

Institut für Architektur von Anwendungssystemen
Universität Stuttgart
Universitätsstraße 38
D-70569 Stuttgart

Fachstudie Nr. 187

Vergleich der Energieeffizienz von Cloud Infrastrukturen

Thomas Michelbach, Oussama Riahi

Studiengang: Softwaretechnik

Prüfer/in: Prof. Dr. Frank Leymann
Betreuer/in: Dipl.-Inf. Sebastian Wagner

Beginn am: 01.05.2013

Beendet am: 01.11.2013

CR-Nummer: H.3.4, H.3.5

Kurzfassung

Im Vergleich zu traditionellen Rechenzentren bietet das Paradigma des Cloud-Computings ein großes Potential für eine ressourcenschonende Nutzung von IT-Infrastrukturen. Hierzu werden in dieser Fachstudie einige der populärsten Cloud-Computing Anbieter, aus dem Private- sowie Public-Cloud-Computing Bereich, hinsichtlich der Konzepte und Möglichkeiten zur Steigerung der Energieeffizienz gegenüber traditionellen Rechenzentren evaluiert. Die Bewertung erfolgt dabei unter Berücksichtigung von technischen sowie nicht-technischen Aspekten.

Inhalt

| | |
|--|-----------|
| 1. Einleitung | 1 |
| 2. Grundlagen | 2 |
| 2.1 Cloud-Computing..... | 2 |
| 2.2 Energieeffizienz | 5 |
| 2.3. Metriken | 7 |
| 2.3.1 Power Usage Effectiveness (PUE)..... | 7 |
| 2.3.2 Green Power Usage Effectiveness (GPUE)..... | 9 |
| 2.3.3 Datacenter infrastructure Efficiency (DCiE) | 9 |
| 2.3.4 Carbon Utilization Effectiveness Standard (CUE) | 10 |
| 3. Aspekte der Energieeffizienz | 11 |
| 3.1 Technische Aspekte | 11 |
| 3.1.1 Autoscaling Richtlinien | 11 |
| 3.1.2 Load Balancing Richtlinien..... | 12 |
| 3.2 Nicht-technische Aspekte..... | 12 |
| 3.2.1 Infrastruktur des Rechenzentrums..... | 12 |
| 3.2.2 Standort | 13 |
| 4. Herausforderungen bei der Ermittlung des Energieverbrauchs | 14 |
| 5. Untersuchung von Private- und Public-Cloud Anbieter..... | 18 |
| 5.1 Amazon Elastic-Compute Cloud (EC2) | 18 |
| 5.2 Google Compute Engine..... | 22 |
| 5.3 Microsoft Windows Azure | 25 |
| 5.4 GreenQloud Computeqloud..... | 28 |
| 5.5 VMware vCloud | 31 |
| 5.6 OpenStack | 33 |
| 6. Zusammenfassung und Ausblick | 38 |
| Literaturverzeichnis..... | 40 |

Abbildungsverzeichnis

| | |
|--|----|
| Abbildung 1: Zusammenhang zwischen Energieverbrauch und Auslastung | 6 |
| Abbildung 2: Mastodon C Dashboard | 16 |
| Abbildung 3: Amazon CloudWatch Monitoring Optionen | 20 |
| Abbildung 4: Definition eines Amazon CloudWatch Alarms | 20 |
| Abbildung 5: Windows Azure Autoscaling Optionen..... | 26 |
| Abbildung 6: GreenQloud Truly Green Dashboard | 29 |
| Abbildung 7: OpenStack Ceilometer | 35 |
| Abbildung 8: Kwapi RRD Visualisierung | 36 |

1. Einleitung

Im Gegensatz zu traditionellen Rechenzentren bietet das Paradigma des Cloud-Computings ein großes Potential für eine umweltschonende Nutzung von IT-Infrastrukturen. In Abhängigkeit des gewählten Cloud-Computing Anbieters fallen diese Potentiale unterschiedlich hoch aus und differieren stark je nach Private- und Public-Cloud-Anbieter.

In der folgenden Fachstudie werden die gebotenen Möglichkeiten und eingesetzten Konzepte zur Steigerung der Energieeffizienz ausgewählter Cloud-Computing Anbieter kritisch betrachtet.

Die vorliegende Fachstudie wurde zu diesem Zweck in mehrere Kapitel gegliedert. Zunächst erfolgt eine Einführung in die Grundlagen des Cloud-Computings sowie in die Begrifflichkeit der Energieeffizienz. Hierfür wird nach einer allgemeinen Einführung auch auf die wichtigsten Metriken zur Generierung von Kennzahlen eingegangen. In Kapitel 3 folgt eine Aufstellung von technischen als auch nicht-technischen Aspekten zur potentiellen Effizienzsteigerung von Cloud-Computing Infrastrukturen. Diese dient als Grundlage für die Betrachtung der populärsten Anbieter im Kapitel 5. Zuvor wird in Kapitel 4 auf gegenwärtige Hindernisse und potentielle Problemfelder bei der Ermittlung des Energieverbrauchs eingegangen. Abschließend folgt eine Zusammenfassung dieser Arbeit sowie ein Ausblick.

Bei den betrachteten Cloud-Computing Anbietern, die im Rahmen dieser Arbeit untersucht wurden, handelt es sich um:

- Amazon Elastic Compute Cloud (EC2)
- Google Compute Engine
- Microsoft Azure
- GreenCloud ComputeCloud
- VMware vCloud
- OpenStack

2. Grundlagen

Im nachfolgenden Kapitel werden die Grundlagen für ein tiefergehendes Verständnis des Cloud-Computing Begriffs geschaffen. Hierfür wird der Begriff des Cloud-Computings vom traditionellen Begriff eines Rechenzentrums abgegrenzt. Darüber hinaus wird der Begriff unter dem Gesichtspunkt der Energieeffizienz in einen Zusammenhang mit dem Paradigma der Green-IT gebracht.

2.1 Cloud-Computing

Mit steigender Popularität des Internets in den letzten Jahren steigt gleichermaßen auch die Nachfrage nach angemessenen Konzepten zur Bereitstellung von IT-Infrastrukturen. Als Alternative zum klassischen IT-Outsourcing und zum Eigenbetrieb hat sich Cloud-Computing als effizienter Ansatz zur dynamischen und bedarfsgerechten Bereitstellung von Rechenressourcen etabliert.

Im Gegensatz zu herkömmlichen Rechenzentren wird beim Konzept des Cloud-Computings die IT-Infrastruktur, je nach Bereitstellungsmethode, nicht mehr im eigenen Rechenzentrum betrieben, sondern in die Cloud zu einem Cloud-Computing Anbieter ausgelagert. Da es sich beim Cloud-Computing um einen jungen Begriff handelt, existiert bis dato jedoch keine allgemein gültige Definition oder Abgrenzung zu verwandten Begriffen. Eine weit verbreitete und damit allgemein anerkannte Begriffsdefinition stammt von der US-amerikanischen Standardisierungsstelle NIST (National Institute of Standards and Technology) [1]:

"Cloud-Computing ist ein Modell, das es erlaubt bei Bedarf, jederzeit und überall bequem über ein Netz auf einen geteilten Pool von konfigurierbaren Rechnerressourcen (z. B. Netze, Server, Speichersysteme, Anwendungen und Dienste) zuzugreifen, die schnell und mit minimalem Managementaufwand oder geringer Serviceprovider-Interaktion zur Verfügung gestellt werden können."

Bei der Definition des Cloud-Computings werden von der Standardisierungsstelle (NIST) folgende fünf Schlüsseleigenschaften als charakteristisch eingestuft:

On-demand Self Service: Die Bereitstellung von Ressourcen erfolgt vollautomatisch, weitestgehend ohne menschliche Interaktion mit dem Cloud-Computing Anbieter.

Broad Network Access: Die Anbindung an den Cloud-Computing Anbieter erfolgt mittels Standard-Mechanismen über eine Vielzahl heterogener Geräte.

Resource Pooling: Die Ressourcen eines Cloud-Computing Anbieters liegen in einem Pool vor, aus dem die Nutzer automatisch bedient werden. Die genaue Verortung der Ressourcen bleibt dem Nutzer vorenthalten, kann jedoch vertraglich vereinbart werden.

Rapid Elasticity: Skalierbarkeit bietet die Entkopplung von Nutzungs-schwankungen und Infrastrukturbeschränkungen. In vielen Fällen erfolgt diese Skalierung automatisch und bleibt dem Nutzer verborgen.

Measured Services: Die Nutzung von Ressourcen kann automatisch überwacht und optimiert werden. Der Nutzer hat zu jedem Zeitpunkt die Auslastung aller eingesetzten Ressourcen im Blick.

Die technische Grundlage für das Cloud-Computing bildet die sog. **Virtualisierung**. Sie bezeichnet die Erzeugung von (mehreren) virtuellen Instanzen (virtuellen Maschinen) auf einer einzigen physischen Hardware. Der Vorteil der Virtualisierung liegt also in einer Reduktion der physischen Systeme und somit darin, mehr Leistung, also eine größere Anzahl von Instanzen, auf geringerem Raum anbieten zu können.

