Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Amazon Web Services

Seminararbeit im Seminar DBaaS

Autor:
Anne Speerschneider

Matrikel: 6524692

Fachbereich Informatik

Seminarleiter: Felix Gessert

Hamburg, 26. Februar 2017

Einleitung

Seit einigen Jahren unterliegt die Softwareentwicklung in vielen Unternehmen einem Wandel. Wurden bisher Anwendungen „aus einem Guss“ entwickelt, entfernt sich die IT-Branche mehr und mehr von diesem monolithischen Gedanken und wendet sich einer modularen Anwendungsentwicklung zu. Anwendungen bestehen aus Microservices. Kleine Module einer Anwendung, die unabhängig voneinander bestimmte Funktionalitäten der Anwendung abdecken und im Zusammenspiel die Anwendung ausmachen. Diese Module können wiederum aufgrund ihrer Unabhängigkeit erweitert und angepasst werden, ohne dass die Funktionalität der Gesamtanwendung beeinträchtigt wird. Ein großer Vorteil gegenüber der bisherigen monolithischen Anwendung, welche für Anpassungen kurzzeitig komplett nicht zur Verfügung steht. Bei umfangreicheren Anwendungen kann die modulare Aufbauweise jedoch auch zu einer erhöhten Komplexität führen. Das Prinzip der Microservices macht sich Amazon Web Services ebenfalls zunutze. Web Services bieten eine schwach gekoppelte, asynchron und nachrichtenbasierte Kommunikation und bilden damit die Grundlage für ein weltweit verfügares Netz heterogener Ressourcen. [Red], [BKNT11]
Eine weitere Bewegung ist in der Zusammenarbeit von Softwareentwicklern und Systemadministratoren zu beobachten. Letztere stellen IT-Ressourcen bereit, damit die Anwendungen der Entwickler mit guter Performance laufen. Absprachen untereinander und notwendige Systemanpassungen führten zu längeren Releasezyklen einer Anwendung (z.B. monatlich). Durch dieses Vorgehen war es Entwicklern nicht möglich auf Anpassungswünsche der Kunden zeitnah zu reagieren. „DevOps“ änderte die obige Herangehensweise und brachte Entwickler und Administratoren näher zusammen. Im Fokus steht die Automatisierung von Tests und Bereitstellungsprozessen für Anwendungen und IT-Ressourcen. Es sind mehrere Deployments pro Tag möglich, wenn die Bereitstellung der Anwendung bestmöglich automatisch abläuft. Vom Erstellen der Anwendung über automatische Tests bis zur automatischen Installation in der Testumgebung bzw. schlussendlich dem Produktivsystem.
Infrastructure as Code ermöglicht die codebasierte Konfiguration benötigter Ressourcen sowie die Möglichkeit der Versionierung der Konfigurations-Code-Teile. Voraussetzung ist Hardware, die mit diesem Code auch umgehen kann. Amazon Web Services bietet eine solche Hardware und eine entsprechende Schnittstelle, die diesen Code entgegennehmen und zur Verarbeitung geben kann. Somit sind bei Bedarf in wenigen Minuten neue Ressourcen verfügbar. [WW16]
Durch dieses Vorgehen kann ein Unternehmen zeitnah auf Anforderungen reagieren, was sicherlich dazu beitragen kann, um sich am Markt zu behaupten.

Inhaltsverzeichnis

Abbildungsverzeichnis

Kapitel 1

Cloud Computing

Der Begriff Cloud Computing besitzt keine standardisierte Definition, weshalb er vielseitig interpretierbar ist. Eine oft zitierte Definition stammt vom Nationalen Institut für Standards und Technologie (NIST) [BKNT11] und beschreibt

- fünf wesentliche Eigenschaften
- drei verschiedene Dienstklassen
- vier unterschiedliche Betriebsmodelle

die für alle Cloud Computing Angebote gelten. Die fünf Eigenschaften sind:

1. Diensterbringung auf Anforderung
2. Netzwerkbasierter Zugang
3. Ressourcen-Pooling
4. Elastizität
5. Messbare Dienstqualität

Der Begriff „Cloud“ ist als Metapher zu verstehen, welche beschreibt, dass diverse Anbieter über das Internet (oder Intranet eines Unternehmens) ihre Dienste zur Verfügung stellen. Trotz unterschiedlicher Interpretationsmöglichkeiten gibt es grundlegende Ziele des Cloud Computing, die alle einen.

1. Cloud Computing beschreibt die dynamische Bereitstellung und Nutzung von IT-Ressourcen, Plattformen und Anwendungen als elektronisch verfügbare Dienste, unter der Nutzung von Virtualisierung und dem modernen Web.
2. Die bereitgestellten Dienste sollen durch mehrere Nutzer skalierbar verwendbar sein. Das bedeutet, sie sind sowohl auf Abruf als auch nach Bedarf verfügbar.

[BKNT11], [WW16]

Die IT-Ressourcen selbst sind für den Nutzer nicht direkt ersichtlich. Das Abstraktionslevel der Cloud variiert von virtueller Hardware bis hin zu komplexen verteilten Systemen.
Die Nutzung der Cloud-Dienste bietet einige Vorteile für den Anwender. Ein Vorteil ist die dynamische Skalierbarkeit der Dienste, weshalb sie von jungen Startups bis hin zu großen Unternehmen genutzt werden. Die sorgfältige Planung an zukünftig notwendigen IT-Ressourcen weicht dem On-Demand Ansatz. Es werden nur so viele Ressourcen bereitgestellt, wie auch benötigt werden. Abgerechnet wird nach dem 'pay-per-use'–Prinzip, welches ein weiterer Vorteil ist. Nach diesem Prinzip werden nur die tatsächlich genutzten Ressourcen abgerechnet. Nicht mehr und nicht weniger. Daneben ist es ein großer Vorteil, dass die IT-Ressourcen selbst in der Regel virtualisiert sind. Damit gibt es keine zu beachtenden systembedingten Abhängigkeiten. Ebenso entfallen mögliche Zwangsbedingungen für die Anwendungen des Nutzers [BKNT11], [WW16].
Cloud Computing bietet auf lange Sicht die Perspektive, das klassische Rechenzentrum zu einem IT-Servicezentrum umzuwandeln. Durch die immer spezialisierteren Dienste werden Mitarbeiter aus dem Management befähigt, eigenständig benötigte IT-Ressourcen zu kaufen. Dabei kann die mitunter aufwendige Abstimmung mit der internen IT-Abteilung deutlich geringer oder ganz ausfallen, was zu einer Veränderung in der Rolle der IT aber auch des Managements führen kann [BKNT11].

