzeuge, wie CloudWatch oder Cost-Explorer. [95] Diese Kategorien lassen sich daher in eingeschränkter Weise und unter den aktuellen Umständen untersuchen. Es bestehen Empfehlungen für verschiedene AWS-Dienste, unter anderem für EBS, Route 53, RDS und AWS Lambda. In diesem Sinne werden im Folgenden Empfehlungen zu EC2-Instanzen gegeben, da dies der Fokus dieser Arbeit entspricht.[96]

Empfehlungen zur Kostenoptimierung

Sollten EC2-Instanzen mit geringer Auslastung gefunden werden, wird dies bei Trusted Advisor signalisiert. Denn diese Instanzen verursachen Kosten, welche durch die Terminierung oder das Pausieren vermieden werden können. Eine geringe Auslastung wird von AWS definiert, wenn Instanzen in den letzten 14 Tagen eine CPU-Auslastung von 10% oder weniger hatten und wenn der Netzwerkverkehr in den letzten vier Tagen den Wert von 5 MB nicht überschritten hat.

Die Buchungen von *reservierten Instanzen*, die in den letzten 30 Tage abgelaufen sind

[95] Die Empfehlungen für die Kategorien Kostenoptimierung und Servicekontingente werden in der AWS-Dokumentation nur kurz beschrieben und sind in einem Basiskonto nicht zugänglich. (Vgl. AWS Support - Benutzerhandbuch. S.59-65 und S.83-94 [38])

[96] Empfehlungen für Amazon S3-Speichereinheiten sind im Trusted Advisor nicht verfügbar.

oder in den kommenden 30 Tage ablaufen werden, werden hervorgehoben. Auf diese Weise wird vermieden, dass die Buchung von Instanzen vergessen wird oder dass sie erneuert werden müssen, wenn sie bereits abgelaufen sind.

Die Empfehlungen des Cost-Explorers zu Saving Plans werden auch im Trusted Advisor angezeigt. Die Saving Plans sind eine mögliche Sparalternative zu reservierten Instanzen. AWS weist darauf hin, dass nur eine der beiden Maßnahmen zur Instanzreservierung durchgeführt werden sollte.

Trusted Advisor erstellt Simulationen möglicher Kombinationen von reservierten Instanzen und On-Demand-Instanzen. Dies dient dazu, die Auswahl reservierter Instanzen auf der Grundlage von AWS-Simulationen zu erleichtern.

Empfehlungen zur Servicekontingente

In der Kategorie Servicekontingente (auch als Kontingente bekannt) werden Empfehlungen zur Vermeidung von Grenzwertüberschreitungen hervorgehoben. Sich dieser Grenzen bewusst zu sein, sollte die Möglichkeit, rechtzeitig zu handeln und es trägt zu Kostenkontrolle über die AWS-Cloud-Dienste bei.

Für Auto-Scaling-Gruppen wird geprüft, ob deren Nutzung mehr als 80% des Kontingents beträgt. Aufgrund fehlender Informationen in der AWS-Dokumentation wird angenommen, dass eine Auto-Scaling-Gruppe als eine einzelne Recheneinheit betrachtet und eine Auslastung von mehr als 80% als Näherung an die Grenze der Rechenkapazität angesehen wird. Dies wird eine Anpassung der Startkonfiguration für eine bessere Skalierung zur Folge haben.

Die Prüfungen, die die Nutzung eines Kontingents über 80% betragen, werden auch für *On-Demand-Instanzen, reservierte Instanzen, EC2-Classic Elastic IP Addresses und EC2-VPC Elastic IP Addresses* angezeigt.

Trusted Advisor Kostenerwägungen

Es muss berücksichtigt werden, ob es kosteneffizient für Support-Pläne zu zahlen wird, da diese den Zugang zu allen Empfehlungen des Trusted Advisors ermöglichen. Eines der Ziele dieser Arbeit ist es, die Entstehung der Kosten auf eine praktikable Weise zu verstehen (Kostenüberwachung). In diesem Sinne wurde bisher nicht festgestellt, dass die Empfeh-

lungen von Trusted Advisor zu echten Einsparungen führen. So wäre es nicht sinnvoll, Kosten für Pläne wie Business- oder Enterprise Support zu übernehmen, wenn diese die möglichen Einsparungen übersteigen. Die Vorteile von Business- oder Enterprise Support-Plänen beschränken sich nicht auf Kosteneinsparungen und Kostenbegrenzung, sondern tragen auch zur Sicherheit und Leistung bei.[97] Dabei stellt sich die Frage, ob die Empfehlungen aller fünf Kategorien für die aktuelle Situation des Unternehmens benötigt werden.

Die Preise für einen *Business Support*-Plan sind wie folgt definiert:
A) Zwischen 0 USD und 10.000,00 USD: 10% oder 100 USD.
Je nachdem, was größer ist.
B) Zwischen 10.000,00 USD und 80.000,00 USD: 7%.
C) Zwischen 80.000,00 USD und 250.000,00 USD: 5%.
D) Ab 250.000,00 USD: 3%.[98]

Die Preise für einen *Enterprise Support*-Plan sind wie folgt definiert:
A) Zwischen 0 USD und 150.000,00 USD: 10% oder 15.000,00 USD. Je nachdem, was größer ist.
B) Zwischen 150.000,00 USD und 500.000,00 USD: 7%.
C) Zwischen 500.000,00 USD und 1.000.000,00 USD: 5%.
D) Ab 1.000.000,00 USD: 3%.[99]
Die Prozentsätze basieren auf der monatlichen Gebühr für AWS-Dienste.

Ein Unternehmen, das zum Beispiel 85.000 USD pro Monat für Cloud-Dienste zahlt, würde wie folgt für Business Support-Plan bezahlen:
10.000 USD x 10% = 1.000 USD
+ 70.000 USD x 7% = 4.900 USD
+ 5 000 USD x 5% = 250 USD

Gesamtsumme = 6.150,00 USD

Am Ende circa 7,2% von der Gesamtkosten über Cloud-Dienste.

[97] Darüber hinaus bieten beide Pläne rund um die Uhr technischen Support durch AWS-Ingenieure und andere zusätzliche Dienstleistungen, auf die in dieser Arbeit nicht weiter eingegangen wird. Die Preise der Support-Pläne geben einen Hinweis darauf, ob die zu zahlende Empfehlungen von Trusted Advisor zur Kostenoptimierung und -überwachung kosteneffizient würden.
[98] Vgl. AWS Support Plan Pricing - Business Support-Plan, 2021, o.S. [39]
[99] Vgl. AWS Support Plan Pricing - Business Enterprise-Plan, 2021, o.S. [39]

Fazit

In diesem Kapitel wurde nachgewiesen, dass die Anwendung von CloudWatch ermöglicht Alarme auf Basis von Ereignissen einzurichten, die mit Amazon SNS oder externen E-Mail-Adressen kommunizieren. Zudem wurde - aus einem Blickwinkel des Kostenmanagements gezeigt, dass mit Cost-Explorer eine Kostenanalyse der letzten 12 Monate, eine Kosteneinschätzung im aktuellen Monat und eine Kostenprognose für die nächsten Monate möglich ist. - Diese Informationen dient unter anderem zur Erstellung einer operativen Budgetplanung mit genaueren Daten, da Kosten nach Tags und anderen Filtern getrennt werden können. Darüber hinaus wurde Trusted Advisor in eingeschränkter Weise vorgestellt, welcher konkrete Optimierungsempfehlungen gibt und warnt über Leistungsgrenzen. Dies kann mit erheblichen Kosten verbunden sein und ist daher nicht für alle Unternehmen unmittelbar attraktiv. Nun, es ist offensichtlich, dass nicht jedes Unternehmen die Kosten von Trusted Advisor zahlen kann. Nichtsdestotrotz spricht nichts dagegen, die kostenlosen Empfehlungen zur Kostenoptimierung und Servicekontingente zu berücksichtigen, da diese Hinweise für eine kosteneffiziente IT-Infrastruktur erzielen.

