

Technology Arts Sciences TH Köln

Technische Hochschule Köln

Fakultät für Informatik und Ingenieurwissenschaften

BACHELORARBEIT

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

Vorgelegt an der TH Köln Campus Gummersbach
im Studiengang Wirtschaftsinformatik

ausgearbeitet von:

CARLO MENJIVAR 11117929

Erstprüfer: Prof. Dr. Roman Majewski

Zweitprüfer: Thomas Raser

Gummersbach, im Feb<Monat der Abgabe>

1 Abstract

Zusammenfassung

Diese Arbeit beschäftigt sich damit, wie mehr Kontrolle über die Kosten von Cloud-Diensten erhalten wird, indem sie überwacht werden. In Kombination damit werden Maßnahmen und Werkzeuge untersucht, die zu erheblichen Kosteneinsparungen in der Cloud führen.

Angefangen bei der Wahl des richtigen Zahlungsmodells, über das automatische Herunterfahren ungenutzter Instanzen zu bestimmten Zeiten bis hin zur Implementierung von Autoscaling .

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

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

Abstract

Platz für das englische Abstract...

Inhaltsverzeichnis

1	Abstract	1
2	Abkürzungsverzeichnis	5
3	Einleitung	6
3.1	Motivation	6
3.2	Problemstellung	6
3.3	Fragestellung	7
3.4	Zielsetzung	7
3.5	Einschränkungen	7
3.6	Struktur der Arbeit	8
4	Grundlagen	9
4.1	Ökonomie des Cloud Computing	9
4.1.1	Skalierbarkeit	10
4.1.2	Flexibilität und Agilität	10
4.1.3	Selbstbedienung	11
4.1.4	Keine Vorabkosten	11
5	Zahlungsmodelle	12
5.1	On-Demand / Nutzungsabhängige Zahlung	12
5.2	Reservierte Instanzen und Saving Plans	13
5.2.1	Vorauszahlung	14
5.3	Versteigerung? / Spot Instanzen	14
5.4	Wann welches Zahlungsmodell?	15
6	Kostenüberwachung	16
6.1	AWS CloudWatch	16
6.1.1	Fakturierungsalarms mit CloudWatch	18
6.1.2	Alarm bei Hoch- und Runterfahren von EC2-Instanzen	18
6.2	AWS Cost-Explorer	18
6.3	AWS Trusted advisor	19
7	Optimierungsmöglichkeiten	21
7.1	Optimierungsmaßnahmen	21
7.1.1	Auto Scaling Group	21
7.1.2	Automatisiere das Hoch- und Herunterfahren von Instanzen	22
7.1.3	(auto) Tiering	22

7.1.4	Automatisierung mit Lambda Funktionen	22
7.1.5	VERKAUFE DEINE Ungenutzte Kapazität in RI Marketplace . . .	22
8	Zusammenfassung und Ausblick	23
8.1	Bewusstsein in der gesamten Organisation	23
8.2	Die richtige Personen(Owneship verbreiten)	23
9	Glossar	24
10	Quellenverzeichnis	25
10.1	Literatur	25
10.2	Internetquellen	25
A	Anhang	28
A.1	Anhang X	28
	Erklärung über die selbständige Abfassung der Arbeit	29

ABBILDUNGEN

2 Abkürzungsverzeichnis

AWS: Amazon Web Services

Amazon EC2:

Amazon Elastic Compute Cloud

API

Application Programming Interface

AJAX

Asynchronous JavaScript And XML

CI/CD

Continuous Integration/ Continuous Deployment

TCO

Total Cost of Ownership

3 Einleitung

3.1 Motivation

Amazon Web Services, kurz AWS, wurde unter anderem für diese Arbeit ausgewählt wegen seiner frühen Präsenz (2006) als Cloudanbieter und seines großen Angebotes an Dienstleistungen, welche für zahlreiche Anwendungsfälle geeignet sind.

Eine Recherche von Gartner positioniert AWS als Marktführer in der Magic Quadrant für Cloud-Infrastruktur und Plattform-Services 2021. [22] Kostenoptimierung für Cloud-Dienste ist ein wichtiger Punkt, da man ohne Optimierungsmaßnahmen mit höheren Kosten rechnen muss als bei On-Premise Systemen.