Diese Einsparung von physischer Hardware und alle damit verbundenen Auswirkungen bilden das technische Fundament des Cloud-Computings und machen die Virtualisierung zu einer wichtigen Maßnahme zur Steigerung der Energieeffizienz gegenüber dem traditionellen Betrieb im Rechenzentrum.

Da sich die Anforderungen unterschiedlicher Kunden im Bezug auf Sicherheit und Erreichbarkeit stark unterscheiden, wird bei den Bereitstellungsmodellen (engl. Deployment-Models) zwischen Private-, Public-, Community- sowie Hybrid-Cloud unterschieden [2].

Private-Cloud Infrastrukturen sind für die exklusive Nutzung durch eine einzelne Organisation ausgelegt und können entsprechend durch die Organisation selbst bereitgestellt und organisiert werden. In diesem Fall spricht man von On-site Private-Cloud. Bei derartigen Cloud Strukturen stehen demzufolge, je nach Größe des eigenen Rechenzentrums, nur begrenzte Ressourcen zur Verfügung. Als Vorteil dieser Bereitstellungsmethode kann angeführt werden, dass die Organisation die volle Kontrolle über die eigenen Daten behält, da kein Dritter Zugang zu diesen besitzt. Lediglich die für den Betrieb notwendige Software wird dabei vom Cloud-Computing Anbieter bereitgestellt. Stellt der Anbieter oder ein anderer Dritter auch die Hardware-Ressourcen bereit, wird von outsourced bzw. off-site Private-Cloud gesprochen [2].

In einer **Public-Cloud** hingegen stellt ein Cloud-Computing Anbieter Zugang zu IT-Infrastrukturen für die Allgemeinheit zur Verfügung. Bei dieser in der Praxis häufigsten Form des Cloud-Computing erfolgt sowohl der Betrieb als auch die Bereitstellung durch einen Cloud-Computing Anbieter. Im Gegensatz zu einer On-site Private-Cloud stellt der Cloud-Computing Anbieter bei Nutzung eines Public-Cloud Angebots demnach die gesamte Infrastruktur zur Verfügung. Public-Cloud-Computing Anbieter erlauben ihren Kunden demzufolge, die benötigte IT-Infrastruktur zu mieten und diese auf Grundlage des tatsächlich anfallenden Grades der Nutzung zu finanzieren.

Dieses Finanzierungskonzept wird in Anlehnung an die Vertragsgestaltung als „Pay-As-You-Go“ [3] bezeichnet und soll den Kunden davor schützen, vor schnell zu hohe Investitionen in IT-Infrastruktur tätigen zu müssen. Als Nachteil gegenüber einer Private-Cloud lässt sich anführen, dass der Kunde den Einfluss auf die Sicherheit der Daten verliert und in die Hände des Anbieters abgibt. Im Vertragsverhältnis zwischen den Parteien soll die Datensicherheit über die sog. Service Level Agreements sichergestellt werden.

In einer **Community-Cloud** wird die Infrastruktur mehrerer Institutionen mit ähnlichem Interesse geteilt. Eine solche Cloud kann den o.g. Ausführungen entsprechend von einer dieser Institutionen (on-site) oder einem Dritten (off-site bzw. outsourced) betrieben werden [2].

Der Begriff der **Hybrid-Cloud** beschreibt den Betrieb mehrerer unabhängiger Cloud Infrastrukturen mittels einer standardisierten Schnittstelle, über die Anwendungen und Daten ausgetauscht werden können. Hierbei werden datenschutzkritische Anwendungen (und Daten) in einer Private-Cloud verarbeitet und anderen weniger sensitive Anwendungen öffentlich.

Durch eine Trennung der Sphären wird in diesem Modell versucht die Nachteile der jeweiligen Bereitstellungsmethode zu minimieren.

2.2 Energieeffizienz

Im Allgemeinen bezeichnet die Energieeffizienz eine Größe um den Energieaufwand zu bemessen, der für einen festgelegten Nutzen benötigt wird. Ein Ereignis kann als effizient bezeichnet werden, wenn ein vordefinierter Output mit einem möglichst geringen Energieaufwand erreicht wird [4]. Bezogen auf die IT-Infrastruktur eines Unternehmens definiert die Energieeffizienz demzufolge die benötigte Energie zu deren Bereitstellung in Relation zu anderen möglichen Formen der Realisierung. Die Quelle der Energie ist dabei zunächst nicht von Belang.

In der IT-Fachwelt hat sich in den letzten Jahren der Begriff der sog. „Green IT“ etabliert. Green-IT bezeichnet den Einsatz von modernen Informations- und Telekommunikationstechnologien unter Berücksichtigung einer Entlastung für die Umwelt. Die Begrifflichkeit Green IT vereint demzufolge Energieeffizienz und den Einsatz von erneuerbare Energien wie bspw. Solar- oder Windenergie. Per Definition ist folglich nicht nur die Quelle der gewählten Energie entscheidend sondern auch die Effizienz der eingesetzten IT-Infrastruktur.

Unter „schmutziger Energie“ hingegen versteht man den Einsatz von fossilen Brennstoffen mit niedrigem Wirkungsgrad verbunden mit einer Freisetzung großer Mengen von umweltschädlichem Kohlenstoffdioxid. In vielen Fällen lässt sich diese Freisetzung nicht vermeiden, jedoch durch die regelmäßige Modernisierung und Wartung der Rechenzentren zumindest reduzieren [5].

Unter der Prämisse, dass die Cloud-Computing Anbieter umweltbewusst arbeiten, lässt sich durch das Konzept des Cloud-Computings, in einem Rechenzentrum, die IT-Infrastruktur vieler Firmen effizient bündeln. Beim off-site Cloud-Computing muss nicht jedes Unternehmen eigene Hardware anschaffen und betreiben. Cloud-Computing kann somit eine positive Wirkung auf die Umwelt haben. Einsparungen sind bei der Produktion, der Distribution sowie auch bei der Entsorgung von Hardware möglich [6]. Ferner sorgt die Elastizität der IT-Ressourcen dafür, dass Rechenleistung je nach Auslastung zur Verfügung steht und keine Instanzen betrieben werden, die unnötig Energie verbrauchen

Durch die Virtualisierung lasten beim Cloud-Computing mehrere Instanzen einen physischen Host aus. Der Betrieb eines Servers mit mehreren Instanzen führt dabei zwar zu einem höheren Energieverbrauch, jedoch wird dies durch einen relativ höheren Performancegewinn kompensiert (siehe Abb. 1) [6].

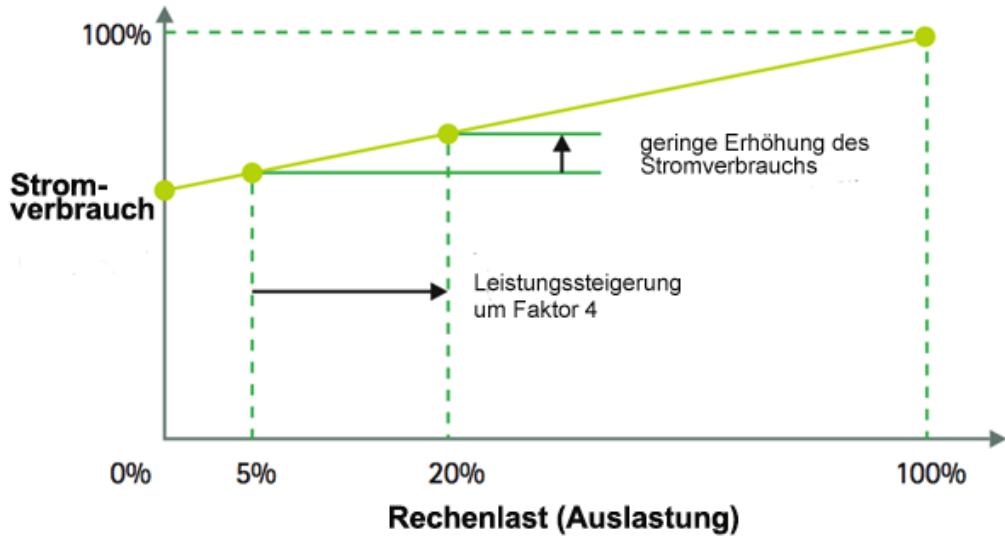


Abbildung 1: Zusammenhang zwischen Energieverbrauch und Auslastung

In der Unternehmenspraxis ist seit einiger Zeit ein Paradigmenwechsel zu Green-IT zu beobachten. Realisiert wird dieser Wechsel oftmals durch den Abschluss von Langzeitverträgen zum Einkauf von erneuerbaren Energien sowie durch den direkten Anschluss der Rechenzentren an energieeffiziente Quellen [7].

2.3. Metriken

Um den Begriff der Energieeffizienz messbar und vergleichbar zu machen, werden sog. Metriken zur Generierung von Kennzahlen genutzt. Da Rechenzentren äußerst energie- und kostenintensiv sind, existieren mehrere Ansätze zur Bewertung der Energieeffizienz. Bei den im folgenden Kapitel aufgeführten Metriken handelt es sich um die populärsten Vertreter im Zusammenhang mit der Betrachtung der Energieeffizienz von Rechenzentren.