Kapitel 2

Amazon Web Services

Amazon Web Services (AWS) gehört zum amerikanischen Online-Versandhändler Amazon und beschreibt die seit 2006 entwickelten Infrastrukturdienstleistungen, welche für andere Unternehmen in einer öffentlichen Cloud angeboten werden [Amak]. Ihren Ursprung haben die Dienste in den Versuchen Amazons, Kosten einzusparen. Als Online-Versandhändler unterliegt das Unternehmen einem dynamischen Nutzungsaufkommen. Gerade zu saisonalen Ereignissen wie Weihnachten sind die Anfragen an die Webseiten und damit an die bereitgestellten IT-Ressourcen gut zehnmal höher als in der restlichen Zeit des Jahres. Damit die Ressourcen in dieser Zeit nicht ungenutzt bleiben und nur Geld kosten, entstand die Idee, die freien Kapazitäten an Dritte zu verkaufen. Dabei nutzt Amazon den Pooling-Effekt: Ungenutzte Ressourcen landen in einem gedachten Pool und können je nach Bedarf weitergenutzt werden. Hierdurch gelingt es Amazon ein für Nutzer sehr attraktives Modell zu schaffen, durch welches sie sich je nach Bedarf flexible Ressourcen und Kapazitäten zusammenstellen können [BKNT11].

AWS ist eine öffentliche Cloud (weitere Cloudtypen vgl. [WW16], [BKNT11]). Das bedeutet, sie wird durch eine Organisation verwaltet und steht der Öffentlichkeit zur Verfügung. Mit seinem Angebot deckt AWS folgende Klassifizierungen für Cloud Computing Dienste ab.

1. Infrastructure as a Service (IaaS)
   AWS bietet grundlegende Ressourcen wie Berechnung (computing), Speicherung (storage) und Netzwerk Kapazitäten (network capabilities). Ein Kerndienst hierfür ist Elastic Compute Cloud (EC2). Weitere sind Dynamo, S3, SimpleDB, CloudFront und SQS.

2. Platform as a Service (PaaS)
   AWS bietet beispielsweise über Elastic Beanstalk eine Plattform, über welche kundenspezifische Anwendungen in der Cloud bereitgestellt werden können.

3. Software as a Service (SaaS)
   SaaS kombiniert die vorhandene Infrastruktur mit der verfügbaren Software in der Cloud. Dies bietet AWS zum Beispiel mit dem Dienst Workspaces, welcher es ermöglicht, über einen virtuellen Desktop in der Cloud zu verfügen.

4. Humans as a Service (HuaaS)
   Hierbei geht es um Dienste, bei denen der Mensch als Ressource ins Spiel kommt,

da dieser der Maschine in einigen Bereichen deutlich überlegen ist. Zum Beispiel in Übersetzungs- oder Design-Aufgaben. Bei Amazon Mechanical Turk übernimmt eine Gruppe von Menschen Aufgaben unterschiedlicher Größe und Komplexität und erhält dafür je Kopf eine entsprechende Entlohnung. Damit entspricht der Dienst einem Marktplatz für CrowdSourcing-Angebote.

[WW16], [BKNT11]

Amazon dominiert den Markt im Bereich SaaS deutlich (45% Marktanteil), was sich auch in den Umsatzzahlen zeigt. Im dritten Quartal 2016 konnte AWS ein Umsatzplus von 55% auf 3,2 Milliarden US-Dollar für sich verbuchen. Das macht etwa 10% des Gesamtumsatzes von Amazon aus. [Bri]

Die, über AWS, bereitgestellten Dienste können grob in nachfolgende Gruppen unterteilt werden. Dabei beschränkt sich die Liste auf die wesentlichen der aktuell 70 verfügbaren Dienste [Sen], [Ram].

- Berechnungs-Dienste
  Beinhaltet die Bereitstellung von Rechenleistung und Speicherplatz z.B. Virtuelle Server.

- Speicher
  Hierbei wird das Sammeln, Persistieren und Archivieren von Daten betrachtet.

- Datenbank-Speicher
  Die genannten Dienste bieten gegenüber der „einfachen“ Speicheroption einige Vorteile, wenn es ums Managen strukturierter Daten geht. Es werden relationale und NoSQL-Systeme untersützt.

- Migrations-Dienste
  Diese Dienste unterstützen den Anwender zum Beispiel bei der Migration seiner Daten und Datenbanken in die AWS Cloud.

- Netzwerk und Content-Delivery
  Die Gruppe Netzwerk und Content Delivery beinhaltet Dienste, die es zum Beispiel ermöglichen private Netzwerke zu definieren, ein Domain Name System (DNS) für seine Anwendung einzurichten oder einen Load Balancer zu konfigurieren.

- Entwicklungs- und Administrations-Dienste
  Diese Dienste basieren auf den bereits oben genannten Diensten und sind hilfreich bei Themen wie Zugangsberechtigungen vergeben und einrichten, virtuelle Server monitoren und dem Bereitstellen von Anwendungen.

- Kommunikations-Dienste
  Diese Dienste bieten Lösungen für z.B. Queueing oder das Senden und Empfangen von E-Mails und Push-Notifications.

- Dienste für das Unternehmen
  Hiermit sind unahängige Dienste gemeint, die z.B. Videokonferenzen und das unternehmensweite Speichern und Teilen von Daten ermöglichen.

[WW16], [Amaf]

Amazon betreibt und verwaltet die über ein Netzwerk miteinander verbundene Hardware, welche für die korrekte Funktion der Anwendungsservices benötigt wird, sowie die benötigten Ressourcen, welche über eine Webanwendung bereitgestellt und genutzt werden. Mit seinem Angebot zählt AWS zu den bedeutendsten internationalen Angeboten im Cloud Computing.

2.1 Regionen und Availability Zones

AWS verfügt derzeit über 42 Availability Zones (Verfügbarkeitszonen) in 16 geografischen Regionen weltweit verteilt. Verschiedene Dienste, darunter EC2 und S3, sind in Regionen eingeordnet.

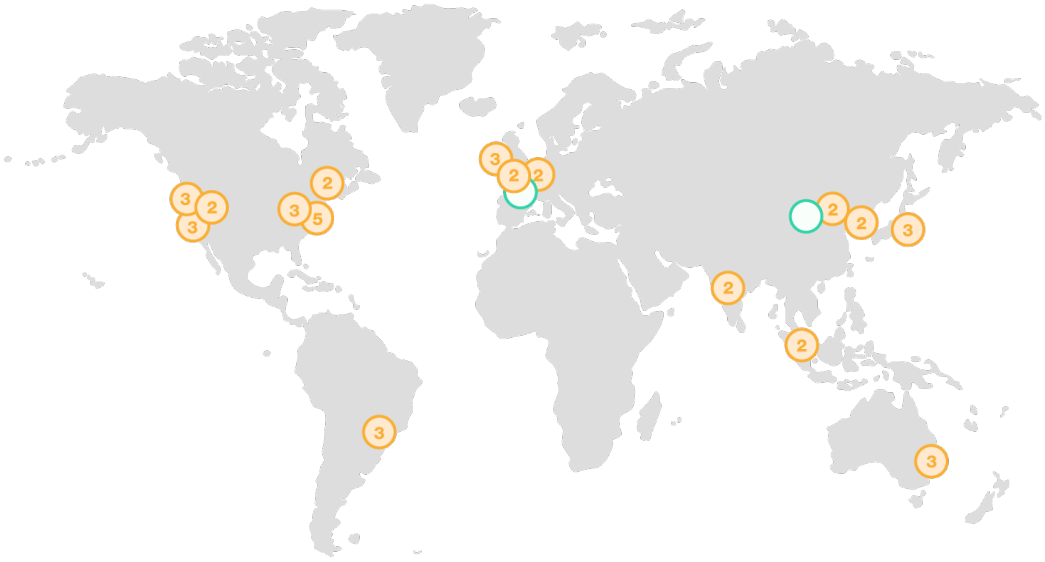

Abbildung 2.1: weltweite Infrastruktur (orange: Region mit x AZs, weiß/grün: geplante Region) [Amaj]