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

3.2 Problemstellung

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

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

„In its Stratecast Predictions 2018, Frost & Sullivan noted that 53% of IT leaders surveyed cited “managing costs to run cloud workloads” as a huge obstacle, and over 50% have difficulty justifying the expenses of some public cloud workloads.“ [23]

Diese Bachelorarbeit beschäftigt sich mit ebendieser Problematik, um herauszufinden, wie Unternehmen mit den passenden Werkzeugen die Kosten ihrer Cloud-Dienste überwachen und im Blick behalten können.

Zum Beispiel können frühzeitige Benachrichtigungen alarmieren, wenn Ressourcen mehr Kosten verursachen als geplant.

¹[2], Seite 152

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

3.3 Fragestellung

In dieser Arbeit wird versucht, die folgenden Fragen beantworten.

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

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

3.4 Zielsetzung

Daraus ergeben sich für die Arbeit die folgenden Ziele:

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

3.5 Einschränkungen

Der Schwerpunkt dieser Arbeit liegt auf EC2-Instanzen, da diese in der Regel den größten Anteil an der Rechnung ausmacht. An zweiter Stelle stehen S3-Speichereinheiten, weil sie einen erheblichen Teil der Kosten darstellen.

[9, 10]

Diese Arbeit legt den Fokus auf die Optimierung der oben genannten Dienste. Als Überwachungswerkzeuge für die Kosten werden die AWS CloudWatch, der AWS Cost-Explorer und der AWS Trusted Advisor untersucht.

3.6 Struktur der Arbeit

Diese Bachelorarbeit ist in folgende Kapitel unterteilt:

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

Kapitel 4 zeigt die verschiedenen Zahlungsmodelle für Amazon Web Services. Es werden Kriterien vorgestellt, die helfen, sich für das richtige Zahlungsmodell bei verschiedenen Szenarien zu entscheiden.

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

Kapitel 6 befasst sich mit Optimierungsmaßnahmen insbesondere für EC2-Instanzen und S3 Speichereinheiten.

4 Grundlagen

In diesem Grundlagenkapitel werden Erfolgschancen für Unternehmen aufgelistet, die Cloud-Dienste in ihre Geschäftsprozesse integrieren.

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

Folgende Ergebnisse können erreicht werden, wenn Überwachungs- und Optimierungsmaßnahmen eingeführt werden:

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

4.1 Ökonomie des Cloud Computing

[Date last review: 05.11 Sarah]

Cloud Economics auf Englisch, basierend auf dem On-Demand Prinzip, besitzt die Flexibilität, die Rechenkapazität je nach Bedarf anzupassen. Es entfällt die Notwendigkeit, hohe Investitionen in Hardware zu tätigen, wie bei On-Premise-Systemen. Durch den Verzicht auf Hardware entfallen die Kosten für Reparatur, Wartung und eventuell damit verbundene Lizenzen.

Der Cloud-Anbieter übernimmt viele Verwaltungsaufgaben. Das führt zu einer Abnahme der nötigen Fachkraft. [21]

Die Nutzung von Cloud-Diensten ist in unabhängiger Weise möglich. HIER GUT FORMULIEREN DASS AWS NICHT INVOLVIERT SEIN MUSS. Cloud-Dienste können auch in Selbstbedienung genutzt werden. Das bedeutet jedoch gleichzeitig, dass der Nutzer Verantwortung für die anfallenden Kosten übernimmt³.

[Grafik der Kosten On-Premise/Demand?]

²TCO: Total Cost if Ownership

³Nutzer von Cloud-Diensten

4.1.1 Skalierbarkeit

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

Mit Auto Scaling wird sichergestellt, dass die Rechenkapazität in Zeiträumen von hoher Nachfrage automatisch hochskaliert werden kann.

Auf diese Weise kann Zeit mit der Verwaltung von IT - Ressourcen gespart werden, um sich auf die wesentlichen Geschäftsaktivitäten zu konzentrieren⁴.

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

„By getting out of the business of managing datacenters, WBA[Walgreens Boots Alliance] can spend less time worrying about managing IT resources and more time focusing on what it’s really good at—delivering great health-care 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.,,

[18]

4.1.2 Flexibilität und Agilität

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

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

Durch die Vielzahl der verfügbaren Optionen ist es möglich, Prototypen und Experimente in kurzer Zeit durchzuführen. ⁵.