2.3.1 Power Usage Effectiveness (PUE)

Der PUE-Wert, der durch das Green Grid Konsortium vorgeschlagen wurde und sich seither etabliert hat, beschreibt das Verhältnis des gesamten Energieverbrauchs eines Rechenzentrums zum ausschließlichen Energieverbrauch der IT-Geräte [8]:

$$PUE = \frac{\text{Gesamtenergieverbrauch des Rechenzentrums}}{\text{Energieverbrauch der IT-Geräte}}.$$

Unter dem Gesamtstromverbrauch wird in diesem Zusammenhang die elektrische Energie verstanden, die für den Betrieb des Rechenzentrums (Server, IT-Geräte, Belüftung usw.) insgesamt benötigt wird. Der Stromverbrauch der IT-Geräte repräsentiert die elektrische Energie, die nur für die Server und die IT-Geräte benötigt wird. Der PUE-Wert drückt folglich lediglich ein Verhältnis aus und trifft keinerlei Aussage über den tatsächlichen und absoluten Verbrauch der Infrastruktur.

Um die Vergleichbarkeit unterschiedlicher Messungen gewährleisten zu können, wurde der PUE durch ein Konsortium aus GreenGrid, ASHRAE, EnergyStar, Uptime Institute sowie weiteren Beteiligten in vier Kategorien unterteilt. Die Kategorien unterscheiden sich in der Auslegung der Messvorschriften sowie der Definition der Energiebeträge [9].

In einer Messung der **Kategorie 0** werden sowohl für den gesamten Stromverbrauch eines Rechenzentrums als auch für den Energieverbrauch der IT-Geräte die Jahreshöchstwerte in Betracht gezogen und mit 12 (Monaten) multipliziert. Die Messung erfolgt, per Definition, am USV (Uninterruptible Power Supply) Ausgang.

Messungen der **Kategorie 1** und **Kategorie 2** unterscheiden sich im Ort der Energiemessung der IT-Geräte. Während Messungen der Kategorie 1 am Ausgang der USV vorgenommen werden, wird der Energieverbrauch der IT-Geräte in **Kategorie 2** am Ausgang der PDU (Power Distribution Unit) bestimmt. Diese Messungen der Kategorie 2 repräsentieren somit exaktere Messungen als die der Kategorie 1, da sämtliche Verluste bis zum Ausgang der PDU dem Gesamtverbrauch des Rechenzentrums zugeschrieben werden. Dieser Wert sowie der Gesamtverbrauch der IT-Geräte wird für diese beiden Kategorien und für Kategorie 3, jeweils über den Zeitraum eines Jahres betrachtet – es kommen also keine Spitzenwerte zum Einsatz.

Die exaktesten Messungen sind **der Kategorie 3** zuzuordnen. Der Energieverbrauch der IT-Geräte wird hierbei am Eingang der Geräte gemessen, somit werden sämtliche Verluste dem Gesamtenergieverbrauch des Rechenzentrums zugeschrieben.

Für alle Kategorien gilt: je geringer der PUE-Wert ist, desto effektiver ist das Rechenzentrum. Im Idealfall fließt die komplette eingesetzte Energie verlustfrei in die IT-Infrastruktur, der PUE wäre demnach 1.0. Laut einer Umfrage des Uptime Instituts aus dem Jahr 2013 nutzen 66% der Befragten den PUE als Messgröße. Er liegt bei den Befragten, eigenen Aussagen zufolge, durchschnittlich bei 1.65 [10].

Jedoch trifft selbst dieser Idealwert keinerlei Aussage über die Herkunft sowie die absolute Höhe des Verbrauchs der Energie. Dieses Defizit versucht der GPUE auszugleichen.

2.3.2 Green Power Usage Effectiveness (GPUE)

Die Green Power Usage Effectiveness (GPUE) Metrik wurde vom Cloud-Computing Anbieter GreenQloud (siehe Kapitel 5) als Erweiterung des PUE vorgeschlagen [11]. Sie soll Aufschluss darüber geben, woher die eingesetzte Energie stammt und wie effizient diese eingesetzt wird.

$$G = \sum(Gewichtung\ des\ Energieträgers + 1) * Prozentualer\ Anteil$$

G ist hierbei die gewichtete Summe der eingesetzten Energieträger. Als Gewichte wurden die g CO₂/hWh Werte aus der Metaanalyse von Benjamin K. Sovacool [12] in kg CO₂/kWh übernommen. Die Gewichtung der einzelnen Energieträger wird mit 1 addiert, damit das Resultat nicht kleiner gleich dem PUE ist. Der GPUE kann anschließend durch folgende Formel aus dem PUE berechnet werden:

$$GPUE = G * PUE$$

Der Idealwert für den GPUE, unter Berücksichtigung der vorliegenden g CO₂/kWh Werte aus der Metaanalyse von Benjamin K. Sovacool, liegt bei 1.009. Dies setzt einen Idealwert für den PUE von 1.0 sowie die alleinige Nutzung von regenerativer Windenergie voraus.

2.3.3 Datacenter infrastructure Efficiency (DCiE)

Der DCiE-Wert hingegen bewertet dagegen die Energieeffizienz des Rechenzentrums, also den prozentualen Wirkungsgrad der im Rechenzentrum eingesetzten Energie [8]. Als Kehrwert des PUE berechnet er sich aus:

$$DCiE = \frac{1}{PUE}$$

Der ideale DCiE Wert ist also 100% und entspricht einem PUE von 1. In diesem Fall wäre das volle Potential der Rechenzentrum-Infrastruktur durch die IT-Komponenten ausgenutzt. Für den DCiE gelten die bereits im PUE erläuterten Einschränkungen hinsichtlich der Aussagefähigkeit.

2.3.4 Carbon Utilization Effectiveness Standard (CUE)

Der CUE-Wert gibt Auskunft über die CO₂-Bilanz eines Rechenzentrums. Im Gegensatz zu den bisher dargestellten Metriken handelt es sich um eine absolute Messgröße mit Einheit. Der CUE wird in Kilogramm Kohlenstoffdioxid pro Stunde (kg CO₂/kWh) angegeben und beschreibt den absoluten Kohlenstoffdioxidausstoß pro kWh [13]:

$$CUE = \frac{\text{Totale CO}_2\text{-Emission des Rechenzentrums}}{\text{IT-Geräte Energie}}$$

Hierbei liegt der Idealwert bei 0.0, was darauf hinweist, dass beim Betrieb des Rechenzentrums kein Kohlenstoffdioxid ausgestoßen wird.

Einschränkend soll nicht unerwähnt bleiben, dass alle dargestellten Metriken lediglich die Auswirkungen auf die Umwelt während des Betriebs beschreiben und damit nicht den gesamten Produktlebenszyklus der IT-Geräte abdecken. Nicht berücksichtigt sind somit bspw. die CO₂-Emission während der Produktion und Anlieferung der Infrastruktur sowie etwaige Auswirkungen durch den Einsatz von umweltbelastenden Materialien.

3. Aspekte der Energieeffizienz

Im Rahmen dieser Arbeit sollen mögliche technische und nicht-technische Aspekte unterschiedlicher Cloud-Computing Anbieter hinsichtlich der Energieeffizienz gesammelt und bewertet werden. Im folgenden Kapitel sollen hierfür zunächst charakteristische technische Aspekte einer Cloud-Computing Infrastruktur erläutert werden, die zu deren Energieeffizienz beitragen (können). Darüber hinaus werden auch mögliche nicht-technisch Kriterien zur Verbesserung der Energieeffizienz aufgeführt. Die Aufzählung ist dabei nicht als abschließend zu verstehen. Die genannten Aspekte repräsentieren jedoch die wesentlichsten Punkte, die bei der späteren Untersuchung der verschiedenen Anbieter als Rahmen dienen sollen.

3.1 Technische Aspekte

3.1.1 Autoscaling Richtlinien

Die Virtualisierung trennt die virtuellen Maschinen in Form der einzelnen Instanzen von der physischen Hardware. Die Elastizitätseigenschaften einer Cloud kommen erst durch den Einsatz von sog. Autoscaling-Richtlinien zu stande, die es erlauben die Anzahl der (virtuellen) Instanzen in einer Cloud-Computing Umgebung zu skalieren. In vielen Fällen geschieht dies vollautomatisch auf Grundlage der Beobachtung vorgegebener Parameter (CPU-Auslastung, Speicher-Auslastung). Bei steigender Auslastung einzelner Instanzen wird die Arbeit auf eine größere Anzahl von Instanzen verteilt. Bei sinkender Auslastung werden entsprechend Instanzen automatisch reduziert. Autoscaling ist demgemäß eine Form der Lastverteilung, die sich automatisch an die Auslastungsmuster anpasst.

Da die Bezahlung der Cloud-Computing Anbieter meist auf Grundlage der Nutzungsdauer erfolgt, führt diese automatische Skalierung der laufenden Instanzen zu einem enormen finanziellen Einsparungspotential. Ebenso kann die Energieeffizienz durch den Einsatz einer angemessenen Anzahl von Instanzen gesteigert werden.

Im Rahmen der Untersuchung im Kapitel 5 soll insbesondere darauf geachtet werden, ob und in welchem Maß die Möglichkeit besteht Autoscaling einzusetzen und welche Parameter hier als Grundlage dienen (können).