Eine Region entspricht einem physischen Ort auf der Welt, welcher mehrere Availability Zones (AZs) beherbergen kann. Bei einer AZ handelt es sich um ein oder mehrere unabhängige Rechenzentren, wobei jedes eine redundante Energieversorgung, Netzwerk und Konnektivität besitzt.

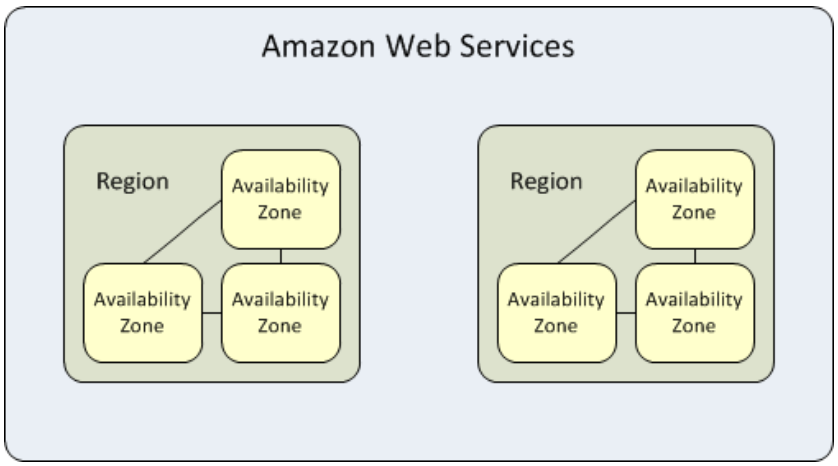

Abbildung 2.2: Regionen mit zugehörigen Availability Zones [Amal]

Dies erhöht die Ausfallsicherheit im Falle physischer Schäden z.B. Stürme und ist ein Alleinstellungsmerkmal gegenüber fast allen anderen Anbietern für technologische Infrastrukturen. Es ist daneben auch möglich, Daten zwischen mehreren AZs auszutauschen, um dem Ausfall der Anwendung bei einem Ausfall von AZs in einer Region vorzubeugen. Die AZs sind dafür mit schnellen, privaten Glasfasernetzwerken verbunden. Durch die Replikation der Daten über mehrere geografische Regionen hinweg, kann die Redundanz und Fehlertoleranz noch zusätzlich erhöht werden.

Die weltweit verteilten Datenzentren sind gerade für international agierende Unternehmen interessant. Je näher ein Datenzentrum dem Endkunden einer Anwendung ist, welche bei AWS gehostet wird, desto geringer fallen die Latenzzeiten aus. Darüberhinaus punktet Amazon mit diesem Konzept beim Thema Datenschutz. In Frankfurt am Main sind 2014 zwei AZs entstanden um Bedenken deutscher Unternehmen hinsichtlich Datenschutz auszuräumen. Kunden können definieren, dass ihre Daten ausschließlich in deutschen Rechenzentren gehalten und bearbeitet werden. Damit unterliegen die dort gehosteten Daten den deutschen Datenschutz-Vorgaben.

Aktuell sind fünf weitere Availability Zones und zwei weitere Regionen geplant.
[Red], [Amaj], [Amal]

2.2 Vorteile

Warum ein Einsatz der Amazon Web Services für Unternehmen jeder Größe sinnvoll sein kann, zeigen die Vorteile:

1. Kostenersparnis
   An oberster Stelle der Vorteile eines Einsatzes von IT-Ressourcen in der Cloud

gegenüber einem klassischen Rechenzentrum stehen die Kosten. Statt einer Aufstellung der IT-Ressourcen für die nächsten Jahre, können Ressourcen bei AWS nach Bedarf beansprucht werden. Ist die Zugriffsrate auf die Webseite gerade sehr hoch, können in wenigen Minuten weitere Server bereitgestellt werden, um die Gesamtlast aufzuteilen. Bezahlt wird dabei nach dem „Pay-per-use"-Prinzip. Es werden nur Kosten für Ressourcen erhoben, die auch tatsächlich verwendet wurden.

2. Hohe Innovations-Geschwindigkeit
   Im Jahr 2015 hat Amazon 722 neue Services und Features umgesetzt. 40% mehr als im Jahr zuvor. Wöchentlich werden neue Features und Verbesserungen veröffentlicht. Möglich wird das durch viele kleine Teams, die unabhängig voneinander arbeiten. Dabei stehen die Umsetzung der Wünsche der mittlerweile über eine Millionen Kunden im Vordergrund.

3. Weltweite Infrastruktur
   AWS Nutzer können auf ein Netzwerk weltweit verteilter Datenzentren zurückgreifen. Siehe Sektion 2.1

4. Arbeitserleichterung
   Auf Wunsch übernimmt AWS mit seinem enormen Dienstangebot notwendige Arbeiten der Nutzer, die sie sonst in Eigenregie erledigen bzw. verwalten müssten. Zum Beispiel Load Balancing zwischen den Servern oder das Aufsetzen eines E-Mail-Service.

5. Automatisierung
   Viele Dinge, die einen gewissen manuellen Aufwand und ab einer gewissen Komplexität auch eine kognitive Herausforderung bedeuten, können über Skripte automatisiert werden. Via Code können Instanzen aufgesetzt oder Container und Datenbanken bereitgestellt werden. Um Abhängigkeiten muss sich der Anwender dabei nicht sorgen, denn das übernimmt der Computer ganz automatisch. Die Infrastruktur kann beliebig flexibel definiert werden ohne dass manuelle Einstellungen auf der AWS Webseite getätigt werden müssen. Und das Beste daran ist, dass diese Skripte über ein Versionierungs-Tool versioniert werden können und im Fall eines Komplettabsturzes des Systems selbiges in sehr kurzer Zeit neu aufgesetzt werden kann. [1]

6. Skalierbarkeit
   Neben den Kosten ist Skalierbarkeit wohl das stärkste Argument für eine IT-Infrastruktur in der Cloud. Im Gegensatz zur „klassischen" IT, bei der weit in die Zukunft geplant werden und zukünftige Zugriffszahlen abgeschätzt werden müssen, können bei AWS Ressourcen innerhalb von Minuten hinzugeschaltet werden. Sollte der Ansturm vorbei sein und Ressourcen nicht mehr benötigt werden, können diese ebenso schnell wieder abgeschaltet werden. Das ist agil, umweltfreundlich

[1] vgl. Infrastructure as Code [WW16]

und spart auch noch Geld. Nicht nur zum Auffangen schwankender Zugriffszahlen, auch für das schnelle Bereitstellen von Testsystemen stellt eine hohe Skalierbarkeit der IT-Ressourcen einen großen (Geschwindigkeits-)Vorteil dar.