5 Optimierungsmaßnahmen

In diesem Kapitel werden die Optimierungsmaßnahmen für EC2-Instanzen und Amazon S3 vorgestellt. Im Abschluss, mithilfe von CloudWatch und Cost-Explorer, besteht die Möglichkeit, die Effekte der Optimierungsmaßnahmen zu messen. Somit kann ein Vergleich zwischen den Kosten vor und nach der Optimierungsmaßnahmen durchgeführt werden.

5.1 EC2 Auto Scaling

Das *Auto Scaling* oder die automatische Skalierung von Instanzen dient dazu die richtige Anzahl von EC2-Instanzen zur Verfügung zu haben, um die Anwendungslast dynamisch abzudecken.[100] Diese Fähigkeit wird als *horizontale Skalierung* bezeichnet.[101]

Die Abbildung 14 zeigt das wechselnde Verhalten einer Beispielanwendung, die vor allem unter der Woche genutzt wird. Am Wochenende sinkt die Nachfrage nach Rechnerkapazität auf weniger als 25% und lässt den Rest der Kapazitäten ungenutzt.

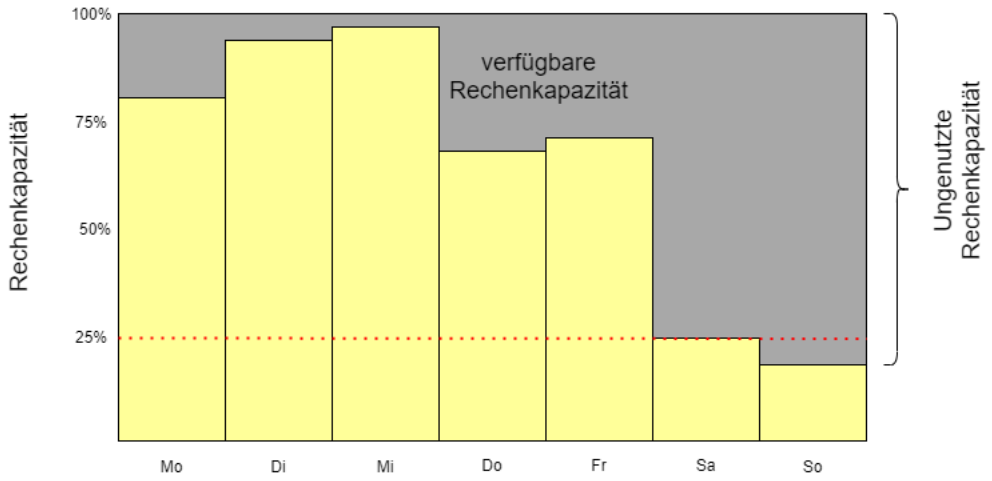


Abbildung 14
Ungenutzte Rechenkapazität ohne automatische Skalierung.
Quelle: Eigene Darstellung mit fiktiven Angaben.

Die gelben Säulen stellen die tägliche genutzte Rechenkapazität dar. Die graue Zone ent-

[100] Vgl. Was ist Amazon EC2 Auto Scaling? S.9[33]
[101] Vgl. Die Grundbedeutung der horizontalen Skalierung ist, dass Systeme durch zusätzliche Komponenten erweitert werden. Im Gegensatz dazu bedeutet der Begriff "vertikale Skalierung", dass einer einzelnen Komponente zusätzliche Leistungsfähigkeiten und Ressourcen hinzugefügt werden. o.S.[63]

spricht ungenutzte Rechenkapazität und beträgt etwa ein Drittel der wöchentlichen Rechnerkapazität.

Auto Scaling Group

Die Instanzen, die zur Deckung der erforderlichen Rechenkapazität zur Verfügung stehen, werden in einer *Auto-Scaling-Gruppe (Auto Scaling Group)* zusammengefasst. Diese Gruppe von Instanzen wird in AWS als Auto-Scaling-Gruppe bezeichnet. Bei der Erstellung einer Auto-Scaling-Gruppe wird somit eine minimale, gewünschte und maximale Anzahl von Instanzen definiert.

Die Abbildung 15 zeigt die gewünschte Instanzen einer Auto-Scaling-Gruppe, welche beim Start der Auto-Scaling-Gruppe gestartet werden. Die minimale und maximale Anzahl von Instanzen sind die Grenzwerte für die Auto-Scaling-Gruppe.

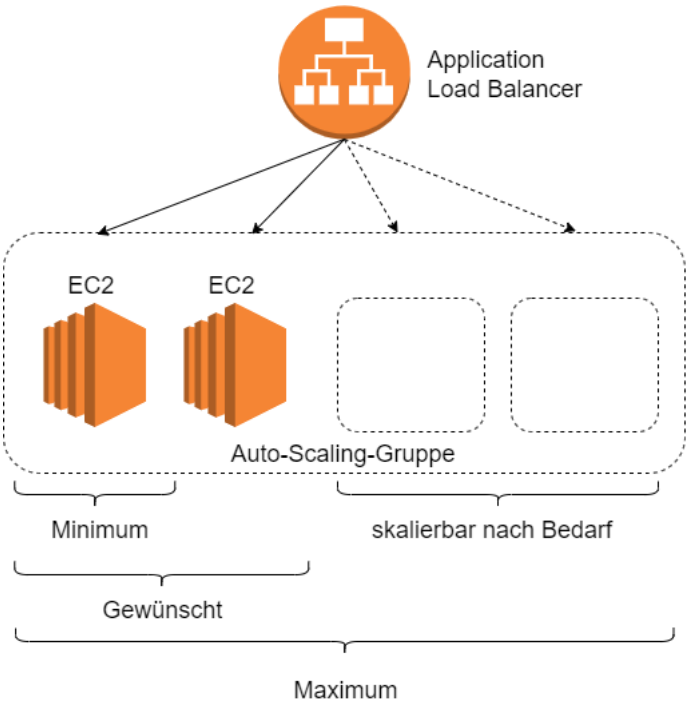


Abbildung 15
Auto-Scaling-Gruppe nach den Anzahl der Instanzen und
die Umleitung der Datenverkehr durch dem Application Load Balancer.
Quelle: Eigene Darstellung basiert auf Amazon
EC2 Auto Scaling - Benutzerhandbuch. S.9[33].

Elastic Load Balancing

Ein *Elastic Load Balancer* ist für die Verwaltung eingehender Anfragen zuständig, indem es den Datenverkehr auf alle laufenden EC2-Instanzen umleitet.[102] Dies sorgt dafür, dass die Instanzen mit einem ausgeglichener CPU-Auslastung arbeiten. Im diesem Sinne zeigt die Abbildung 15 einen Application-Load-Balancer, welcher den Datenverkehr auf die Instanzen einer Auto-Scaling-Gruppe verteilt.