3.1.2 Load Balancing Richtlinien

Autoscaling sorgt wie beschrieben dafür, dass eine angemessene Anzahl von Instanzen zur Verfügung steht. Load Balancing Richtlinien stellen sicher, dass die einkommenden Anfragen gleichmäßig auf alle laufenden Instanzen verteilt werden.

Mit Hilfe von Load Balancing Richtlinien kann also sichergestellt werden, dass der Gesamtenergieverbrauch weitestgehend gleichmäßig auf die zur Verfügung stehenden Instanzen verteilt wird. Im Idealfall lassen sich dadurch folglich, unter Berücksichtigung des Energieverbrauchs in der Ruhephase sowie der variablen Energiekosten jeder weiteren Rechenoperation der jeweiligen Instanzen, entsprechend effiziente Strukturen realisieren. Der Energieverbrauch ist dabei abhängig davon, wie eine Anwendung eine Komponente auslastet.

Die größten dynamischen energieverbrauchenden Ressourcen eines Servers sind die CPU, der Speicher und die Festplatten. Bezuglich der verschiedenen Möglichkeiten, den Energieverbrauch zu bestimmen wird an dieser Stelle auf Kapitel 4 der vorliegenden Arbeit verwiesen.

Ob und mit welchen Möglichkeiten Load Balancing zur Verfügung steht wird für die unterschiedlichen Cloud-Computing Anbieter in Kapitel 5 evaluiert.

3.2 Nicht-technische Aspekte

3.2.1 Infrastruktur des Rechenzentrums

Die Infrastruktur eines Rechenzentrums ist sowohl für die Bereitstellung der notwendigen Energie, als auch für die Abfuhr der entstehenden Wärme verantwortlich. Ein zu hoher PUE-Wert deutet oft auf ein hohes Optimierungspotential diesbezüglich hin.

Durch Modernisierung sowie die Nutzung von Geothermie, solaren Kühlungen oder Absorptionskälteanlagen kann bspw. der Anteil der Kühlung am Gesamtenergieverbrauch des Rechenzentrums stark gesenkt werden. Die beschriebenen Möglichkeiten sind jedoch stark an den Standort des Rechenzentrums gebunden.

3.2.2 Standort

Die Energiequelle schlägt sich als dominierender Einflussfaktor in der CO₂-Bilanz eines Rechenzentrums nieder. Die Gesamtmenge der Kohlendioxid-Emissionen lässt sich auf die Zusammensetzung der Energieträger und deren Wirkungsgrad zurückführen.

Die U.S. Energy Information Administration (EIA) sammelt und analysiert ein breites Spektrum an Informationen zur Energiegewinnung und -nutzung. Diese Daten lassen sich geografisch filtern und enthalten länderspezifische Auswertungen zur Zusammensetzung der Energieträger, zum Energieverbrauch, dem Anteil erneuerbarer Energien sowie der (absoluten) Menge des CO₂-Ausstoßes.

Die zur Verfügung gestellten Daten können potentiell sowohl von Cloud-Computing Anbietern als auch von deren Kunden genutzt werden. Während die Anbieter die entsprechenden Daten bei der Standortwahl für neue Rechenzentren mit in das Kalkül einbeziehen können, stellen diese Informationen für entsprechend sensibilisierte Kunden möglicherweise ein Entscheidungskriterium bei der Wahl des Anbieters dar.

In Kapitel 5 wird der Standort des Rechenzentrums der Cloud-Computing Anbieter als Faktor betrachtet. Darüber hinaus wird untersucht, inwiefern dem Nutzer Möglichkeiten zur Bestimmung bzw. Wahl des Standorts eingeräumt werden.

4. Herausforderungen bei der Ermittlung des Energieverbrauchs

Um Cloud Infrastrukturen hinsichtlich ihrer Effizienz untersuchen zu können ist es unabdingbar, den Status quo des Energieverbrauchs ermitteln zu können. Im nachfolgenden Kapitel werden die Herausforderungen bei der Ermittlung des Energieverbrauchs in Cloud-Computing Umgebungen in der Praxis aufgezeigt, da diese Problemstellungen im Rahmen der Analyse der einzelnen Anbieter in Kapitel 5 der vorliegenden Arbeit zu berücksichtigen sind (sog. Energietransparenz) [14].

Der Energieverbrauch eines Servers lässt sich in zwei Phasen unterteilen. Die Phase, in der ein Server keine Anfragen erhält, wird als Idle-Phase (Ruhephase) bezeichnet. Der Energieverbrauch eines Servers ist in dieser Phase konstant und lässt sich einfach hardwareseitig bestimmen. Sobald Anfragen den Server erreichen, werden verschiedene Ressourcen (CPU, Speicher, Festplatten), je nach Anfrage und Anwendung, unterschiedlich stark beansprucht. Somit ist in der zweiten Phase ein variabler Energieverbrauch zu verzeichnen. Da die Anfragen ungleichmäßig verteilt sind, möglicherweise sogar parallel auftreten und von kurzer Dauer sind, sind hardwareseitige Messungen für diese zweite Phase nicht besonders effektiv [14].

Eine Alternative zu traditionellen Messungen bieten deshalb Approximationen des Energieverbrauchs. Bei der nachfolgend vorgestellten Methode wird versucht, den gesamten Energieverbrauch zu approximieren, indem der Auslastungszustand verschiedener Hardware Ressourcen wie CPU, Speicher oder Festplatte(n) im Zusammenhang mit dem tatsächlichen Energieverbrauch betrachtet wird [14]:

$$P = P_{idle} + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

Der gesamte Energieverbrauch P wird hierbei als Linearkombination der einzelnen Komponenten plus dem konstanten Ruheverbrauch P_{idle} modelliert. Zudem wird jede Komponente X_k mit dem zugehörigen Gewicht β_k multipliziert, welches angibt, in welchem Maß eine Komponente zum Gesamtenergieverbrauch beiträgt.

4. Herausforderungen bei der Ermittlung des Energieverbrauchs

Der Energieverbrauch der ersten Phase, also der Ruheverbrauch, kann im un- ausgelasteten Zustand mit Hilfe eines Messgeräts einfach ermittelt werden.

Während einer Trainingsphase, in der die maximale Auslastung isoliert auf verschiedene Ressourcen simuliert wird, erfolgt eine Messung des gesamten Energieverbrauchs P zusammen mit der Auslastung der jeweiligen Komponenten. Aus diesen Datenpunkten werden auf Grundlage einer Regressionsmethode (kleinste Quadrate) die Gewichte β_1 bis β_n ermittelt

Nach abgeschlossener Trainingsphase charakterisieren die Gewichte das Energieverbrauchsmodell. Das Messgerät ist für künftige Approximationen des Energieverbrauchs damit nicht mehr notwendig. Zu diesem Zeitpunkt lässt sich für bestimmte und bekannte Auslastungszustände der Hardware Ressourcen (CPU, Speicher und Festplatten) ein Energieverbrauch approximieren. Diese Modelle weisen einen durchschnittlichen Fehler von unter 1.25% auf und erreichen somit eine relativ hohe Genauigkeit [14].

Da es sich bei Cloud Infrastrukturen meist jedoch um virtualisierte Hardware handelt, sind für eine o.g. Trainingsphase notwendige Messungen des Energieverbrauchs nur für den Host möglich. Direkte Messungen des Energieverbrauchs sind ausschließlich für physische Hardware möglich, sobald mehrere virtuelle Maschinen auf einer physischen Maschine betrieben werden sind derartige Messungen auf Instanzebene, wie notwendig, nicht möglich.

Mit dem Joulemeter Tool [15] wird versucht die einzelnen Ressourcen der virtuellen Maschinen separat zu betrachten. Das beschriebene Vorgehen ist dasselbe wie oben dargestellt, jedoch werden die für jede virtuelle Maschine notwendigen Auslastungszustände softwareseitig getrennt betrachtet.

Diese Methode erreicht einen Fehlerbereich von 0.4W bis 2.4W [15].

Die Mehrzahl der im Rahmen dieser Fachstudie untersuchten Public-Cloud Anbieter tätigt keine Aussagen über die eingesetzte Hardware und erschwert damit Rückschlüsse auf den Energieverbrauch des Rechenzentrums. Zudem sind im off-site Betrieb keine hardwareseitigen Messungen des Energieverbrauchs möglich. Somit ist es nahezu unmöglich an belastbare Daten über den tatsächlichen Energieverbrauch des Hosts, einer einzelnen Instanz oder gar des Rechenzentrums zu gelangen.

4. Herausforderungen bei der Ermittlung des Energieverbrauchs

Diese mangelnde Transparenz schützt die Anbieter vor der Konkurrenz. Eine zu transparente Spezifikation der eingesetzten Hardware ermöglicht Einblicke auf Skalierbarkeit sowie auf die Kostenstruktur und ermöglicht einen einfachen Leistungsvergleich zwischen den Anbietern [16].