7. Ausfallsicherheit
   Die meisten angebotenen Dienste implizieren bereits eine Ausfallsicherheit, so dass sich der Anwender um dieses Problem auch nicht mehr kümmern muss.

8. Schnelle Anpassungsfähigkeit
   Durch die bereits erwähnte hohe Skalierbarkeit ist auch eine flinke Anpassung an sich ändernde Anforderungen z.B. durch den sich schnell wandelnden Markt gegeben. Entwicklungszyklen können deutlich kürzer ausfallen, da Testsysteme schneller bereitgestellt und Tests schneller durchgeführt werden können.

9. Die Menge macht's
   Je mehr Kunden AWS nutzen, desto günstiger werden die Dienste für den Einzelnen. Denn um die stetig wachsende Nutzerzahl mit gleichbleibendem Service bedienen zu können, müssen die unterliegenden Prozesse so optimal wie möglich sein.

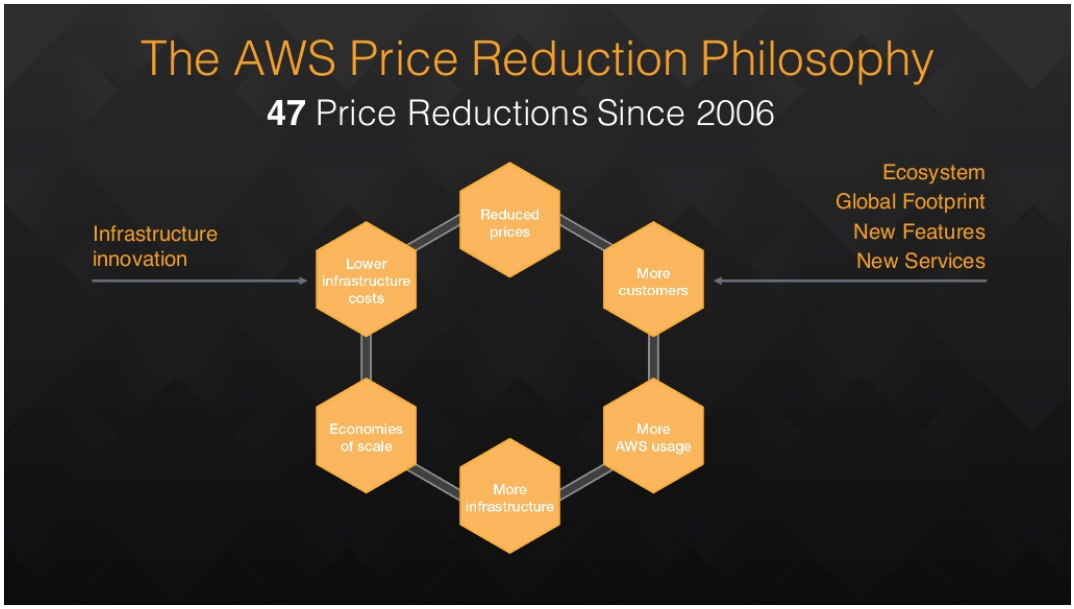

Abbildung 2.3: Preis-Reduktions-Philosophy [Ram]

10. Professionalität
    AWS setzt verschiedene Standards ein, um z.B. Zahlungssicherheit und Datensicherheit zu gewährleisten. [2]

[Ram], [WW16], [vVP11]

[2] weitere Informationen zu eingesetzten Standards siehe Kapitel 1.3.9 [WW16]

2.3 Kosten

Die Kosten für AWS sind abhängig von der Nutzung („pay-per-use“-Prinzip). Zum Beispiel wie viel Speicherplatz verbraucht wird oder wie viel die Laufzeit eines virtuellen Servers beträgt. Dieses Konzept ist gerade für kleine Unternehmen mit einem geringen Budget interessant, da sie weitaus bessere Möglichkeiten haben von Beginn an fehlertolerante Systeme aufzusetzen. Der Betrieb eines großen Servers kostet am Monatsende ebenso viel wie der Betrieb zweier kleinerer Server mit der selben Kapazität wie der große Server. Jedoch bietet die Infrastruktur mit zwei Servern die Möglichkeit redundanter Datenhaltung.
Für zukünftig geplante Installationen bietet AWS auf seiner Seite einen Preisrechner [Amae].
Seit 2010 bietet Amazon die Möglichkeit, in begrenztem Maße beliebte Dienste und Rechenleistung (750 Stunden) für 12 Monate kostenlos auszuprobieren. Dieses Programm nennt sich „Free Usage Tier“. [Wik], [WW16]

2.4 Sicherheit

Die Sicherheit der eigenen gehosteten Daten ist einer der wichtigsten Punkte, die Amazon gewährleisten muss. Um deutschen Datenschutz- und Datensicherheitsstandards gerecht werden zu können, wurden zwei Datenzentren in Frankfurt am Main eröffnet. Die Daten werden auf Wunsch der Kunden nicht auf amerikanischen Servern gespiegelt und es steht Entwicklern ausdrücklich frei, die hinterlegten Daten zu verschlüsseln. Allerdings war in der Vergangenheit auch immer wieder von Sicherheitslücken die Rede und auch die Auftragsvergabe einer public cloud der CIA an Amazon wurde hinsichtlich des Datenschutzes kritisch beurteilt [Wen], [Ber].

Kapitel 3

AWS am Beispiel einer Sharing Energy Plattform

Nachfolgend möchte ich anhand eines Beispiels aus dem Energiemarkt einige Dienste von AWS näher betrachten und ihre Wirkungsweise im Zusammenhang darstellen.