5.1.1 Zeitgesteuerte Skalierung

In einem On-Premise-System würde es, wenn überhaupt, nur einen geringen Kostenunterschied ausmachen, wenn Instanzen die ganze Zeit aktiv bleiben.[103] Im Gegensatz dazu, ist es bei On-Demand-Zahlungsmodelle sinnvoll Zeiträume zu definieren, in denen Instanzen abgeschaltet werden können, um deren Nutzung zu reduzieren. Bei Systemen, die nur tagsüber und unter der Woche in Betrieb sein müssen, kann sogar dies zu einer Einsparung von bis zu 67% der Kosten für Instanzen führen. Im Falle, das zum Beispiel Test- und Beta-Umgebungen von Montag bis Freitag von 7 bis 20 Uhr laufen würden.

Die Abbildung 16 zeigt die Kostenberechnung einer nicht produktiven Umgebung (z.B. Test, Dev oder Beta) mit On-Demand-Instanzen. Diese Umgebung wird nur von Montag bis Freitag von 7:00 bis 20:00 Uhr genutzt. In der rechten Spalte werden die Kosten für Instanzen berechnet, wenn sie immer aktiv bleiben. In der linken Spalte wurde eine Berechnung durchgeführt, bei der die Instanzen nur dann eingeschaltet werden, wenn sie nach einem Zeitplan gesteuert würden. Darüber hinaus veranschaulicht die Abbildung 16 am Ende den Prozentsatz und den Betrag (in Euros) der möglichen Einsparungen, wenn die Instanzen nach einem determinierten Zeitplan gesteuert wird.

[102] Vgl. In diesem Fall beschränkt auf den Application Load Balancer. Amazon Elastic Container Service Entwicklerhandbuch - Load Balancer-Typen - S.617[40]
[103] Anders Lisdorf, 2021, S. 153[4]

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 16
Berechnung für ein nicht-produktive Umgebung mit zeitgesteuerter Skalierung.
Quelle: Eigene Darstellung.

Quelle des Stundensatzes: AWS Pricing Calculator.[104]

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

5.1.2 Dynamisches Auto Scaling

Es kann jedoch zu schnellen und kontinuierlichen Änderungen im Verhalten von Applikationen kommen, häufig innerhalb von wenige Minuten. Bei solchen Szenarien ist es daher 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 17 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 passend, wenn das Ziel darin besteht, ungenutzte Cloud-Dienste abzuschalten. Als Konsequenz der Abschaltung von ungenutzten Cloud-Diensten folgt die Reduzierung von Kosten.

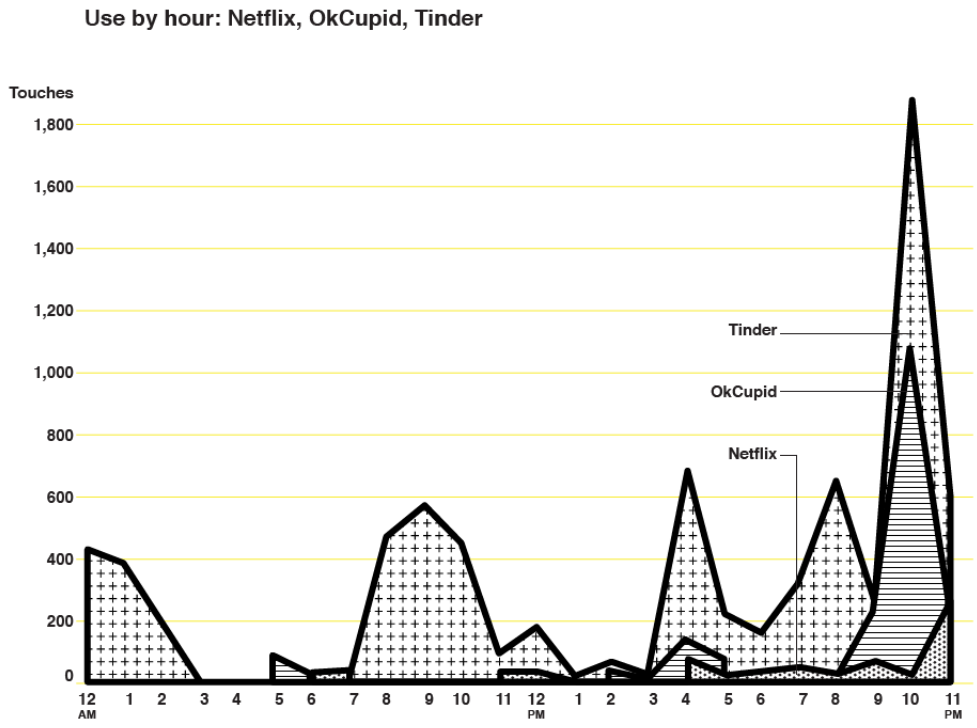


Abbildung 17
DScout's Study: "Putting a Finger on Our Phone Obsession".
Nutzung pro Stunde von Netflix, OkCupid und Tinder während des Tages[68].
Mit Touches sind die Anzahl der Klicks, Swipes oder einfachen Interaktionen mit der Applikation gemeint.

Eine der Metriken, die von AWS benutzt wird, ist die gesamte CPU-Auslastung (CPU-Utilization).[105] 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

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, eine Anpassung in der Rechenkapazität benötigt wird, besteht die Möglichkeit die Rechenkapazität manuell zu steuern. Dies geschieht, ohne dass die aktiven Instanzen unterbrochen werden.

5.1.4 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. Anders als zeitgesteuerte Skalierung ist es nicht notwendig, die Verhaltensmuster der Anwendungen zu analysieren.

5.2 S3 Optimierung

In diesem Unterkapitel werden Maßnahmen zur Speicheroptimierung für Amazon S3 beschrieben. Jedem Objekt in Amazon S3 ist eine Speicherklasse zugewiesen. Die Speicherklassen werden nach der Zugriffshäufigkeit auf die Objekte unterschieden und sind für verschiedene Szenarien konzipiert. Es gibt Speicherklassen für den häufigen und den seltenen Zugriff.[106] Der Preis bei Amazon S3 wird pro GB berechnet und ist umso niedriger, je geringer der Zugriff auf die Objekte ist.[107] Um die Speicherkosten zu optimieren, ist es daher notwendig, die richtige Speicherklassen für die jeweilige Applikation zu wählen, weil die Speicherkosten durch ihre Klasse berechnet werden.