Nicht unerwähnt bleiben soll an dieser Stelle jedoch die Website des Unternehmens Mastodon C, einem Spezialisten für die Analyse von großen Daten. Mittels eines Dashboards werden auf deren Internetpräsenz [17] die CO₂-Emissionen der Rechenzentren populärer Public-Cloud Anbieter Amazon, Microsoft, Rackspace sowie GreenQloud in Form von Heatmaps dargestellt. Die Ergebnisse können der nachfolgenden Abbildung entnommen werden:

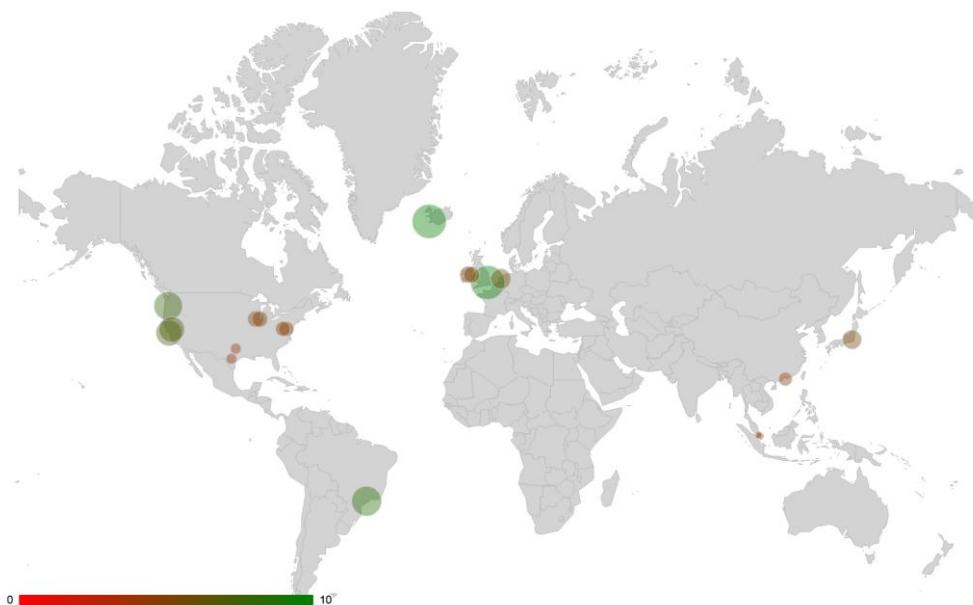


Abbildung 2: Mastodon C Dashboard

4. Herausforderungen bei der Ermittlung des Energieverbrauchs

Infolge der mangelnden Transparenz der Cloud-Computing Anbieter im Bezug auf den Energieverbrauch basiert diese Visualisierung jedoch auf Annahmen, die, zusammen mit öffentlich zugänglichen Daten, in einem mathematischen Modell, zusammengeführt werden. Bei den öffentlich zugänglichen Daten handelt es sich in diesem Zusammenhang um Daten des eGrid (Emissions & Generation Resource Integrated Database) die über AMEE bereitgestellt werden [18]. Darüber hinaus werden Annahmen zur eingesetzten Hardware getroffen (ein marktübliches 380 Watt Netzteil in einem x86 1U Server), um daraus den Energieverbrauch ableiten zu können. Der PUE wird in Abhängigkeit der aktuellen lokalen Temperaturen ebenfalls nur modelliert.

Die Energietransparenz ist ein wesentlicher Parameter bei der Betrachtung der Effizienz von Cloud-Computing Infrastrukturen. Insofern wurde dieser zusätzliche Aspekt als Kriterium in die in Kapitel 5 dargestellten Untersuchungen mit aufgenommen.

5. Untersuchung von Private- und Public-Cloud Anbieter

Im folgenden Kapitel werden die populärsten Cloud-Computing Anbieter hinsichtlich der im Kapitel 3 erläuterten Konzepte untersucht. Im vorangehenden Kapitel wurde zudem erörtert, dass es durch die inhärenten Eigenschaften des Cloud-Computings grundsätzlich an einer Möglichkeit mangelt den tatsächlichen Energiebedarf mit angemessenem Aufwand und Präzision zu bestimmen. Dennoch soll nachfolgend betrachtet werden, wie sich die verschiedenen Anbieter bezüglich der Energietransparenz verhalten. Damit einhergehend gilt es zu eruieren, welche Möglichkeiten existieren, die zur Verfügung gestellten Informationen zur Approximation des Verbrauchs und damit als Grundlage zur Steigerung der Energieeffizienz zu nutzen.

5.1 Amazon Elastic-Compute Cloud (EC2)

Amazon Elastic Compute Cloud (EC2) ist ein Service von Amazon Web Services (AWS) des Online-Unternehmens Amazon. Neben der EC2 existiert ein breites Spektrum an weiteren PaaS (Platform as a Service) Angeboten. Im Bereich des Public-Cloud-Computings ist Amazon gleichermaßen Pionier und Marktführer.

Standorte

Amazon verfügt weltweit über neun, geografisch getrennte, Standorte: Nord-Virginia, Oregon, Nordkalifornien, Irland, Singapur, Tokio, Sydney, São Paulo sowie einen US Standort mit dem Namen „GovCloud“ [19]. Zudem sind die neun Standorte (sog. „Regions“) in jeweils mehrere Verfügbarkeitszonen („Availability Zones“) unterteilt. Jede Availability Zone wird auf einer separaten physischen Infrastruktur ausgeführt. Diese Maßnahme führt zu geringer Latenzzeiten innerhalb einer Availability Zone und zur erhöhten Abschottung gegenüber anderen Availability Zonen. Dies bedeutet, dass Instanzen am selben Standort durch das Konzept der Availability Zones vor einem Komplettausfall des gesamten Standortes geschützt sind. Dem Anwender steht es, in Abhängigkeit der gewünschten Anforderung, offen in welcher Region Instanzen betrieben werden.

5. Untersuchung von Private- und Public-Cloud Anbieter

Die beiden Standorte Oregon sowie die „GovCloud“ werden 100% CO₂-neutral betrieben [20]. In Oregon wird dies durch den Einsatz von Wasserkraft entlang des Lake Wallula Reservoirs erreicht. Die Energiequelle der „GovCloud“ wird hingegen nicht näher durch Amazon spezifiziert [21]. Die Kosten für einen Betrieb in der umweltfreundlichen Region Oregon unterscheiden sich dabei nicht signifikant von den Kosten, die für den Betrieb in einer anderen Region anfallen und stimmen für die meisten Instanztypen mit den Kosten für den Betrieb in Nord-Virginia überein.

Energietransparenz

Der Dokumentation des Unternehmens lässt sich entnehmen, dass eine „Elastic Compute Unit“ (ECU), also die Abstraktion einer Prozessoreinheit im AWS-Umfeld, äquivalent leistungsfähig zu einem 1.0/1.2 GHz 2007 Opteron oder 2007 Xeon Prozessor ist [22]. Amazon nutzt diese Aussage jedoch nur als Richtlinie zum Leistungsvergleich und tätigt keine expliziten Aussagen über den Stromverbrauch von Instanzen. Auch bezüglich Informationen zu weiteren Umweltdaten hält sich das Unternehmen bedeckt

Autoscaling

Als Monitoring Instrument kommt bei Amazon der sog. CloudWatch Dienst zum Einsatz, mit dem unterschiedliche Daten im Bezug auf den Zustand einer EC2 Instanz gesammelt und verarbeitet werden können (Abb. 3.). Die verfügbaren Metriken beschränken sich auf die Auslastung der CPU, des Netzwerkverkehrs, der Festplattenoperationen sowie die Anzeige des System-Status. Die Nutzung von Amazon CloudWatch ist kostenlos, jedoch werden Metriken in der kostenfreien Variante nur im 5-Minuten-Takt überwacht. Für eine minütliche Überwachungsfrequenz, sowie die Beobachtung benutzerdefinierte Metriken, fallen pro Instanz zusätzliche monatliche Gebühren an.

5. Untersuchung von Private- und Public-Cloud Anbieter

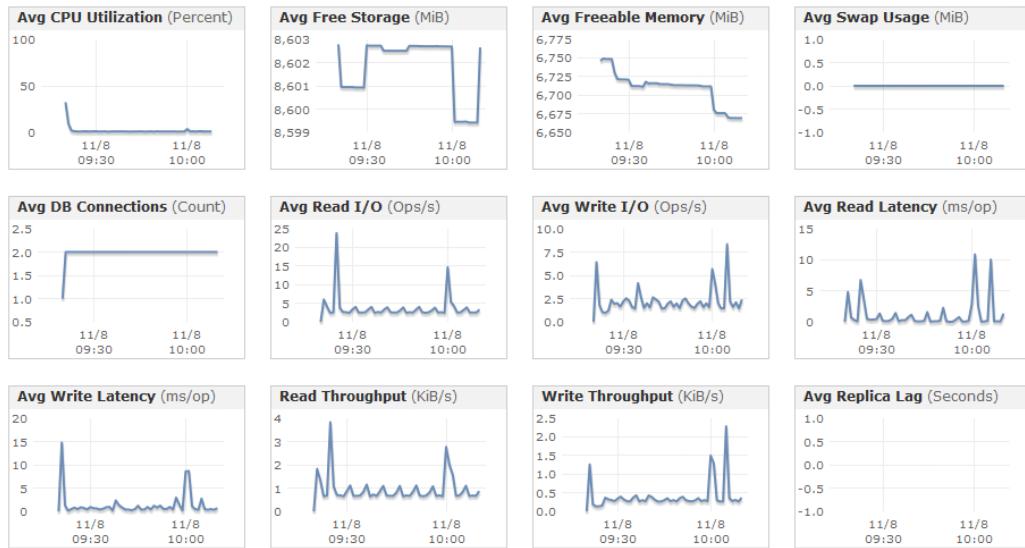


Abbildung 3: Amazon CloudWatch Monitoring Optionen

Die Autoscaling-Funktion nutzt den Amazon CloudWatch Service um eine Amazon EC2 Instanz hinsichtlich verschiedener Amazon CloudWatch Metriken (Abb. 3) zu beobachten. Die beobachteten Daten dienen sodann als Grundlage für mögliche Skalierungen. Somit lässt sich ein CloudWatch Alarm definieren (Abb. 4), der die CPU Auslastung einer Instanz beobachtet und bei zu hoher durchschnittlicher Auslastung (über einen definierten Zeitraum) neue Instanzen startet.