3.1 Das Fallbeispiel

Bei LichtBlick, Deutschlands größtem unabhängigen Ökostromanbieter, arbeitet aktuell ein Team an einer neuen Plattform-Idee. Grundsätzlich soll es möglich sein, dass sich Personen gegenseitig ihren selbst erzeugten Ökostrom verkaufen, ohne dass ein Stromhändler dazwischen hängt. Ähnliche Produkte gibt es bereits auf dem niederländischen und dem australischen Markt, jedoch noch nicht in Deutschland. Grund hierfür sind diverse Regularien, die es Besitzern von Wind-, Wasserkraft-, Photovoltaik- und Kraft-Wärme-Kopplungs-Anlagen erschweren, ihren Strom direkt zum Verkauf an andere Personen anzubieten. Da die Entwicklung in diesem Bereich noch recht unklar ist, dennoch erste Ideen auf dem deutschen Markt vertestet werden wollen, entscheidet sich das Team für eine agile und schlanke Herangehensweise. Daher fiel die Wahl bei der Frage nach der erforderlichen IT-Infrastruktur auf Amazon Web Services.

3.2 IT-Infrastruktur

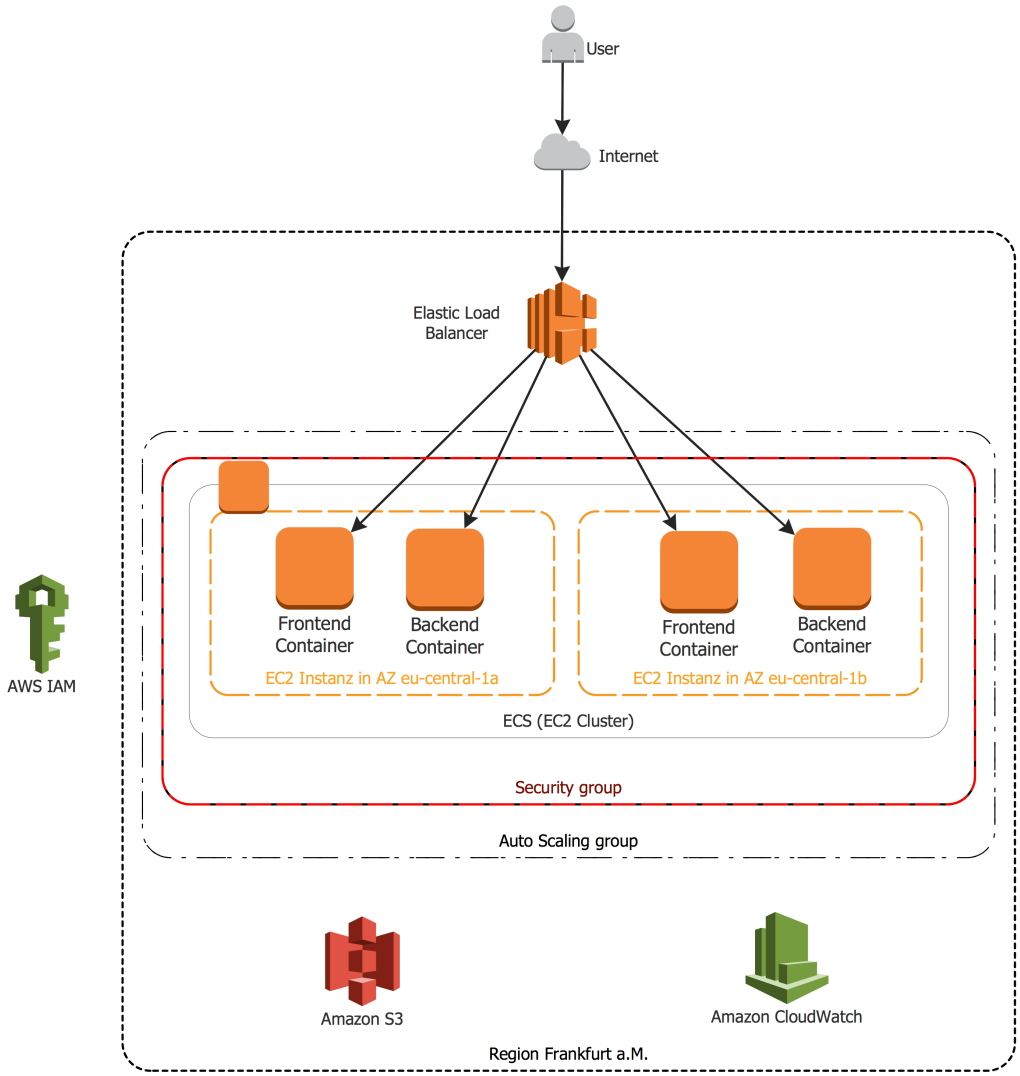

Abbildung 3.1: Infrastruktur für Sharing Energy Plattform

3.3 Virtuelle Server mit Elastic Compute Cloud

Jede Anwendung im Web benötigt einen Server, auf dem sie ausgeführt werden kann. Der Webservice Elastic Compute Cloud (EC2) bietet virtuelle Server und damit Rechenleistung in der Cloud, welche je nach Bedarf anpassbar sind. [Amaa] Der Entwickler hat die Möglichkeit, Instanzen von Servern innerhalb weniger Minuten aufzusetzen. Dabei kann er wählen, auf welchem Betriebssystem die Instanz basieren (Windows oder Linux)

und welche Ressourcenkonfiguration die bereitgestellte Instanz haben soll. [Amaa]

Eine Instanz wird aus einem Amazon Machine Image (AMI) erzeugt, welches das Betriebssystem und die darauf installierte Software vorgibt. Amazon bietet verschiedene vorkonfigurierte Images basierend auf 64-bit Windows oder verschiedenen 64-bit UNIX-Derivaten. Dem Nutzer stehen jedoch auch Images von Drittanbietern wie IBM oder Oracle zur Verfügung. Darüberhinaus besteht die Möglichkeit, sein eigenes Image zu entwickeln. Einzige Bedingung ist, dass es auf XEN Hypervisor[1] ausgeführt werden kann. [BKNT11], [vVPG13]

Die Ressourcenkonfiguration kann sich aufgrund der Leistungsfähigkeit des Prozessors und der Arbeitsspeicher- und Plattengröße unterscheiden. Die Instanzgrößen variieren von einer kleinen Instanz mit 1,7GB Arbeitsspeicher und 1 EC2 Recheneinheit bis hin zu einer High-CPU Extra Large Instanz mit 7GB Arbeitsspeicher und 20 EC2 Recheneinheiten. [vVPG13]

Darüberhinaus kann der Nutzer entscheiden, innerhalb welcher Availability Zone die Instanz bereitgestellt wird und ob die Instanzen zusätzlich mit persistentem Speicherplatz versehen werden sollen. Dabei hat er die Wahl zwischen zwei Varianten.

1. S3-gesichert
   Diese Instanztypen können nur neugestartet oder gelöscht werden, da sie einem flüchtigen Speicher entsprechen. Das root device ist Teil der Instanz selbst.

2. EBS-gesichert
   Dieser Typ wird bevorzugt ausgewählt, da Instanzen auch gestoppt und gestartet werden können. Dies wird möglich, da sich die root/boot disk auf einem separaten Elastic Block Store Volume (EBS) befindet, welches unabhängig der Instanz besteht.