[105] Die für die automatische Skalierung erforderlichen Metriken wurden bereits ausführlicher im Unterkapitel 4.1 erläutert.
[106] Vgl. AWS: Amazon Simple Storage Service - User Guide. S.709.[20]
[107] Vgl. AWS S3 Pricing [10]

5.2.1 Auswahl der passenden Speicherklasse

Um die richtige Wahl zu treffen, müssen die Anforderungen der Applikation verstanden werden. Ärztliche Patientenakten und *Instagram-Stories* sind zwei Beispiele für Daten, die nach deren Erstellung für einen Mindestzeitraum oder auf unbestimmte Zeit aufbewahrt werden.[108] In Deutschland müssen ärztliche Patientenakten mindesten zehn Jahre aufbewahrt werden.[109] *Instagram* verwendet die von seinen Nutzern bereitgestellten Informationen, einschließlich der Metadaten von Bildern, um andere *Instagram-* und *Facebook*-Produkte zu empfehlen.[110] Die Zugriffshäufigkeit und die Aufbewahrungszeit bilden somit die zwei Hauptkriterien für die Verschiebung von Daten zwischen Speicherklassen.[111]

Die Objekte werden in Behältern gespeichert, die bei AWS Buckets genannt werden. Daten werden über einen längeren Zeitraum gespeichert aufgrund der vorgeschriebenen Anforderungen oder weil per Gesetz auf die Informationen zukünftig zugegriffen werden muss. Zusätzlich, wenn auf die Daten kaum zugegriffen wird, sind *Glacier* und *Glacier Deep Archive* passende Speicherklassen. Die richtige Speicherklasse zuwählen, stellt jedoch keine einfache Entscheidung dar. Hinzu kommt, dass nicht alle Daten in einer Applikation immer die gleichen Zugriffsmuster haben. In solchen Fällen, besteht die Möglichkeit, Regeln zu definieren, die Dateien zwischen verschiedenen Speicherklassen in Abhängigkeit ihres Alters zu verschieben.

[108] Vgl. Bei Instagram Stories handelt es sich um einen i.d.R. kurzen visuellen Content in der Regel Bilder oder kurze Videos, die nach 24 Stunden automatisch aus der Applikation Instagram (Stand November 2021).[50]

[109] Vgl. Nach dem Bürgerlichen Gesetzbuch (BGB) § 630f müssen Patientenakten zehn Jahren nach Abschluss der Behandlung aufbewahrt werden, soweit nicht nach anderen Vorschriften andere Aufbewahrungsfristen bestehen. [43]

[110] Vgl. Instagram macht keine genauen Angaben darüber, wie lange die Nutzerdaten aufbewahrt werden, sondern gibt nur an, dass sie so lange wie nötig aufbewahrt werden. Hilfebereich Instagram: VII. Datenspeicherung, Deaktivierung und Löschung von Konten[51].

[111] Vgl. AWS: Amazon Simple Storage Service - User Guide. S.711.[20]

5.2.2 Lebenszyklus-Konfiguration

Die *Lebenszyklus-Konfiguration* oder *lifecycle policy* ist eine Maßnahme zur Optimierung für Amazon S3-Speichereinheiten. Eine S3-Lebenszykluskonfiguration beschreibt in einer XML-Datei Regeln und Aktionen für die Verschiebung von Objekten in unterschiedliche Speicherklassen. Allerdings verursacht diese Verschiebung Kosten. Ein Beispiel von diesen Kosten und mögliche Einsparungen werden in Abbildung 18 vorgestellt.