⁴[17], Seite 29

⁵[21], Seite 7

Softwareprojekte können schnell auf den Markt gebracht werden.

Sollte ein Projekt stillgelegt werden, könnten alle damit verbundenen Kosten ausfallen.

4.1.3 Selbstbedienung

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

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

4.1.4 Keine Vorabkosten

Amazon Web Services bietet ein Pay-as-you-go-Modell für Ressourcen im On-Demand Zahlungsmodell für EC2 und Speichereinheiten wie S3. Wenn man nur für die monatlich verbrauchten Ressourcen bezahlt, verringert sich die Anfangsinvestition in die IT-Infrastruktur insgesamt. Es ist zu bedenken, dass weitere Investitionen wie die technische Schulung für das Personal erforderlich werden.

[ABSCHLUSS DES KAPITELS]

5 Zahlungsmodelle

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

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

Jedes Zahlungsmodell hat seine Vor- und Nachteile und eignet sich für unterschiedliche Anwendungsfälle. Gute Ergebnisse können auch durch die Kombination mehrerer Zahlungsmodelle erzielt werden. Dies wird in Unterkapitel 5.4 behandelt.

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

5.1 On-Demand / Nutzungsabhängige Zahlung

Bei diesem Zahlungsmodell besteht keine Notwendigkeit, ein festes Anfangsbudget festzulegen. Die Kosten richten sich nach dem Verbrauch auf der Grundlage der Nutzungsstunden.

Dieses Modell eignet sich für Projekte, bei denen nicht viel vorhersehbar ist und die Möglichkeit besteht, dass das Projekt in kurzer Zeit abgeschlossen sein wird, so dass es keinen Sinn macht, eine langfristige Verpflichtung einzugehen.

Hier einige Beispiele von EC2-Instanzen im On-Demand Zahlungsmodell.

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

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

363 von 363 verfügbaren Instances werden angezeigt

Q < 1 2 3 4 5 6 7 ... 19 >

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

Abbildung Unterabschnitt 5.1: On-Demand Preise für Amazon EC2 [3]

Der Preis für dieses Modell variiert je nach Instanztyp, Region und der übertragenen Datenmenge. Die aktuellen Preise für die verschiedenen Regionen sind auf der Amazon-Website in der Sektion EC2 - On-Demand-Preise⁶ zu finden. In der Abbildung 5.1 werden die für die Region Ohio verfügbaren Linux-Instanzen gezeigt. Je Es ist zu beachten, dass Instanzen mit denselben Eigenschaften, aber in verschiedenen Regionen, unterschiedliche Preise haben können.

5.2 Reservierte Instanzen und Saving Plans

Die Zahlungsmodelle Reservierte Instanzen und Saving Plans sind sich sehr ähnlich. Beide kommen mit einer gleichbleibenden Nutzungsverpflichtung, die in \$-€/Stunden gemessen wird.

Um die reduzierten Preise zu bekommen, müssen Verträge 1 oder 3 Jahre abgeschlossen werden.

Nachfolgend werden die prozentualen Einsparungen gemäß des jeweiligen Modells gezeigt.

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

[8, 12]

⁶<https://aws.amazon.com/de/ec2/pricing/on-demand/>

Die ersten beiden Optionen in der obigen Tabelle, die Saving Plans, unterscheiden sich dadurch, dass die Compute Savings Plans die Flexibilität bieten, EC2-Instanzen nach Familie⁷, Größe, Availability Zone (AZ), Betriebssystem oder Mandant zu wechseln.

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

Bei den EC2-Instance Saving Plans hingegen muss eine Instance-Familie in einer bestimmten Region ausgewählt werden. Dies reduziert automatisch die Kosten für die ausgewählte Instanz-Familie in der jeweiligen Region, unabhängig von Availability Zone, Größe, Betriebssystem oder Mandant.

REVIEW

Die Festlegung eines festen Stundensatzes über einen langen Zeitraum bietet die Möglichkeit, künftige Kosten zu planen.

Folgenden Attributen definieren den Preis von EC2-Instanzen bei Saving Plans:

- Vertragslaufzeit
- Vorabzahlung
- Betriebssystem
- Region
- Mandant

AUCH FÜR RIs?

WANN SAVING PLANS?

5.2.1 Vorauszahlung

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

Amazon bietet drei verschiedene Optionen an:

keine, teilweise und vollständige Vorauszahlung.