[vVPG13], [Amaa]

Die Instanzen werden in einer virtual private cloud (siehe 3.7) bereitgestellt. Einem logisch isoliertem Netzwerk, welches verschiedene Netzwerk-Sicherheitsmechanismen mitbringt. Zum Beispiel gehört jede Instanz zu einer oder mehreren Security Groups, um sie gegenüber der Außenwelt abzuschirmen. Angesprochen werden die Instanzen über freigegebene Ports oder IP-Masken.
Weiterhin erhalten Instanzen eine öffentliche IP, die sich jedoch mit jeder Bereitstellung der Instanz (z.B. Neustart) ändert. Daher bietet Amazon die Möglichkeit, die öffentliche IPv4-Adresse[2] durch eine Elastic IP Adresse auszutauschen. Elastic IP bietet den

[1] https://www.xenproject.org/
[2] IPv6-Adressen werden derzeit nicht unterstützt

Vorteil, dass sie konstant bestehen bleibt. Sollte eine Instanz ausfallen, kann die Elastic IP Adresse sehr schnell einer gesunden Instanz zugewiesen werden. [vVPG13], [Amah]

In unserem Beispiel verwenden wir aus Gründen der Ausfallsicherheit zwei EC2-Instanzen (Typ: EBS-gesichert) innerhalb eines Clusters, wobei jede zwei identische Services enthält. Den Backend- und den Frontend-Service (siehe 3.2). Die Clusterung dient der logischen Gruppierung der Instanzen, um geeignete Zugriffsparameter setzen zu können.
Jede Instanz liegt in einer anderen Availability Zone in der Region Frankfurt am Main (eu-central-1) und kann auf 4GB Arbeitsspeicher zurückgreifen (Typ: t2.medium). Unterliegend wird ein Linux Image verwendet.
Aufgesetzt werden die Instanzen über die AWS API anhand von vordefinierten Konfigurationen, die über TerraForm[3]-an AWS gesendet werden. Die hierdurch entstehenden Vorteile wurden in Kapitel 2.2 behandelt. Weitere Möglichkeiten sind

1. Amazon Management Konsole
2. Software Development Kits (SDKs)
3. Command Line Interface

[WW16]

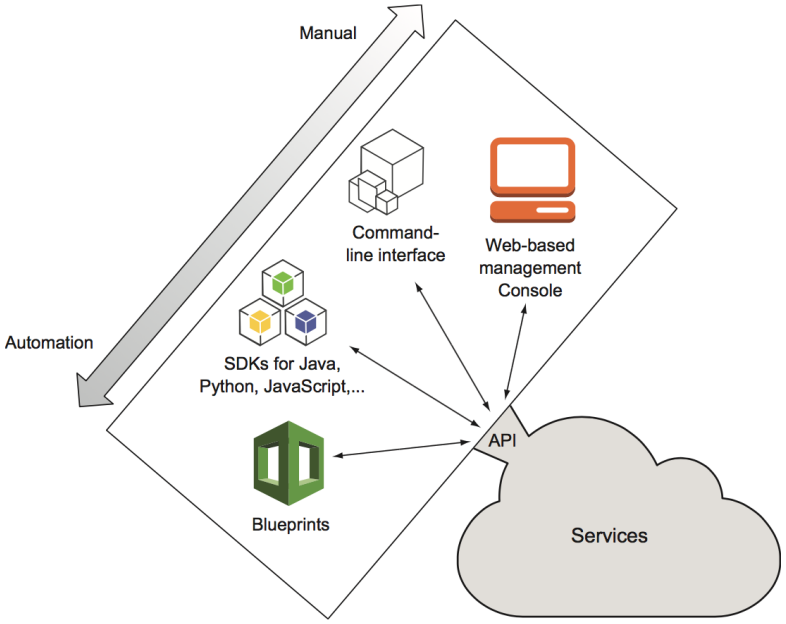

Abbildung 3.2: Interaktionsmöglichkeiten mit der AWS API [WW16]

[3]https://www.terraform.io/intro/

3.4 EC2-Container Service

Da unsere Services innerhalb von Docker-Containern laufen, nutzen wir den von Amazon angebotenen EC2-Container Service (ECS). Dieser Dienst wurde speziell für die Unterstützung von Docker-Containern auf EC2-Instanzen entwickelt und erlaubt neben dem Betrieb auch die Verwaltung der Anwendungen auf einem verwalteten EC2-Instanz-Cluster. Der Betrieb einer eigenen Cluster-Managementinfrastruktur entfällt. Funktionen wie Sicherheitsgruppen, Elastic Load Balancing, EBS-Volumes und IAM-Rollen stehen daneben ebenfalls zur Verfügung [Amab].

3.5 Auto-Scaling

Beide Instanzen laufen innerhalb einer Security Group, welche wiederum in einer Auto-Scaling Group liegt. Letztere ermöglicht eine elastische Instanzgruppe, deren bereitgestellte Ressourcen sich je nach Bedarf anpassen. Das Auto-Scaling basiert dabei auf CloudWatch-Metriken wie z.B. CPU-Auslastung. Beispielsweise könnte eine Metrik besagen, dass zwei neue Instanzen gestartet werden sollen, wenn die CPU Auslastung der bestehenden Instanz(en) für 5 Minuten bei 60% oder höher liegt. CloudWatch ist ein Service, der verschiedene Informationen der verwendeten IT-Ressourcen überwacht und misst. Von der Betrachtung einzelner Instanzen bis hin zur Überwachung aller Instanzen einer Region ist nahezu alles möglich. Es ist jedoch auch denkbar, dass die Ressourcenanpassung der Gruppe auf Amazons verteilten Queue-System SQS basiert. Dann würden sich die Ressourcen nach der Anzahl der Items in der Queue richten. [vVPG13], [vVP11]

In unserem Projekt haben wir noch keine konkreten Metriken definiert, die für das AutoScaling herangezogen werden sollen. Allerdings würden wir uns vermutlich im ersten Schritt auf die CPU-Auslastung stützen.