Um konkretere Regeln zu definieren, ist es möglich Tags zu verwenden und somit eine Unterscheidung zwischen Objekten mit verschiedenen Tags zu gewährleisten zu können. Es ist wie in dem folgenden Codebeispiel möglich, alle Objekte mit dem Tag-Wert: *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.
```

112

5.2.3 Anwendungsbeispiel für eine Lebenszyklus-Konfiguration

Zur Veranschaulichung der Verschiebung von Objekten zwischen Speicherklassen wird der folgende Anwendungsfall vorgestellt[113].

Ein Sicherheitsunternehmen muss Sicherheitsvideos speichern, die aktuell im Summe 120 TB groß sind. Viele von ihnen werden mindestens 5 Jahre lang aufbewahrt, falls sie vor Gericht als Beweismittel dienen müssen. 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. Jedes Video 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 zweites Szenario betrachtet, in dem die Videos je nach Alter verschoben werden. Im zweiten Szenario müssen die Kosten für die Verschiebung zwischen Speicherklassen berücksichtigt werden. Die Verschiebung erfolgt durch eine Lebenszyklus-Konfiguration, wie im Unterkapitel 5.2.2 beschrieben. Zum besseren Verständnis wird angenommen, dass 20% der Dateien in S3 Standard Infrequent Access und 30% in S3 Glacier gespeichert werden.

[112] Vgl. AWS: Amazon Simple Storage Service - User Guide. S.701.[20]

[113] In diesem Fall wird der Punkt als Dezimaltrennzeichen und das Komma als Tausendertrennzeichen verwendet.

Durchschnittliche Dateigröße	3.4	GB
Anzahl der Dateien	36,141	Überwachungsvideos
Gesamtspeicher	122,880	GB
	120	TB

Ausschließlich S3-Standard verwenden		
	S3 Standard (erste 51200GB)	**S3 Standard (Nächste 450 TB)**
Speicherplatz in GB	51,200	71,680
Preis pro GB	$0.0245	$0.0235
Speicherverteilung	42%	58%
Anzahl der Dateien	15,059	21,082
Übertragungsgebühr (pro 1.000 Aufrufe)	-	-
Kosten für Verschiebung	0	0
Speicherkosten	$1,254.40	$1,684.48
Monatliche 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	51,200	10,240	24,576	36,864
Preis pro GB	$0.0245	$0.0235	$0.0136	$0.0045
Speicherverteilung	42%	8%	20%	30%
Anzahl der Dateien	15,059	3,012	7,228	10,842
Übertragungsgebühr ($0.01/1,000 Aufrufe)	-	-	$0.0100	$0.0360
Kosten für Verschiebung	0	0	$0.72	$3.90
Speicherkosten	$1,254.40	$240.64	$334.23	$165.89
Monatliche Gesamtkosten	$1,999.79			

Abbildung 18
Kostenvergleich durch Nutzung von unterschiedlichen Speicherklassen.

Quelle: Eigene Darstellung mit Stundensätze der S3-Preise.[114]

Anhand der Berechnungen in der Abbildung 18 wird sichtbar, dass ein Einsparungspotenzial von rund 1,000 (Eintausend) USD pro Monat besteht.[115] Die notwendigen Regeln für die Verschiebung von den Videos erfolgt mit Hilfe einer Lebenszyklus-Konfiguration[116] und somit wird ein Teil der Videos in anderen kostengünstigeren Speicherklassen verschoben.

[114] Vgl. AWS S3 Pricing[10]
[115] Bei der Berechnung wurden die Kosten für das Verschieben von Videos zwischen Speicherklassen berücksichtigt.
[116] Eine ähnliche Lebenszyklus-Konfiguration, wie die von Unterkapitel 5.2.2

5.2.4 Intelligent-Tiering

Das *Intelligent-Tiering* verschiebt Objekte auf der Grundlage von Zugriffsmustern. Diese Speicherklasse ist ideal für Objekte mit wechselnden oder unbekannten Zugriffsmustern. Wie die Senior Product Managerin für Amazon S3 *Ruhi Dang* erklärt, hätten einige Unternehmen weder die Zeit noch das Geld, um eine Person einzustellen, die ihre Daten sortiert und in die richtige Speicherklasse einordnet[117]. Intelligent-Tiering ist eine attraktive Lösung für Unternehmen, die jährlich weniger als 100,000 USD für Speicher ausgeben. [118]

Die Abbildung 19 zeigt, wie Objekte in Abhängigkeit davon, ob auf sie zugegriffen wurde oder nicht, verschieben werden.

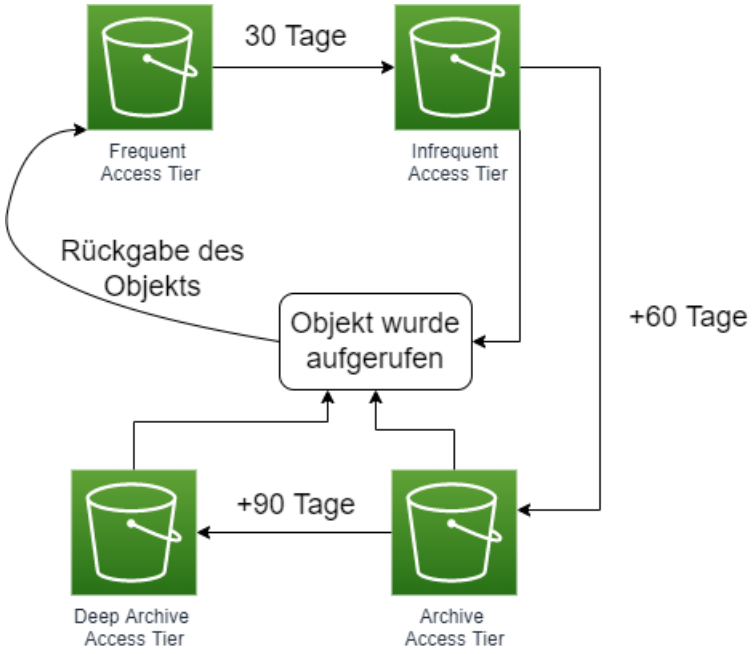


Abbildung 19
Funktionsweise von Intelligent-Tiering

Quelle: Eigene Darstellung auf der Grundlage von der Funktionsweise von Intelligent-Tiering.[119]

[117] Vgl. Ruhi Dang, 2019, AWS re:Invent 2019: Guidelines and design patterns for optimizing cost in Amazon S3. Minute: 20:05 [17]
[118] Vgl. Ruhi Dang, 2019, AWS re:Invent 2019: Guidelines and design patterns for optimizing cost in Amazon S3. Minute: 21:12 [17]
[119] Vgl. Amazon Simple Storage Service - User Guide. S.715[20]

Wird ein Objekt zu einem späteren Zeitpunkt aus der Ebene der seltenen Zugriffe aufgerufen, wird dieses automatisch in eine Speicherklasse der häufigen Zugriffe zurückversetzt.

Anwendungsbeispiel Intelligent-Tiering

Im Folgenden wird eine Berechnung mit Intelligent-Tiering für das Szenario über die Sicherheitsvideos des Unterkapitels 5.2.3 durchgeführt. In diesem Fall wurden ausschließlich die Ebenen *Frequent-Access, Infrequent-Access* und *Instant-Archive-Access* ausgewählt.[120]

Die Speicherzuweisung für die 120 TB in den Speicherebenen wurde wie folgt:
Frequent-Access-Tier: 50%
Infrequent-Access-Tier: 20%
Instant-Archive-Access: 30%

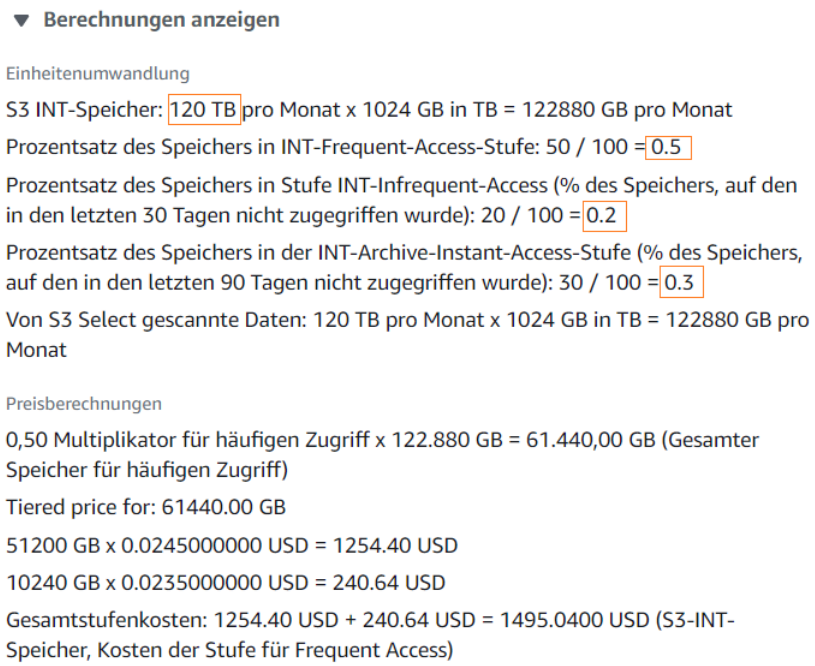


Abbildung 20
Berechnung für die Verwaltung von 120 TB mit AWS Pricing-Calculator für S3 Intelligent-Tiering f(1).
Quelle: eigene Darstellung von AWS Pricing-Calculator [19].

[120]Da bei der Berechnung des Unterkapitels 5.2.3 eine Speicherklasse für häufigen Zugriff, eine für seltenen Zugriff und eine für Archivierung verwendet wurde, wurden vergleichbare Speicherebenen für die Berechnung mit Intelligent-Tiering ausgewählt.

In der Abbildung 21 werden die Kosten gezeigt, die durch jede Speicherebene anfallen (Orange markiert).
Darüber hinaus werden die folgenden Kosten berechnet (Blau markiert):
- Überwachung und Automatisierung von 36.141 Objekten.
- Leseanfragen (GET-Anfragen) von 18.070 Objekten, welche 50% der Gesamtzahl der Videos entsprechen, wie in den Anforderungen im Unterkapitel 5.2.3 definiert.
- Scannen von Objekten, die allen Videos und 120 TB entsprechen.

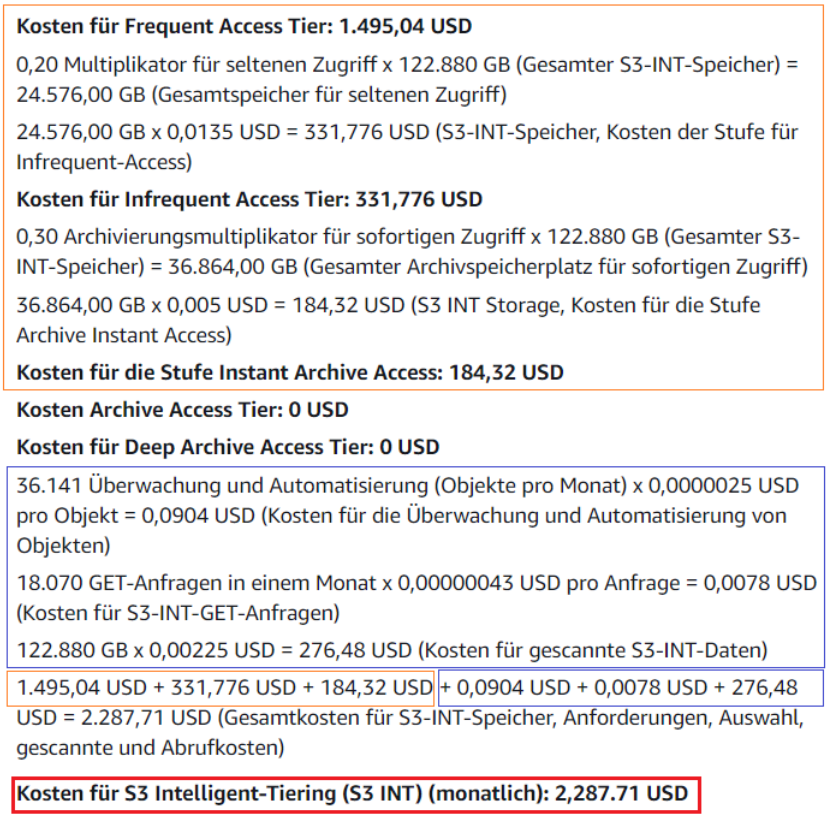


Abbildung 21
Berechnung für die Verwaltung von 120 TB mit
AWS Pricing-Calculator für S3 Intelligent-Tiering (2).
Quelle: eigene Darstellung mit AWS Pricing-Calculator [19].

Vergleich zwischen der Intelligent-Tiering und der Lebenszyklus-Konfiguration

Bezüglich dieser Thematik muss betont werden, dass bei Intelligent-Tiering die Objekte in Ebenen von seltenen Zugriff automatisch auf eine Ebene für häufigen Zugriff zurückgegeben werden. Siehe Abbildung 19.

Ein weiterer Unterschied liegt darin, dass bei einer Lebenszyklus-Konfiguration die Tage der Verschiebung zwischen den Speicherklassen und die Speicherklassen leicht verändert werden können. Diese ist bei Intelligent-Tiering standardmäßig vorab bereits festgelegt. Darüber hinaus besteht die Möglichkeit, Objekte aus Intelligent-Tiering in andere Speicherklassen zu verschieben, indem man eine Lebenszyklus-Konfiguration verwendet.[121] Dadurch wäre es möglich Objekte für einen längeren Zeitraum in einer bestimmten Speicherklasse aufzubewahren und Richlinien von Intelligent-Tiering zu überspringen.