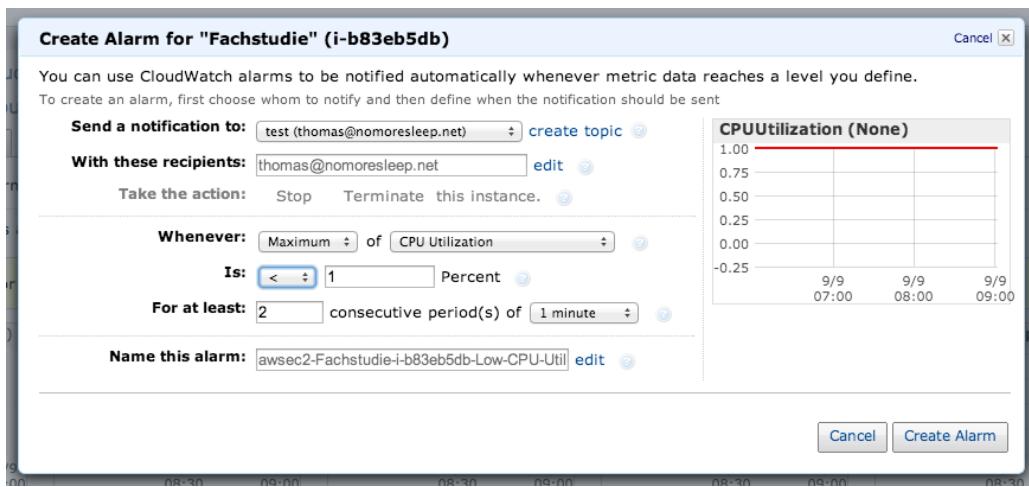


Abbildung 4: Definition eines Amazon CloudWatch Alarms

Load Balancing

Amazon stellt zum Load Balancing den sog. Elastic Load Balancer zur Verfügung. Dieser verteilt den einkommenden Verkehr entsprechend an zugewiesene Instanzen. Mit Elastic Load Balancing wird es möglich, fehlerhafte Instanzen zu erkennen und den einkommenden Verkehr nur auf fehlerfreie Instanzen zu verteilen. Ferner ist es möglich, die Zahl der zugewiesenen Instanzen hinter einem Elastic Load Balancer über eine Autoscaling Richtlinie zu steuern.

Fazit

Von Amazon werden keinerlei Aussagen über den Stromverbrauch getroffen, in Folge dessen sind diese auch nicht über eine CloudWatch Metrik zur Erstellung von Alarmen verfügbar. Jedoch lassen sich über die CloudWatch-API benutzerdefinierte Metriken mit Messwerten befüllen, auf die mit entsprechenden Maßnahmen reagiert werden kann. Wäre es also möglich den Energieverbrauch zu messen, oder zumindest über ein Verfahren anzunähern, könnten diese Werte an CloudWatch gemeldet werden und somit über einen CloudWatch Alarm die Anzahl der laufenden Instanzen an den aktuellen Energieverbrauch angepasst werden.

5.2 Google Compute Engine

Google Compute Engine ist ein IaaS (Infrastructure as a Service) Produkt von Google das sich derzeit noch im Vorschau-Modus [23] befindet. Neben der Google Compute Engine bietet Google u.a. die Google App Engine (GAE) an. Hierbei handelt es sich um eine PaaS Plattform zur Entwicklung und Bereitstellung von Web Anwendungen in den Programmiersprachen Python, Java, PHP oder Go.

Standorte

Google betreibt derzeit weltweit 13 Rechenzentren in Berkeley County, Council Bluffs, Douglas County, Quilicura, Mayes County, Lenoir und The Dalles. Für Asien existieren Rechenzentren in Hongkong, Singapur sowie Taiwan. In Europa befinden sich die Rechenzentren in Hamina, St. Ghislain sowie Dublin [24]. Im Rahmen der eingeschränkten Vorschau, in der sich die Google Compute Engine derzeit befindet, stehen jedoch ausschließlich US-Rechenzentren zur Auswahl [23].

Der Dokumentation [25] lässt sich nicht entnehmen, welche Rechenzentren in Zukunft explizit für die Google Compute Engine eingesetzt werden, und welche für weitere Google Angebote vorgehalten werden. Als Energiequelle des Rechenzentrums in Hamina kommt Windkraft zum Einsatz. Ferner wird Kühlwasser aus dem finnischen Meer für die Wärmetauscher genutzt. Das Rechenzentrum in St. Ghislain kommt, durch eine Kombination aus Verdampfung und Verrieselung von Wasser aus einem Industriekanal, ohne Kältemaschinen aus [26].

Energietransparenz

Google ist bemüht, seine negativen Auswirkungen auf die Umwelt durch die Nutzung erneuerbarer Energien sowie den Einsatz energiesparender Rechenzentren möglichst gering zu halten. Eigenen Aussagen zufolge liegt der durchschnittliche PUE für alle Google Rechenzentren bei 1.12 [27]. Google ist als erster Internet Service Provider mit externer Zertifizierung der Umweltnorm ISO 14001 sowie des Arbeitsschutzstandards OHSAS 18001 für alle US Rechenzentren ausgezeichnet worden. ISO 14001 repräsentiert ein Umweltstandard Framework zur Minimierung der negativen Umweltauswirkungen [28]. Trotz aller Bemühungen tätigt Google jedoch keine Aussagen über den Energieverbrauch seiner Google Compute Engine Instanzen.

Autoscaling

Es steht zum aktuellen Zeitpunkt kein natives Autoscaling zur Verfügung, jedoch besteht die Möglichkeit die Komponenten der Google Infrastruktur so zu orchestrieren, dass ein Autoscaling Effekt erzielt werden kann [29]. Hierfür wird eine Google App Engine Anwendung dazu genutzt, in regelmäßigen Abständen die Status Informationen (CPU Auslastung, Speicher Auslastung) der Google Compute Engine Instanzen zu sammeln. Die hierfür notwendigen Autoscaling-Richtlinien werden in einer Konfiguration auf der Google App Engine Anwendung hinterlegt. In periodischen Abständen werden sodann die akkumulierten Status Informationen mit diesen Richtlinien verglichen. Darauf aufbauend werden Entscheidungen zur Skalierung getroffen. Diese werden anschließend über die Compute Engine REST-API kommuniziert und ausgeführt.

Bei der Nutzung einer Google App Engine Anwendung können weitere Kosten anfallen. Der Java-Quellcode für die Anwendung steht jedoch kostenlos zur Verfügung [30].

Load Balancing

Die Lastverteilung erfolgt über Load Balancing-Richtlinien und der Weiterleitung von eingehendem Verkehr an eine Gruppe von Instanzen. Diese Gruppen werden TargetPools genannt und gruppieren mehrere Instanzen, an die der eingehende Verkehr über Lastverteilungsalgorithmen verteilt wird. Die dabei zum Einsatz kommenden Regeln werden als ForwardingRules bezeichnet.

Fazit

Durch den Einsatz erneuerbarer Energien in energiesparenden Rechenzentren gilt Google als wegweisendes Unternehmen bei der Minimierung negativer Umweltauswirkungen. Beim untersuchten Produkt, der Google Compute Engine, handelt es sich jedoch um ein junges Produkt mit eingeschränktem Funktionsumfang. Auf den ersten Blick macht sich dies u.a. bei den eingeschränkten Möglichkeiten zur Nutzung des Autoscalings bemerkbar. Jedoch deutet die hierfür vorgeschlagene Lösung auf ein großes Potential, durch die Nutzung und Kombination mächtiger APIs zur Verwaltung und Steuerung von Instanzen, hin.

Unter der Prämisse, dass Energieverbrauchswerte zur Verfügung stehen bzw. ein ausreichend akkurate Annäherungsmodell existiert, können diese Daten als Grundlage für Maßnahmen zur Steigerung der Energieeffizienz der Cloud Infrastruktur dienen.

5.3 Microsoft Windows Azure

Windows Azure ist die Public-Cloud-Computing Plattform von Microsoft. Ähnlich wie Amazon bietet Microsoft inzwischen weitere Webservice Angebote an.

Standorte

Microsoft betreibt derzeit acht Rechenzentren in Hong Kong, Singapur, Irland, Amsterdam, Illinois, Texas, Virginia sowie Kalifornien [31]. Der Dokumentation lässt sich jedoch nicht entnehmen, welche Rechenzentren speziell für Windows Azure eingesetzt werden und welche für weitere Webservice Angebote von Microsoft zur Verfügung stehen.