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

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

5.3 Versteigerung? / Spot Instanzen

EC2 Spot-Instances bieten die Möglichkeit aus von anderen Nutzern ungenutzter EC2-Instances zu profitieren. Mit einem Preisvorteil von bis zu 90 % gegenüber normalen On-Demand-Instanzen sind Spot-Instanzen ideal für fehlertolerante Anwendungen wie auf

⁷[17], Seite 95

Containern ausgeführte Workloads, CI/CD, Bigdata-Anwendungen und ähnliches. [IST DER SATZ ZU LANG?]

Es ist zu beachten, dass Spot-Instanzen jederzeit unterbrochen werden können, entweder weil der konfigurierte Maximalpreis überschritten wurde oder weil die Instanz nicht mehr verfügbar ist. Dies geschieht, wenn der Instanzanbieter die Instanz erneut anfordert. Falls eine Spot-Instanz unterbrochen wird, benachrichtigt Amazon EC2 2 Minuten im Voraus. Dieses Ereignis ist verfügbar auf CloudWatch.

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

Um von der Preisvorteile der Spot-Instanzen zu profitieren und Ausfälle zu vermeiden, sollten in Kombination weitere Zahlungsmodelle verwendet werden.

Zum Beispiel eine Kombination aus Spot-Instanzen für die erwarteten Last und On-Demand-Instanzen für die dynamischen Last.

5.4 Wann welches Zahlungsmodell?

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

[ABSCHLUSS DES KAPITELS/Überbrückung FÜR DAS KOMMENDE KAPITEL]

6 Kostenüberwachung

(CHECK/VERBESSERN Version 10.11)

In diesem Kapitel werden Werkzeuge vorgestellt, mit denen Budgets mit Alarmen erstellt werden, diese informieren, wenn ein bestimmter Prozentsatz des festgelegten Budgets überschritten wurde. Die Erstellung von Budgets trägt zu einer besseren Planung-/Prognose- und Kostenkontrolle. ZITAT? Die Aktivierung von Alarmen[PASSENDEN WORT?] ist eine der Maßnahmen, die es uns ermöglichen, auf aktive Ressourcen aufmerksam zu werden, die möglicherweise noch laufen, weil man vergessen hat, sie abzuschalten. Darüber hinaus ist es mit Werkzeugen wie CloudWatch möglich, die Abschaltung bestimmter Dienste zu automatisieren, wenn eine Budgetschwelle überschritten wurde.

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

TRUSTED ADVISOR??

6.1 AWS CloudWatch

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

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

Eine der Metriken, die mit Amazon CloudWatch überwacht werden kann, ist die CPU-Last von EC2-Instanzen. Basierend auf einem Prozentsatz der CPU-Last können Alar-me?Benachrichtigungen?[WELCHES WORT?] und Aktionen konfiguriert werden

Zum Beispiel, eine dieser Aktionen ist die automatische Einrichtung neuer Instanzen zur Deckung des Kapazitätsbedarfs ⁸. Diese Art von Aktionen werden im Kapitel 7 Optimierungsmöglichkeiten tiefer behandelt.

⁸[17], Seite 185

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

Visualisierung/ Dashboards

Mit Cloud-Watch Dashboards können Boards[BESSERE FORMULIERUNG] mit relevanten Metriken angelegt werden. Sie ermöglichen eine umfassende Visualisierung von Metriken der Ressourcen in der AWS-Umgebung. Durch die Boards[PASSENDES WORT?], können auch Alarmen erstellt werden. Für die Einrichtung der Alarme ist kein technisches Wissen benötigt⁹.

Die von CloudWatch gesammelte Informationen sind nicht nur relevant für System-Administratoren, es könnten auch für andere Personen innerhalb oder außerhalb der Organisation relevant sein. Für sollte Fällen bietet CloudWatch die Möglichkeit, eine gemeinsame Sicht auf kritische Ressourcen- und Anwendungsmessungen zu generieren. Dies ermöglicht einen schnelleren Kommunikationsfluss in Echtzeit. Die Zugriffsverwaltung für geteilte Dashboards wird über AWS Identity and Access Management abgewickelt¹⁰.

Metriken

Eine Metrik stellt eine Reihe von Daten über die Leistung einer Ressource in zeitlicher Reihenfolge dar. Standardmäßig werden viele kostenlose Metriken an CloudWatch übermittelt.