[121] Vgl. AWS: Amazon Simple Storage Service - User Guide. S.724.[20]

Zusammenfassung

Die Untersuchungen, die vorgestellten Anwendungsfälle und die durchgeführten Berechnungen bestätigen die These dieser Arbeit, dass Unternehmen, welche ihre Nutzung von Cloud-Diensten überwachen, besser in der Lage sind, Optimierungsmaßnahmen zu ergreifen und damit ihre Kosten für Cloud-Dienste reduziert werden. In der folgenden Zusammenfassung wird dies konkret aufgelistet.

In Kapitel 2 wurde die Bedeutung von Cloud-basierten Systemen bestätigt und verstäkt, zum einen durch die Statistiken über die Nutzung von Cloud-Systemen weltweit. Zum anderen durch die Anzahl von Unternehmen, einschließlich in Deutschland, mit erfolgreicher Implementierung von Public Cloud-basierten Systemen. Es hat sich gezeigt, dass eine erhebliche Anzahl von Unternehmen derzeit Schwierigkeiten haben, auf die Cloud umzusteigen, weil ihrer Mitarbeitern technische Qualifikation fehlen.

In Kapitel 3 wurden die Zahlungsmodelle von EC2-Instanzen untersucht. Anhand einer Berechnung von Instanzen in dem Zahlungsmodell Saving Plans, wurde vorgestellt, welche Einsparungen die optionalen Vorauszahlungen erzielen. Es hat sich gezeigt, dass sich die Kombination von Zahlungsmodellen zur Kostensenkung anbietet. Dies erfolgt, wenn man für die geplante Nutzung EC2-Flotten anwendet und On-Demand Instanzen für die unvorhergesehene Nutzung einsetzt. In dem Anwendungsfall von Truecar Inc. im Unterkapitel 3.5 wurde bestätigt, dass die korrekte Berechnung der künftig nötigen reservierten Instanzen und deren spätere Überwachung, erhebliche Einsparungen erzielen.

In Kapitel 4 wurden drei Überwachungswerkzeuge untersucht, die eine bessere Überwachung von Cloud-Diensten ermöglichen. Die damit gesammelten Informationen dienen den betriebswirtschaftlichen Abteilungen der Unternehmen und gewähren eine bessere Kontrolle über die Kosten von Cloud-Diensten. Cost-Explorer verschafft einen umfangreichen Überblick der Nutzung und der Kosten. Zwei Einsatzmöglichkeiten sind die operative Budgetplanung und die Verfolgung von Leistungskennzahlen. Die Metriken bei CloudWatch dienen als Ergänzung für die Verfolgung von Leistungskennzahlen. Deren Dashboards und die Benachrichtigungen von relevanten Ereignissen haben darüber hinaus gezeigt, wie man über den Stand der Cloud-Dienste in Echtzeit informiert werden kann. Trusted Advisor wurde aufgrund des fehlenden Zugangs zu einem Business- oder Enterprise-Plan nur eingeschränkt untersucht. Das Referenzhandbuch über Trusted Advisor war nicht ausreichend, um Berechnungen, Prüfungen oder Ähnliches durchführen zu können. Die Informationen,

die dem Referenzhandbuch entnommen werden konnten, geben jedoch Hinweise auf die in dieser Arbeit behandelten Optimierungsmaßnahmen. Zum Beispiel die Verwendung von reservierten Instanzen, das Abschalten von Instanzen mit geringer Auslastung und die Verwendung von Budgets mit Benachrichtigungen.

In Kapitel 5 wurden anhand von definierten Arbeitszeiten mögliche Einsparungen für eine nicht-produktive Entwicklungsumgebung mit On-Demand EC2-Instanzen berechnet. Außerdem wurde erklärt, wie Auto Scaling-Gruppen und Load Balancer die Leistung einer balancierten und vertikal skalierbaren Serverinfrastruktur ermögichen können. Mit Fokus auf Amazon S3 wurde die Verschiebung von Objekten zwischen Speicherklassen mithilfe von Lebenszyklus-Konfiguration und Intelligent-Tiering vorgestellt. Für beide Optimierungsmaßnahmen wurden jeweils Berechnungen zu den möglichen Einsparungen anhand eines Anwendungsbeispiel durchgeführt.

Eines der in dieser Arbeit festgestellten Probleme war der mangelnde Wissensstand über die Entstehung von Kosten durch Cloud-Dienste. Einer der Gründe dafür ist, dass es sich um eine relativ neue Technologie auf dem Markt handelt.[122] Als Lösung für dieses Problem werden Ausbildungsmaßnahmen, insbesondere im Bezug auf Kostenüberwachung und -optimierung, empfohlen. Schlussendlich müsste die IT-Infrastruktur so einzurichtet werden, dass die entstehenden Kosten leicht zugänglich sind, dass Grenzen für detaillierte Budgets festgelegt, und die richtigen Personen bei Budgetüberschreitung informiert werden. Als Ergebnis hätten Unternehmen mehr Kontrolle über die Kosten, die durch die Nutzung von Cloud-Diensten verursacht werden. Sie würden von einer flexiblen Infrastruktur profitieren, die sich an ständig ändernde Geschäftsanforderungen anpassen lässt.

Während der Entwicklung dieser Arbeit wurden Cost-Explorer, CloudWatch und Trusted Advisor (mit Einschränkungen) mit dem kostenlosen AWS-Kontingent getestet. Es wird empfohlen, dass die oben genannten Überwachungswerkzeuge in einer echten IT-Infrastruktur angewendet und die entsprechenden Optimierungsmaßnahmen für Amazon S3 und EC2-Instanzen ergriffen werden. Um damit, die Ergebnisse dieser Arbeit unter realen Bedingungen zu erproben.

[122] AWS gibt es seit 2006, Azure seit 2010 und Google Cloud seit 2008. (Vgl. Anders Lisdorf, 2021, Cloud Computing Basics: a Non.-Technical Introduction, S.72, S.80 und S.91 [4])

Quellenverzeichnis

Literatur

[1] Mark Wilkins (2021): *AWS Certified Solutions Architect - Associate (SAA-C02)*

ISBN: 9780137325160

[2] Marceil Schweitzer und Ernst Troßmann (1998): *Break-even-Analysen Methodik und Einsatz.*
https://www.wiso-net.de/document/DUHU__9783428490882522
ISBN: PDF 978-3-428-49088-2

[3] V.Bach, & H. Österle (1999): *Business Knowledge Management: Wertschöpfung durch Wissensportale.*
ISBN: 3-540-42804-6

[4] Anders Lisdorf (2021): *Cloud Computing Basics: a Non.-Technical Introduction.*
Apress.
ISBN-13 (pbk): 978-1-4842-6920-6

[5] Helmut Krcmar (2015): *Informationsmanagement.* 6. Auflage.
ISBN: 978-3-662-45863-1 (eBook)

[6] Theo PetersNicole Schelter (2021): *Kompakte Einführung in das Projektmanagement.* https://link.springer.com/book/10.1007%2F978-3-658-31194-0
ISBN: 978-3-658-31194-0

[7] Marc Opitz (2019): *Prozessorientiertes Reporting*
https://content-select.com/de/portal/media/view/5e419784-8730-4de1-a69d-561eb0dd2d03?forceauth=1
(Abgerufen am 11.12.2021)
ISBN: 9783791046556

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 (Vollständiger Bericht auf Englisch)
https://www.accenture.com/_acnmedia/PDF-139/Accenture-Cloud-Outcomes-Exec-Summary.pdf#zoom=40
(Veröffentlicht Juli-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 Pricing
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 EC2
https://calculator.aws/#/createCalculator/EC2 (Abgerufen am 23.11.2021)

[19] AWS Pricing Calculator S3
https://calculator.aws/#/createCalculator/S3 (Abgerufen am 13.12.2021)

[20] 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)