Energietransparenz

Microsoft äußert sich ebenfalls nicht zum Energieverbrauch seiner Instanzen. Das Unternehmen hat es sich jedoch zum Ziel gesetzt, die Infrastruktur ab dem (Geschäfts-) Jahr 2013 CO₂-neutral zu betreiben. Dies soll durch Investitionen in erneuerbare Energien kompensiert sowie durch den Einsatz von energieeffizienteren Infrastrukturen erreicht werden [32].

Autoscaling

Microsoft Azure verfügt über eine native Möglichkeit zur automatisches Skalierung von Instanzen. Diese ist innerhalb des Windows Azure Portals über eine grafische Oberfläche (Abb. 5) konfigurierbar. Zum aktuellen Zeitpunkt befindet sich dieses Feature in der Vorschau und unterstützt ausschließlich die beiden Metriken CPU (Auslastung) sowie Queue (Length). Bei letzterem handelt es sich um die Größe eines Stacks bei Verwendung des Windows Azure Queue Storage, einem Webservice Angebot zum asynchronen Nachrichtenversand im Warteschlangenprinzip.

5. Untersuchung von Private- und Public-Cloud Anbieter

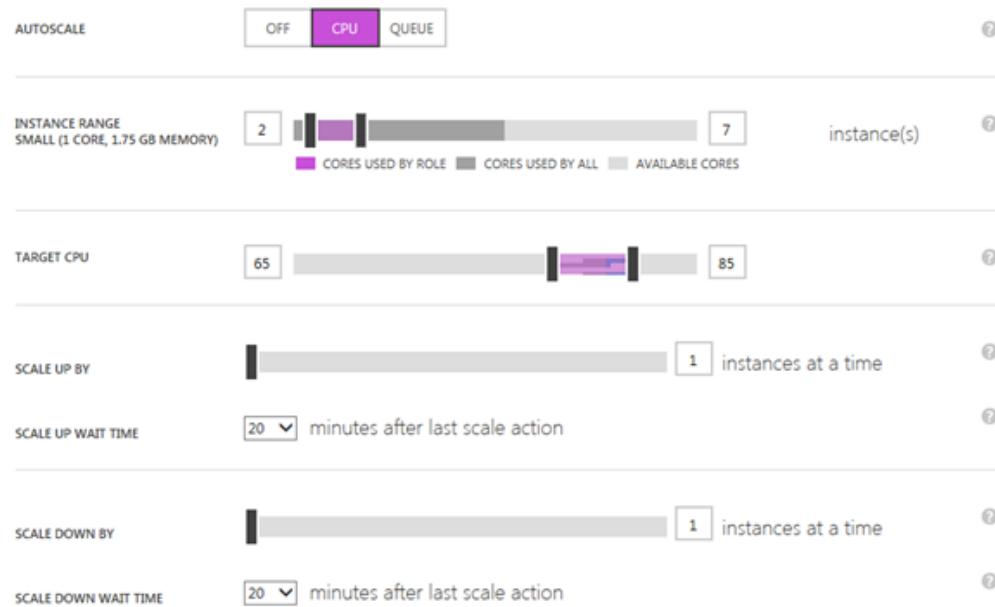


Abbildung 5: Windows Azure Autoscaling Optionen

Die Einführung des nativen Autoscaling macht die Nutzung des früher zum Einsatz kommenden Autoscaling Application-Block (WASABI) überflüssig, über den ein ähnliches Verhalten erreicht werden kann [33]. Dieser Application Block muss jedoch, im Vergleich zur grafischen Oberfläche, auf Grundlage von Regeln formuliert sowie über einen Windows Azure Worker separat installiert werden. Bei diesen Regeln wird zwischen Constraint- sowie Reactive Autoscaling Regeln unterschieden.

Constraint Autoscaling Regeln beschränken die Anzahl der gleichzeitig aktiven Instanzen auf ein Minimum bzw. Maximum. Reactive Autoscaling Regeln passen die Anzahl der Instanzen abhängig vom Status einer gewählten Metrik an.

Load Balancing

Mehrere Instanzen können in Windows Azure über einen sog. Endpoint mit Load Balancing Funktion zusammengefasst werden. Hierfür werden einkommende Anfragen eines speziellen Protokolls (TCP/UDP) über einen konfigurierbaren Port an die einzelnen Instanzen weitergeleitet. Als Algorithmus steht hierfür ausschließlich Round Robin zur Verfügung [34].

Fazit

Eine API zur Beobachtung eigener Metriken, wie bspw. der Amazon CloudWatch, die für das Autoscaling genutzt werden kann, steht derzeit bei Microsoft nicht zur Verfügung. Es ist jedoch denkbar, hierfür den Windows Azure Queue Storage zu nutzen, diesen mit Nachrichten zu füllen und über dessen Größe die automatische Skalierung über die GUI zu konfigurieren. Mit Hilfe einer XML Konfiguration können, bei Einsatz des Autoscaling Application-Blocks, alternativ Regeln angelegt werden, die auf Grundlage eigener Metriken das Skalierungsverhalten steuern.

5.4 GreenQloud ComputeQloud

Im Report „*How Clean is Your Cloud*“ von Greenpeace [35] wird GreenQloud als grüne Alternative zu Amazon Web Services aufgeführt. GreenQloud wurde im Jahr 2010 gegründet und ist somit, verglichen mit den weiteren Public-Cloud Anbietern, ein relativ junges Unternehmen. Der Sitz des Unternehmens sowie auch die betriebenen Rechenzentren liegen in Island. Das Angebot umfasst einen EC2 ähnlichen IaaS Dienst namens ComputeQloud sowie einen Data Storage mit dem Namen StorageQloud, ähnlich zum Amazon Simple Storage Service (S3). Das Angebot der GreenQloud ComputeQloud umfasst acht unterschiedlich leistungsfähige Instanztypen bei Linuxnutzung. Bei Nutzung von Windows als Betriebssystem stehen hingegen nur die sechs leistungsstärksten der Instanzen zur Verfügung. Im Vergleich zu anderen Cloud-Computing Anbietern fallen auch weniger leistungsfähige Instanztypen (bspw. mit 256 bzw. 512 MB RAM sowie 10 bzw. 20 GB Festplattenspeicher) in das Angebotsspektrum der GreenQloud ComputeQloud. Ebenfalls ist es, im Gegensatz zu anderen Cloud-Computing Anbietern, (noch) nicht möglich eigene Images zu nutzen.

Standorte

In Island wird Strom fast ausschließlich CO₂-neutral aus Erdwärme (Geothermie) sowie Wasserkraft gewonnen [36]. Zusätzlich zur umweltschonenden Gewinnung von Energie sorgt die arktische Kälte für eine optimale natürliche Kühlung der Server Infrastruktur. Die beiden Rechenzentren, in denen GreenQloud betrieben wird, sind Verne Global und Thor DC (Advania). Diese werden zu 100% durch erneuerbare Energien bewirtschaftet [37][38]. Thor DC (Advania) wirbt mit einem durchschnittlichen PUE von nur 1.16 [38], Verne Global von höchstens 1.2 [39]. Eigenen Aussagen zufolge arbeitet GreenQloud an einer ISO 14001 Zertifizierung [40].

Ähnlich der 100% CO₂-neutralen EC2 Region Oregon fallen im Vergleich zu anderen Cloud-Computing Anbietern keine nennenswerten Mehrkosten für den umweltschonenden Betrieb in Island an.

Energietransparenz

Im Gegensatz zu den anderen Public-Cloud Anbietern ermöglicht GreenQloud seinen Kunden den Zugriff auf Energieverbrauchswerte. Das Truly Green Statistics Dashboard (Abb. 6) zeigt die CO₂-Ersparnisse im Vergleich zum Betrieb im Heimatland an. Diese Werte aktualisieren sich zwar nicht in Echtzeit, beinhalten aber neben der CO₂-Ersparnis auch den totalen Energieverbrauch gemessen in kWh. Eigenen Aussagen zufolge basieren diese stündlich aktualisierten Energieverbrauchswerte jedoch nur auf Hochrechnungen und spiegeln nicht das Resultat tatsächlicher Messungen wieder.

Der Betreiber des Rechenzentrums stellt auf der eigenen Webseite ebenfalls ein Tool zur Berechnung der Energieersparnis sowie der Kosteneinsparung gegenüber amerikanischen Staaten zur Verfügung [41].