3.6 Load Balancing

Der gesamten Gruppe vorgesetzt ist ein Application Load Balancer (ALB), der die Instanzen vom öffentlichen Internet loslöst. Ein ALB ist eine von zwei Lastenverteilungsoptionen[4] des Services Elastic Load Balancing. Der Elastic Load Balancer (ELB) sitzt vor einer Gruppe von EC2-Instanzen und kann jede Art von TCP Traffic verteilen. Bei der Kommunikation mit den Instanzen nutzt der ELB HTTP und wandelt ggf. vorher eingegangenes HTTPS in HTTP um. Dieses Vorgehen reduziert die Belastung der Instanzen und erhöht damit die Fehlertoleranz. Bei der Verteilung des Traffics achtet der

[4] neben Classic Load Balancer;
vgl. Funktionen https://aws.amazon.com/de/elasticloadbalancing/classicloadbalancer/faqs/

ELB auf eine gleichmäßige Verteilung auf alle Availability Zonen und in jeder AZ auf die gleichmäßige Aufteilung auf die vorhandenen Instanzen. Dies erhöht die Sicherheit und reduziert die Komplexität der Infrastruktur. Dem ELB kann ein Hostname zugeordnet werden, worüber dieser direkt angesprochen werden kann.
Wird der ELB zusammen mit einer Auto-Scaling Group betrieben, kann dieser je nach Bedarf automatisch Instanzen registrieren und bei Nicht-Gebrauch auch wieder abmelden. Eine manuelle Zuordnung der IPs zu den jeweiligen Instanzen ist damit nicht mehr nötig. Bei der Abmeldung einer bestehenden Instanz und nachfolgender Neu-Registrierung einer neuen Instanz kann es unter Umständen kurzzeitig dazu kommen, dass der Dienst nicht zur Verfügung steht. Dieses Problem ist bekannt und entsteht dadurch, dass die alte Instanz erst beendet wird, ehe die neue Instanz hochgefahren werden kann. [vVPG13], [Amai]

3.6.1 Application Load Balancer

Bei dem zuvor erwähnten ALB, welchen wir im Einsatz haben, verteilt sich der eingehende Traffic auf erweiterte Anwendungsinformationen, die die Inhalte der Anfrage enthalten. Damit eignet sich dieser Typ besonders für Anwendungen mit der Notwendigkeit erweiterter Routing-Funktionen, Microservices und Container-basierten Architekturen. Ein ALB ermöglicht die Lastenverteilung auf mehrere Ports einer EC2-Instanz, aber auch die Umleitung auf mehrere Services. Letzteres wird durch Target Groups (siehe 3.6.2) und inhaltsbasiertes Routing ermöglicht, welches Anfragen basierend auf ihrem Inhalt an den entsprechenden Service weiterleitet.[5] [Amai], [Amad]

3.6.2 Target-Group

Eine Target Group ist eine Gruppierung von Instanzen oder Microservices innerhalb der Instanzen, damit Anfragen durch den ALB (siehe 3.6.1) gezielt an den zuständigen Dienst geroutet werden können. Hierfür wird eine Target Group als eine Regel für einen Listener des Load Balancers definiert. Jede Target Group nutzt die Standard Health Check- Einstellungen eines Load Balancers. Auf Basis dessen kann der Load Balancer ausschließlich gesunde Targets ansprechen. [Amam]

In unserem Fallbeispiel gibt es zwei Target Groups. In beiden befinden sich beide verfügbaren Instanzen als Targets. Wie bereits in Kapitel 3.3 beschrieben, befinden sich in jeder Instanz jeweils zwei Services, weshalb die eigentliche Logik der Target Groups bei uns (noch) nicht korrekt greift. Richtiger wäre eine Target Group für alle Backend-Services und eine Target Group für alle Frontend-Services. Je nach Auslastung würde

[5] weitere Funktionen des ALBs können [Amad] entnommen werden

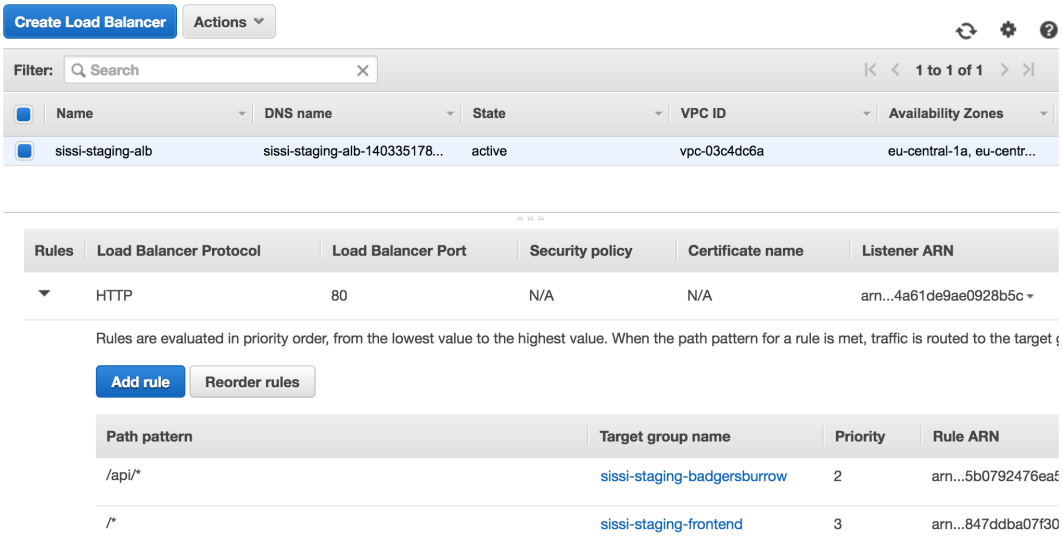

Abbildung 3.3: Screenshot aus AWS-Entwicklerkonsole mit Listener-Regeln des ALB für inhaltsbasiertes Routing

dann der eine über den anderen Service innerhalb der Target Group bevorzugt angesprochen. Unser Aufbau ist derzeit noch der Tatsache geschuldet, dass wir lediglich zwei Instanzen innerhalb unseres Clusters haben und der Scheduler die Container für einen Service bestmöglich gleichmäßig darauf verteilt.

3.7 Virtual Private Cloud

Der im vorherigen Kapitel erwähnte ALB ist der einzige Kontakt zur Außenwelt. Dies dient zum einen der Sicherheit und zum anderen dem Pragmatismus, da nicht für jeden einzelnen virtuellen Server ein DNS Name vergeben und verwaltet werden muss. Der ALB bekommt einen eindeutigen Hostnamen, über den der ALB und damit alle dahinterliegenden Services innerhalb der definierten Security Groups erreichbar sind. Die Security Groups bieten jedoch nicht ausreichend Kontrollmöglichkeiten für das sichere Handling der Anfragen. Zum Beispiel ist es in normalen Security Groups nicht möglich, eingehende Anfragen zu beschränken. Um eingehenden und ausgehenden Datenverkehr kontrollieren zu können, bietet Amazon die Virtual Private Cloud (VPC) an.
Der Einsatz der VPC ermöglicht es uns, die beiden EC2-Instanzen in einem privaten Sub-Netzwerk unterzubringen. Der ALB befindet sich im öffentlichen Netzwerk, welches von außen sichtbar ist. Anfragen gelangen per HTTPS über den Hostnamen zum ALB, wo sie umgewandelt und per HTTP an die jeweiligen Services in den EC2-Instanzen weitergeleitet werden. [vVPG13], [WW16], [Amag]