[21] 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)

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

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

[24] AWS Cost Management Pricing
https://aws.amazon.com/de/aws-cost-management/pricing/
(Abgerufen am 30.11.2021)

[25] Amazon EC2 Reserved Instance Marketplace
https://aws.amazon.com/de/ec2/purchasing-options/reserved-instances/marketplace/
(Abgerufen am 30.11.2021 - Veröffentlicht: 13.05.2020)

[26] AWS by Ben Peven: Running Web Applications on Amazon EC2 Spot Instances
https://aws.amazon.com/de/blogs/compute/running-web-applications-on-amazon-e
(Abgerufen am 01.12.2021)

[27] AWS EC2 Spot Instanzen-Anfragen und Preisverlauf
https://console.aws.amazon.com/ec2sp/v1/spot/home?
(Abgerufen am 01.12.2021)

[28] Amazon Elastic Compute Cloud - Benutzerhandbuch für Linux-Instances
https://docs.aws.amazon.com/de_de/AWSEC2/latest/UserGuide/ec2-ug.pdf#spot-best-practices (Abgerufen am 01.12.2021)

[29] AWS X-Ray Developer Guide: What is AWS X-Ray?
https://docs.aws.amazon.com/xray/latest/devguide/xray-guide.pdf#aws-xray
(Abgerufen am 03.12.2021)

[30] AWS CloudTrail User Guide Version 1.0: What Is AWS CloudTrail?
https://docs.aws.amazon.com/awscloudtrail/latest/userguide/awscloudtrail-ug.pdf#cloudtrail-user-guide
(Abgerufen am 03.12.2021)

[31] AWS – Allgemeine Referenz - Referenzhandbuch
https://docs.aws.amazon.com/general/latest/gr/aws-general.pdf#aws_tagging (Abgerufen am 04.12.2021)

[32] AWS – Amazon SNS
https://aws.amazon.com/de/sns/ (Abgerufen am 04.12.2021)

[33] Amazon EC2 Auto Scaling - Benutzerhandbuch
https://docs.aws.amazon.com/de_de/autoscaling/ec2/userguide/as-dg.pdf#what-is-amazon-ec2-auto-scaling
(Abgerufen am 05.12.2021)

[34] AWS CloudFormation - Benutzerhandbuch
https://docs.aws.amazon.com/de_de/AWSCloudFormation/latest/UserGuide/cfn-ug.pdf#quickref-cloudwatch
(Abgerufen am 05.12.2021)

[35] AWS Single Sign-On
https://aws.amazon.com/single-sign-on/?nc1=h_ls
(Abgerufen am 05.12.2021)

[36] AWS Marketplace
https://aws.amazon.com/mp/marketplace-service/overview/ (Abgerufen am 06.12.2021)

[37] Erin Carlson and Alee Whitman. Getting Started: Tracking AWS Cost Management Metrics
https://aws.amazon.com/blogs/aws-cloud-financial-management/getting-started-tracking-aws-cost-management-metrics/
(Abgerufen am 06.12.2021)

[38] AWS Support - Benutzerhandbuch
https://docs.aws.amazon.com/de_de/awssupport/latest/user/support-ug.pdf#trusted-advisor
(Abgerufen am 07.12.2021)

[39] AWS Support Plan Pricing
https://aws.amazon.com/premiumsupport/pricing/?nc1=h_ls
(Abgerufen am 09.12.2021)

[40] Amazon Elastic Container Service Entwicklerhandbuch - Load Balancer-Typen
https://docs.aws.amazon.com/de_de/AmazonECS/latest/developerguide/ecs-dg.pdf#load-balancer-types
(Abgerufen am 09.12.2021)

[41] 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 10.06.2020)

[42] Definition von Ausgabe im Rechnungswesen
https://wirtschaftslexikon.gabler.de/definition/ausgaben-31469#head1 (Abgerufen am 11.12.2021)

[43] Bürgerlichen Gesetzbuch (BGB) § 630f
https://www.gesetze-im-internet.de/bgb/__630f.html
(Abgerufen am 08.12.2021)

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

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

[46] Ubuntu, delivered by Canonical:A business guide to hybrid/multi-cloud
https://ubuntu.com/engage/multi-cloud-business-guide?utm_source=google_ad&utm_medium=cpc&utm_campaign=7014K000000mSwp&gclid=CjOKCQiAtJeNBhCVARIsANJUJ2Fb2Xp3WST3woFmmIl1ZfqsMTRzvLVld-BlPEOyKVxdhm4tgxMkIwcB
(Abgerufen am 29.11.2021)

[47] The NIST Definition of Cloud Computing
National Institute of Standards and Technology(NIST) https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf
(Abgerufen am 09.12.2021)

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

[49] Instagram: Wann verschwindet meine Instagram Story?
https://help.instagram.com/1729008150678239 (Abgerufen am 08.12.2021)

[50] Online Marketing: Definition von Instagram Story?
https://onlinemarketing.de/lexikon/definition-instagram-story (Abgerufen am 09.12.2021)

[51] Hilfebereich Instagram: VII. Datenspeicherung, Deaktivierung und Löschung von Konten.
https://help.instagram.com/519522125107875 (Abgerufen am 12.12.2021)

[52] 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)

[53] Definition von Customer Acquisition Cost (CAC)
https://onlinemarketing.de/lexikon/definition-customer-acquisition-cost-cac
(Abgerufen am 06.12.2021)

[54] Definition von Freemium
https://onlinemarketing.de/lexikon/definition-freemium (Abgerufen am 06.12.2021)

[55] Definition von Cost-per-Action (CPA)
https://onlinemarketing.de/lexikon/definition-cost-per-action-cpa
(Abgerufen am 06.12.2021)

[56] 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)

[57] 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)

[58] SAP: Definition von maschinellen Lernen
https://www.sap.com/germany/insights/what-is-machine-learning.html (Abgerufen am 09.12.2021)

[59] Medium: How TrueCar Saves 40% on AWS with EC2 Reserved Instances
https://medium.com/driven-by-code/how-truecar-saves-40-on-aws-with-ec2-reser (Abgerufen am 02.12.2021)

[60] Techterms Definition Metadata.
https://techterms.com/definition/metadata
(Abgerufen am 08.12.2021)

[61] 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)

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

[63] Definition Horizontal Scaling
https://www.techopedia.com/definition/7594/horizontal-scaling?ref=wellarchitected (Abgerufen am 09.12.2021)

[64] Definition von TCO
https://www.gartner.com/en/information-technology/glossary/total-cost-of-ownership-tco (Abgerufen am 18.12.2021)

[65] Definition Slack
https://slack.com/intl/de-de/help/articles/115004071768-Was-ist-Slack- (Abgerufen am 11.12.2021)

[66] 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)

[67] Spot by NetApp, What are AWS spot instances?
https://spot.io/what-are-ec2-spot-instances/
(Abgerufen am 01.12.2021)

[68] 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)

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

[70] 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)

[71] 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)

[72] 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)

[73] 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-deutschland
(Abgerufen am 29.11.2021)

[74] Statista: Amazon ist die Nummer 1 in der Cloud
https://de.statista.com/infografik/20802/weltweiter-marktanteil-von-cloud-in
(Abgerufen am 08.12.2021)

[75] Ashish G. Revar, Madhuri D. Bhavsar. Securing User Authentication using Single SignOn in Cloud Computing.
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6153227
(Abgerufen am 05.12.2021)

[76] YAML Org. Definition of YAML https://yaml.org/ (Abgerufen am 12.12.2021)

Anhang

Vorlage für einer Fakturierungsalarme in CloudWatch

JSON