Metriken bleiben wie folgt verfügbar:

Auflösung/Zeitraum:

1 Minute - 15 Tage

5 Minuten - 63 Tage.

1 Stunde - 15 Monate ¹¹

Zum Beispiel kann der Durchschnitt von einer bestimmten API pro Stunde untersucht werden. Für eine detailliertere Überwachung ist es möglich, benutzerdefinierte Metriken zu konfigurieren, die eine Auflösung von bis zu 1 Sekunde zulassen. Ein praktisches Beispiel für benutzerdefinierte Metriken ist die Messung der Ladezeit einer Website. [EIN BEISPIEL MIT BEZUG AUF K:OPTIMIERUNG?oder WIESO IST DIE AUFLÖSUNG RELEVANT?]

Ereignisse / Events

Bei CloudWatch ist ein Ereignis eine Änderung bei einer Ressource in der AWS-Umgebung. AWS-Ressourcen können Ereignisse erzeugen, wenn sich ihr Status ändert.

Beispielsweise, ein Ereignis wird erzeugt, wenn Amazon EC2 Auto Scaling, Instanzen

⁹[15], Seite 28

¹⁰[15], Seite 18 und 39

¹¹[16], Wie lange werden die verfügbaren Metriken aufbewahrt?

startet oder beendet ¹².

Regel

Eine Regel ordnet eintreffende Ereignisse zu und leitet diese zur Verarbeitung an Ziele weiter. Eine einzelne Regel kann an mehrere Ziele weiterleiten, die alle parallel verarbeitet werden¹³.

Target / Ziele

Ziele sind Ressourcen, die aufgerufen werden, wenn eine Regel ausgelöst wird. EC2 instances, AWS Lambda functions und Amazon SNS topics sind unter anderem mögliche Ziele. Die Ziele einer Regel müssen sich in derselben Region wie die Regel befinden ¹⁴.

Alarmer Benachrichtigungen?

Benachrichtigt zu werden ist es wichtig, um relevante Ereignisse nicht zu verpassen und rechtzeitig Maßnahmen zu ergreifen. Mit CloudWatch können Alarmer eingerichtet werden, die durch Metriken wie die CPU-Last und auch Gebühren[ANDERES WORT] auf AWS-Rechnungen ausgelöst werden.

Benachrichtigungen können durch Amazon SNS oder zu einer E-Mail-Adresse geschickt werden.[NUR?] Autoscaling group seted with cooldown periods to avoid too much instances to by launched.

6.1.1 Fakturierungsalarms mit CloudWatch

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

6.1.2 Alarm bei Hoch- und Runterfahren von EC2-Instanzen

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

6.2 AWS Cost-Explorer

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

¹²[14], Seite 1

¹³[14], Seite 2

¹⁴[14], Seite 2

die nächsten Monate erstellt. Die Prognose basieren auf die Kosten der vergangenen Monaten. Die Nutzung des Cost Explorers ist kostenlos, nur API-Aufrufe sind kostenpflichtig¹⁵.

Es gibt drei Arten von Berichten, die der Cost Explorer bereitstellt: -reporting on the usage and cost incurred in the past 12 months. Bericht über die Nutzung und die in den letzten 12 Monaten entstandenen Kosten

-forecasting how much you are going to spend in the next three months based on your past usage. Prognostizieren Sie Ihre voraussichtlichen Kosten. Dient zur Planung...

- and getting recommendations as well on the type of instances to use. Empfehlungen zu Reserved Instances. Amazon analysiert die bisherige Nutzung der Instanzen und gibt Empfehlungen zur Kostensenkung durch den Wechsel von EC2-Instanzen zu reservierten Instanzen. Amazon generiert Empfehlungen automatisch, indem es die Nutzung der EC2-Instanzen analysiert. Diese ignorieren Kapazität, die bereits von anderen reservierten Instanzen abgedeckt wurden.

Budgetplanung

Die Prognose der kommenden Kosten, dienen zu einer guten operativen Budgetplanung.