3.8 Simple Storage Service

Zur dauerhaften Speicherung erhobener Daten der Nutzer verwenden wir den neben EC2 zweiten Basisdienst für Webanwendungen: Amazons Simple Storage Service (S3). Ein virtueller Objekt-Speicher zur Speicherung einer unbegrenzten Datenmenge. Amazon proklamiert eine 99,999999999% Haltbarkeit und 99,99% Verfügbarkeit der Daten. Es soll nahezu unmöglich sein, dass Dateien kaputt oder verloren gehen. Amazons Chief Evangelist Jeff Barr beschreibt es so

> „If you store 10.000 objects with us, on average we may lose one of them every 10 million years or so.“

[vVP11]

Sichergestellt wird dies über ein Service Level Agreement. [vVP11]

Zu speichernde Objekte werden in buckets vorgehalten. Diese können in jeder Region erstellt werden und eine unbegrenzte Anzahl Objekte mit Größen zwischen 1Byte und 5TB enthalten. Zur Speicherung statischer Informationen von Webseiten oder mobilen Anwendungen ist S3 weniger geeignet. Hierfür kann der Nutzer auf CloudFront ausweichen. Einem Content Distribution Network, das dem Verteilen statischer, dynamischer und gestreamter Inhalte über die gesamte Welt dient. [vVPG13], [Amac]

3.9 Identity Access Management

Zur Einschränkung der Zugriffsmöglichkeiten für Nutzer auf die IT-Ressourcen in AWS und damit zur Erhöhung der Sicherheit, bietet AWS den Identity and Access Management Service (IAM). Dieser Service prüft jede Anfrage auf deren Zulässigkeit. Zugriffe können über policies Benutzer- oder Gruppenscharf organisiert werden. Die Authentifizierung der Nutzer gegenüber der Services erfolgt über die Anmeldedaten der Benutzer. IAM bietet neben der Zugriffsverwaltung auch

- Multi Factor Authentification (MFA) als zweiten Authentifizierungsschritt für bestimmte Operationen
- das Hinzufügen von Rollen zu EC2 Instanzen, wodurch die Festlegungen für die Rolle die Zugriffsmöglichkeiten auf die Instanz bestimmen

[vVPG13], [WW16]

Fazit

Amazon bietet mit seinem Angebot der Web Services einige gute Aspekte, die Unternehmen bei der Behauptung am Markt und der Verwaltung der benötigten IT-Ressourcen eine große Hilfe sein können. Die in dieser Arbeit aufgezeigten Punkte zeigen die Flexibilität und Anpassungsfähigkeit Amazons' für nahezu jeden Anwendungsfall der Softwareentwicklung. Vom kleinen Startup, welches gerade erst beginnt, sein Business Modell in die Tat umzusetzen, bis zum gestandenen Unternehmen mit mehreren bereitgestellten Softwareanwendungen kann jeder durch AWS in seiner Arbeit sinnvoll unterstützt werden und sich auf das Wesentliche seines Tuns konzentrieren - dem Kunden qualitativ hochwertige Software zur Verfügung stellen.
Sicherlich gibt es auch Anwendungsfälle, bei denen Eigenlösungen sinnvoller sind und eine Lösung über AWS zu deutlich mehr Kosten führen würde. Dies bleibt den Unternehmen jeweils abzuwägen. Daneben sollten sich Nutzer von AWS darüber bewusst sein, dass sie sich in eine gewisse Abhängigkeit zu den bereitgestellten Services von Amazon begeben, die unter Umständen nicht so einfach aufgelöst werden kann.

Literaturverzeichnis

[Amaa] Amazon. Amazon ec2. https://aws.amazon.com/de/ec2/. 11, 12

[Amab] Amazon. Amazon ecs. https://aws.amazon.com/de/ecs/. 14

[Amac] Amazon. Amazon s3. https://aws.amazon.com/s3/. 17

[Amad] Amazon. Application load balancer - details. https://aws.amazon.com/de/elasticloadbalancing/applicationloadbalancer/. 15

[Amae] Amazon. calculator. http://ws.amazon.com/calculator. 9

[Amaf] Amazon. Cloud products. https://aws.amazon.com/products/. 5

[Amag] Amazon. Default vpc and default subnets. http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/default-vpc.html. 16

[Amah] Amazon. Elastic ip addresses. http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ip-addresses-eip.html. 13

[Amai] Amazon. Elastic load balancer - cloud load balancer. https://aws.amazon.com/de/elasticloadbalancing/. 15

[Amaj] Amazon. Globale AWS infrastruktur. https://aws.amazon.com/de/about-aws/global-infrastructure/. iv, 5, 6

[Amak] Amazon. Informationen zu AWS. https://aws.amazon.com/de/about-aws/. 3

[Amal] Amazon. Regions and availability zones. http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-regions-availability-zones.html. iv, 6

[Amam] Amazon. Target groups for your application load balancers. http://docs.aws.amazon.com/elasticloadbalancing/latest/application/load-balancer-target-groups.html. 15

[Ber] Varinia Bernau. Eher heiter als wolkig. http://www.sueddeutsche.de/wirtschaft/geschaeft-zwischen-amazon-und-der-cia-eher-heiter-als-wolkig-1.1728299. 9

[BKNT11] Christian Baun, Marcel Kunze, Jens Nimis, and Stefan Tai. *Cloud Computing*. Springer Verlag Berlin Heidelberg, second edition, 2011. ii, 1, 2, 3, 4, 12

[Bri] Jörn Brien. Cloud-dienste: Aws deutlich marktführer vor google, microsoft und ibm. http://t3n.de/news/cloud-dienste-aws-marktfuehrer-762113/. 4

[Ram] Gladys Rama. Amazon ceo: Aws is a $10 billion business. http://de.slideshare.net/AmazonWebServices/london-re-play-pace-of-innovation-at-aws. iv, 4, 8

[Red] Bernd Reder. Amazon web services - viel cloud für wenig geld. http://www.computerwoche.de/a/amazon-web-services-viel-cloud-fuer-wenig-geld,3223095. ii, 6

[Sen] SendCheckIt. Amazon web services in plain english. https://www.expeditedssl.com/aws-in-plain-english. 4

[vVP11] Jurg van Vliet and Flavia Paganelli. *Programming Amazon EC2*. O'Reilly, 2011. 8, 14, 17

[vVPG13] Jurg van Vliet, Flavia Paganelli, and Jasper Geurtsen. *Resilience and Reliability on AWS*. O'Reilly, 2013. 12, 13, 14, 15, 16, 17

[Wen] Tobias Wendehost. Sicherheitslücken in der amazon-cloud entdeckt. http://www.computerwoche.de/a/sicherheitsluecken-in-der-amazon-cloud-entdeckt,2498479. 9

[Wik] Wikipedia. Amazon web services. https://de.wikipedia.org/wiki/Amazon_Web_Services. 9

[WW16] Andreas Wittig and Michael Wittig. *Amazon WebServices in Action*. Manning Publications Co., 2016. ii, iv, 2, 3, 4, 5, 7, 8, 9, 13, 16, 17