```
    "SpendingAlarm": {
 "Type": "AWS::CloudWatch::Alarm",
 "Properties": {
 "AlarmDescription": { "Fn::Join": ["", [
 "Alarm if AWS spending is over $",
 { "Ref": "AlarmThreshold" }
 ]]},
 "Namespace": "AWS/Billing",
 "MetricName": "EstimatedCharges",
 "Dimensions": [{
 "Name": "Currency",
 "Value" : "USD"
 }],
 "Statistic": "Maximum",
 "Period": "21600",
 "EvaluationPeriods": "1",
 "Threshold": { "Ref": "AlarmThreshold" },
 "ComparisonOperator": "GreaterThanThreshold",
 "AlarmActions": [{
 "Ref": "BillingAlarmNotification"
 }],
 "InsufficientDataActions": [{
 "Ref": "BillingAlarmNotification"
 }]
 }
}
```

YAML

```
    SpendingAlarm:
    Type: AWS::CloudWatch::Alarm
    Properties:
```

```
AlarmDescription:
'Fn::Join':
- ''
- - Alarm if AWS spending is over $
- Ref: AlarmThreshold
Namespace: AWS/Billing
MetricName: EstimatedCharges
Dimensions:
- Name: Currency
Value: USD
Statistic: Maximum
Period: '21600'
EvaluationPeriods: '1'
Threshold:
Ref: "AlarmThreshold"
ComparisonOperator: GreaterThanThreshold
AlarmActions:
- Ref: "BillingAlarmNotification"
InsufficientDataActions:
- Ref: "BillingAlarmNotification"
```

[123]

[123] AWS CloudFormation - Benutzerhandbuch. S.481.[34]

Alarm für die monatliche Kosten anhand eines Budgets

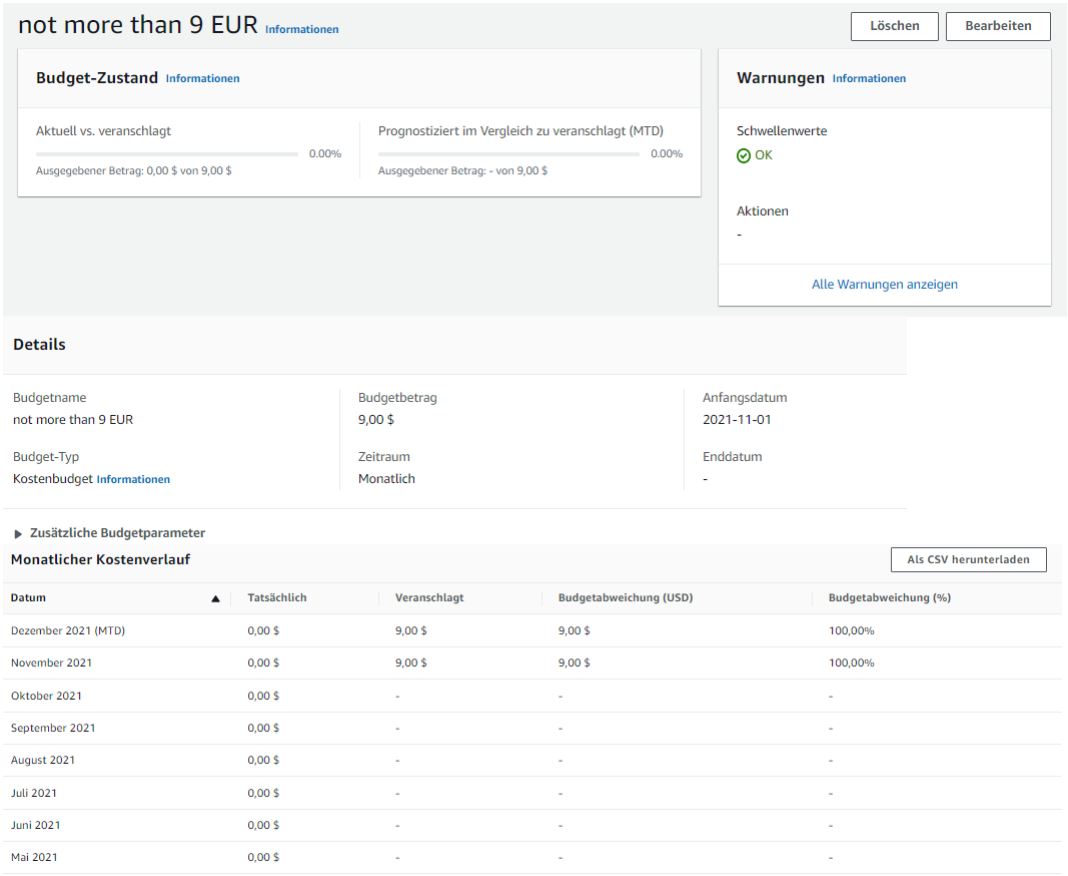


Abbildung 22
Eigene Darstellung von Test AWS-Konto.

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)