Aktuelle Budgets vs erwartet

6.3 AWS Trusted advisor

AWS Trusted Advisor ist ein Werkzeug, welches hilft, Kosten zu senken, Systemverfügbarkeit und -leistung zu verbessern und Sicherheit zu erhöhen. Es analysiert die Nutzung des AWS-Kontos und gibt Best-Practice-Empfehlungen. Es werden die Kategorien Leistungsgrenzen und Kostenoptimierung betrachtet, da diese am relevanten für die vorliegende Arbeit sind. [BILD DER KATEGORIEN]

Kostenoptimierung

Die Empfehlungen zur Kostenoptimierung konzentrieren sich auf Möglichkeiten zur Kostensenkung, indem ungenutzte Ressourcen hervorgehoben werden.

Sollten EC2-Instanzen mit geringer Auslastung gefunden werden, wird es diese bei Trusted Advisor signalisiert. Denn diese Instanzen verbrauchen Ressourcen und können terminiert oder pausiert werden. Nicht zugewiesene Elastic IP-Adressen erzeugen Kosten. Diese können gegebenenfalls von Trusted Advisor gefunden werden.

Leistungsgrenzen

In dieser Kategorie werden Empfehlungen zur Vermeidung von Grenzwertüberschreitungen hervorgehoben. Es wird zum Beispiel nach einer Nutzung gesucht, die mehr als 80 %

¹⁵<https://aws.amazon.com/de/aws-cost-management/pricing/>

des Leistungsgrenzwerten für wichtige Dienste beträgt, wie Amazon EC2, Auto Scaling, Elastic Block Store, Simple Email Service, AWS CloudFormation.

Sich dieser Grenzen bewusst zu sein, gibt die Möglichkeit, rechtzeitig zu handeln und es trägt zu Kostenüberwachung bei. [SAGEN OB DIESE GRENZE DIE GLEICHE WIE BEI CloudWatch SIND]

Summary/Fazit

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

7 Optimierungsmöglichkeiten

Mit dem Explorer wurden Ressourcen gefunden, die Möglichkeiten zur Optimierung bieten. Mit CloudWatch wurde es gezeigt, ob man On-Demand zu reservierten Instanzen wechseln sollte.

Die mit den Überwachungswerkzeuge gesammelte Informationen, bilden die Grundlage für die Optimierungsmaßnahmen.[HIER KONKRETER WERDEN]

7.1 Optimierungsmaßnahmen

7.1.1 Auto Scaling Group

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

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

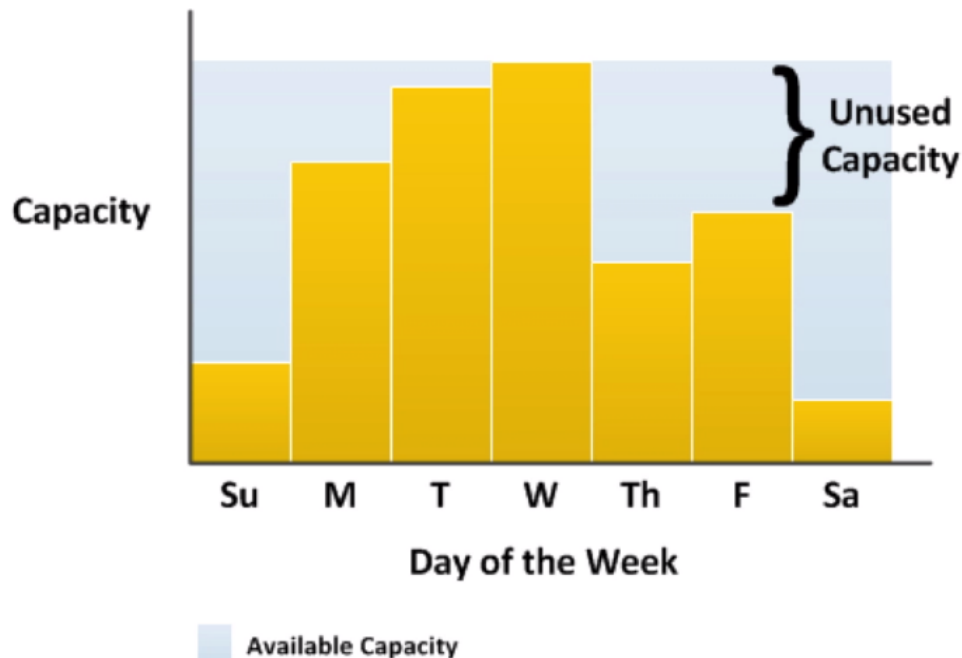


Abbildung Unterunterabschnitt 7.1.1: Ungenutzte Ressourcen ¹⁶

TODO:Es gibt verschiedene Möglichkeiten für die Auto-Skalierung.

- Maintain current Instance Levels: kann die spezifizierte oder minimale Anzahl von laufenden Instanzen zu jeder Zeit konfiguriert werden.

¹⁶Vgl. u.a.[2]

- Manual Scaling: Es wird benötigt, die minimale, die maximale oder die gewünschte Kapazität Ihrer Auto-Scaling-Gruppe zu definieren. Dies ist nützlich für sehr seltene Ereignisse.
- Scheduled Scaling: Wenn dies genau bekannt ist, muss die Anzahl der Instanzen in der Gruppe verringert oder erhöht werden. Wie bei Ereignissen wie Weihnachten oder Black Friday.
- Dynamic Scaling: DIESES WOLLEN WIE ERKLÄREN Mehrere Dienste bieten standardmäßig kostenlose Metriken für Ressourcen wie EC2, EBS und RDS-DB-Instanzen. Zu erklären: cooldown period and scaling policy.

7.1.2 Automatisiere das Hoch- und Herunterfahren von Instanzen

Grund: weil i.d.R., kein Entwickler 24/7 arbeitet. Wie?: mit Tagging, Lambda oder mit Auto Scaling Groups. HIER LESEN [2], Seite 153

7.1.3 (auto) Tiering

Grund: nicht alle Dateien brauchen eine hohe Verfügbarkeit.

7.1.4 Automatisierung mit Lambda Funktionen

Grund: einmal programmiert, funktioniert es für immer.

(To-Do:) Möglichkeiten untersuchen, bewerten und die passende Auswählen.
warum ist das empfehlungswert?

7.1.5 VERKAUFE DEINE Ungenutzte Kapazität in RI Marketplace

8 Zusammenfassung und Ausblick

(To-Do:)

8.1 Bewusstsein in der gesamten Organisation

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

8.2 Die richtige Personen(Owneship verbreiten)

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

9 Glossar

Cloud-Computing:

...

Cloud-Dienste:

...

On-Demand:

...

On-Premise:

...

Region:

Die Region ist ein völlig unabhängiges und eigenständiges geografisches Gebiet. Jede Region hat mehrere, physisch getrennte und isolierte Standorte, die als Availability Zones bekannt sind. Beispiele für Regionen sind London, Dublin, Sydney, usw [17], Seite 42.

Availability Zone:

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

Instance family:

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

¹⁷[17], Seite 42

¹⁸[17], Seite 95

10 Quellenverzeichnis

10.1 Literatur

- [1] Stickel-Wolf, Christine; Wolf, Joachim (2011): Wissenschaftliches Lernen und Lerntechniken. Erfolgreich studieren—gewusst wie!. Wiesbaden: Gabler.
- [2] Anders Lisdorf (2021): Cloud Computing Basics: a Non.-Technical Introduction. Apress.

10.2 Internetquellen

- [1] Accenture Dienstleistungen GmbH. (Veröffentlicht am 13.11.2020, abgerufen am 12.04.2021). Hohe Erwartungen an die Cloud: Hürden meistern, Mehrwert maximieren
<https://www.accenture.com/de-de/insights/technology/maximize-cloud-value>
- [2] AWS Introduction to EC2 Auto Scaling
<https://www.aws.training/Details/Video?id=16387> (Abgerufen am 23.09.2021)
- [3] AWS On-Demand Instances
<https://aws.amazon.com/de/ec2/pricing/on-demand/> (Abgerufen am 20.10.2021)
- [4] AWS-Entwicklerzentrum
<https://aws.amazon.com/de/developer/> (Abgerufen am 21.10.2021)
- [5] AWS Entwicklung kostenloser Websites und Webanwendungen
<https://aws.amazon.com/de/free/webapps/> (Abgerufen am 21.10.2021)
- [6] AWS Instance Scheduler(Abgerufen am 04.2021)
<https://aws.amazon.com/de/solutions/implementations/instance-scheduler/>
- [7] AWS S3 Intelligent-Tiering Adds Archive Access Tiers
<https://aws.amazon.com/de/blogs/aws/s3-intelligent-tiering-adds-archive-acce>

-
- [#:~:text=What%20is%20S3%20Intelligent%2DTiering](#) (Veröffentlicht am 09.11.2020)
- [8] AWS Reserved Instances Pricing
<https://aws.amazon.com/de/ec2/pricing/reserved-instances/> (Abgerufen am 22.10.2021)
- [9] AWS für Amazon EC2 Spot Instances
<https://aws.amazon.com/de/ec2/spot/pricing/> (Abgerufen am 25.10.2021)
- [10] AWS S3 Pricing
<https://aws.amazon.com/de/s3/pricing/> (Abgerufen am 25.10.2021)
- [11] AWS Databases
<https://aws.amazon.com/de/products/databases/learn/> (Abgerufen am 28.10.2021)
- [12] AWS Saving Plans Pricing
<https://aws.amazon.com/de/savingsplans/compute-pricing/> (Abgerufen am 02.11.2021)
- [13] AWS Cloud Watch Features
<https://aws.amazon.com/de/cloudwatch/features/> (Abgerufen am 03.11.2021)
- [14] AWS Cloud Watch Events: User Guide
<https://docs.aws.amazon.com/AmazonCloudWatch/latest/events/cwe-ug.pdf#WhatIsCloudWatchEvents> (Abgerufen am 04.11.2021)
- [15] AWS Cloud Watch : User Guide
https://docs.aws.amazon.com/AmazonCloudWatch/latest/monitoring/acw-ug.pdf#CloudWatch_Automatic_Dashboards_Focus_Service (Abgerufen am 04.11.2021)
- [16] AWS Cloud Watch F.A.Q.
<https://aws.amazon.com/de/cloudwatch/faqs/> (Abgerufen am 07.11.2021)
- pero es
- [17] AWS Certified Solutions Architect - Associate (SAA-C02)
[https://books.google.de/books?id=Dp__DwAAQBAJ&lpg=PA29&ots=T5WqfT25mA&dq=Increase%20efficiencies%3A%20Use%20automation%20to%20reduce%20or%20eliminate%20IT%20management%20activities%20that%](https://books.google.de/books?id=Dp__DwAAQBAJ&lpg=PA29&ots=T5WqfT25mA&dq=Increase%20efficiencies%3A%20Use%20automation%20to%20reduce%20or%20eliminate%20IT%20management%20activities%20that%20)

-
- [20waste%20time%20and%20resources.&pg=PA29#v=onepage&q&f=false](#) (Abgerufen am 02.11.2021)
- [18] Microsoft Customer Story-Walgreens Boots Alliance delivers superior customer service with SAP solutions on Azure
<https://customers.microsoft.com/en-us/story/792289-walgreens-boots-alliance-retailers-azure-sap-migration>
(Veröffentlicht am 10. Juni 2020)
- [19] Bertelsmeier, Birgit (o. J.): Tipps zum Schreiben einer Abschlussarbeit. Fachhochschule Köln-Campus Gummersbach, Institut für Informatik.
<http://lwibs01.gm.fh-koeln.de/blogs/bertelsmeier/files/2008/05/abschlussarbeitsbetreuung.pdf> (29.10.2013).
- [20] Halfmann, Marion; Rühmann, Hans (2008): Merkblatt zur Anfertigung von Projekt-, Bachelor-, Master- und Diplomarbeiten der Fakultät 10. Fachhochschule Köln-Campus Gummersbach.
<http://www.f10.fh-koeln.de/imperia/md/content/pdfs/studium/tipps/anleitungda270108.pdf> (29.10.2013).
- [21] 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)
- [22] 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/1-271W10SP-DEU.html> (Abgerufen am 23.09.2021 / Veröffentlicht am 27. Juli 2021)
- [23] 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)

A Anhang

(To-Do:)

A.1 Anhang X

Erklärung über die selbständige Abfassung der Arbeit

Ich versichere, die von mir vorgelegte Arbeit selbständig verfasst zu haben. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Arbeiten anderer entnommen sind, habe ich als entnommen kenntlich gemacht.

Sämtliche Quellen und Hilfsmittel, die ich für die Arbeit benutzt habe, sind angegeben. Die Arbeit hat mit gleichem Inhalt bzw. in wesentlichen Teilen noch keiner anderen Prüfungsbehörde vorgelegen.

(Ort, Datum, Unterschrift)

Hinweise zur obigen *Erklärung*

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