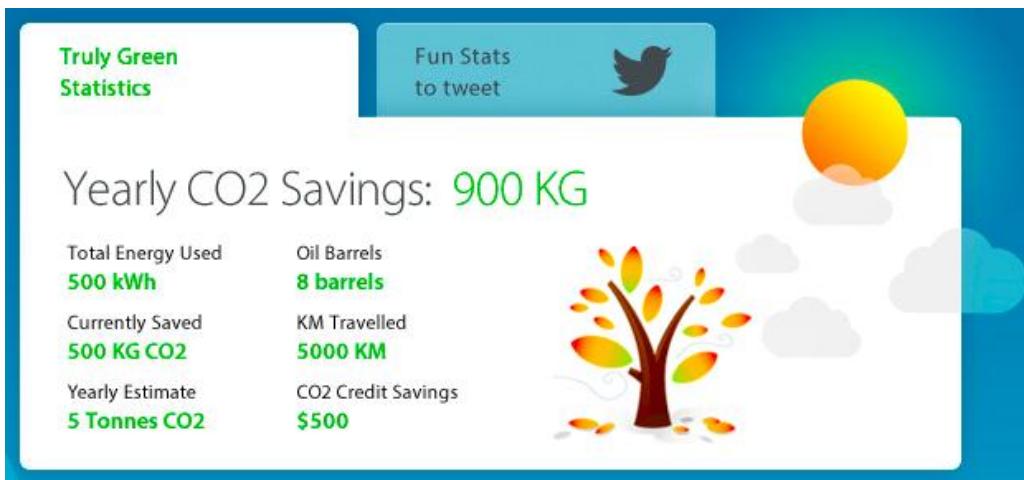


Abbildung 6: GreenQloud Truly Green Dashboard

Load Balancing

Über die GUI ist es möglich, mehrere Instanzen in einer Load Balancing Gruppe zusammenzufassen und so die Auslastung auf die einzelnen Maschinen zu verringern. Hierfür stehen drei unterschiedliche Algorithmen („Round Robin“, „Last Conn.“ und „Source“) zur Verfügung.

Autoscaling

Steht derzeit (noch) nicht zur Verfügung.

5. Untersuchung von Private- und Public-Cloud Anbieter

Fazit

Derzeit existiert leider keine API Schnittstelle um auf die Informationen des Truly Green Dashboards zuzugreifen, diese auszuwerten und entsprechend zu reagieren. Die existierenden API Schnittstellen für die ComputeQloud als auch für den Qloud Storage sind jedoch Amazon EC2 kompatibel. Dies ermöglicht eine unkomplizierte Migration der bestehenden IT-Infrastruktur in die Cloud-Computing Plattform von GreenQloud und damit eine enorme Einsparung an CO₂-Emissionen. Durch die Angabe des Standorts der ehemaligen EC2 Region (als Heimatland) im GreenQloud Benutzerkonto werden die Ersparnisse durch die Migration im Truly Green Dashboard bezogen auf den Standort approximiert.

Im Gegensatz zu anderen Cloud-Computing Anbietern hat GreenQloud einen klaren Standortvorteil und kann dadurch auf die Bereitstellung einer oder mehrerer spezieller umweltfreundlicher Regionen verzichten.

5.5 VMware vCloud

Bei den beiden nachfolgend vorgestellten Cloud-Computing Anbietern handelt es sich um Softwarelösungen zur Bereitstellung von Cloud-Infrastrukturen. Da die im Kapitel 3 vorgestellten Konzepte dadurch weitestgehend in den Verantwortungsbereich des Nutzers fallen, wird nachfolgend der Fokus auf den Komponenten der jeweiligen Lösung und die durch die Komponenten oder die Orchestrierung einzelner Komponenten gegebenen Möglichkeiten zur Steigerung der Energieeffizienz liegen.

Die vCloud ist eine Cloud-Computing Plattform von VMware, einer Softwarefirma die sich auf den Bereich der Virtualisierung spezialisiert hat. Es handelt sich um ein kommerzielles Produkt zum Betrieb einer Private- oder Public-Cloud auf Grundlage der VMware Virtualisierungstechnologie. vCloud wird in den drei Varianten „vCloud Suite Standard“, „vCloud Suite Advanced“ sowie „vCloud Suite Enterprise“ angeboten. Diese unterscheiden sich primär im Funktionsumfang. So stehen beispielsweise mächtige Komponenten zur Überwachung und automatisierten Bereitstellung erst in den Versionen „vCloud Suite Advanced“ bzw. „vCloud Suite Enterprise“ zur Verfügung.

Die vCloud vereint eine große Anzahl an Komponenten mit unterschiedlichen Funktionen die in ihrer Gesamtheit eine vCloud repräsentieren [42]. Die folgend vorgestellten Komponenten haben sich, im Rahmen der Untersuchung, als besonders interessant erwiesen.

Die **VMware vSphere** Komponente ist für die Virtualisierung der Hardware verantwortlich und bildet das Fundament einer vCloud. Es handelt sich, eigenen Angaben zufolge, hierbei um die weltweit führende Servervirtualisierungsplattform. Beim Einsatz von ESX(i) auf dem Hostsystem steht das Kommandozeilentool „esxtop“ zur Verfügung. Bei esxtop handelt es sich um eine spezielle Variante des top Befehls mit dem in Echtzeit Informationen zur Ressourcennutzung des ESXi Hypervisors geliefert werden können. In Abhängigkeit der BIOS Einstellungen steht hier die Option POWER zur Verfügung die den aktuellen CPU-Energieverbrauch (in Watt) angibt.

5. Untersuchung von Private- und Public-Cloud Anbieter

Die **vCloud Director** Komponente abstrahiert vSphere Ressourcen. Die Kommunikation mit der vCloud Installation erfolgt über die REST basierte **vCloud API** Komponente. Die **vCenter Orchestrator** Komponenten dient der Orchestrierung von operativen Abläufen, die mehrere Komponenten einer vCloud Installation involvieren. Die optionale **vCenter Chargeback** Komponenten ermöglicht Messungen des Ressourcenverbrauchs und ermöglicht eine Weiterberechnung von Kosten an Kunden.

Über die **vCenter Operations Management** Komponenten lassen sich Performance und verfügbare Kapazitäten der gesamten verwalteten Cloud-Computing Umgebung oder einzelner Instanzen ermitteln. Diese überwachten Metriken werden im Dashboard zusammengetragen und zusätzlich in die sogenannten Supermetriken Health, Risk und Efficiency organisiert.

Health spiegelt den aktuellen Zustand der vCloud Installation wieder und aggregiert die momentane Auslastung der CPU, des Speichers sowie die Anzahl der I/O Operationen der Festplatte und des Netzwerks.

Risk zeigt mögliche Risiken in der Zukunft auf in dem Ressourcen ermittelt werden, die beim aktuellen Grad der Nutzung in Zukunft zu Engpässen führen können.

Efficiency befasst sich, ähnlich wie Risk, mit Aussagen im Bezug auf die Zukunft, geht jedoch nicht auf Risiken ein sondern ermittelt Möglichkeiten für künftige Optimierungen.

Fazit

Da es sich bei der VMware vCloud um ein kommerzielles Produkt handelt fällt eine exakte Evaluierung des gesamten Funktionsumfangs ohne eine vorliegende Installation sehr schwer. Dadurch dass jedoch sowohl die Virtualisierungslösung als auch die Software zum Betrieb der Cloud Infrastruktur aus derselben Hand kommen, ergeben sich mächtige Möglichkeiten. Jedoch ist hierfür ein hoher Preis zu zahlen. Die Lizenzierung erfolgt pro Prozessor und versteht sich exklusive Support. Eine kostenfreie Alternative hierzu bietet das nachfolgend vorgestellte Open Source Projekt OpenStack.

5.6 OpenStack

Bei OpenStack handelt es sich um ein Open Source Projekt zur Bereitstellung einer IaaS Lösung. Das Projekt wurde im Juli 2012 von Rackspace zusammen mit der NASA ins Leben gerufen und hat mit über 200 Unternehmen mittlerweile eine Vielzahl Unterstützer.

Um eine eigene Private-Cloud OpenStack Installation zu betreiben, stehen Software Pakete für die gängigsten Linux Distributionen wie Debian, SUSE oder Red Hat zur Verfügung. Zudem existieren spezielle Cloud-Computing Anbieter, welche Public-Clouds auf Basis der OpenStack Technologie anbieten. Einer dieser Anbieter ist u.a. HP mit der HP Cloud. Der Betrieb einer Public-Cloud auf Grundlage einer OpenSource Lösung wie OpenStack sorgt für eine hohe Portabilität und Unabhängigkeit von unterschiedlichen Cloud-Computing Anbietern.

Neue OpenStack Versionen werden in einem Release-Zyklus von sechs Monaten (jeweils im Oktober und April) veröffentlicht. Die aktuelle Version, die der Untersuchung zu Grunde lag, trägt den Namen „Havana“. OpenStack besteht aus verschiedenen modularen Komponenten, auch als Projekte bezeichnet, die mit unterschiedlichen Funktionen über ein gemeinsames sog. Dashboard verwaltet werden können. Über dieses Dashboard können dabei u.a. Ressourcen sowie Nutzer verwaltet werden. Zudem ist es möglich diese Funktionen über die OpenStack API zu nutzen. Diese ist Amazon EC2 kompatibel. Ein Umstand, der eine Portierung mit minimalem Aufwand möglich macht.

Horizon stellt ein komfortables Web-Interface zur zentralen Konfiguration und Verwaltung einer OpenStack Installation in Form eines Dashboards zur Verfügung. Über dieses Dashboard können Instanzen, Festplattenimages und Netzwerkressourcen verwaltet werden.

5. Untersuchung von Private- und Public-Cloud Anbieter

Bei der **Nova** Komponente einer OpenStack Installation handelt es sich um eine Kernkomponente zur Instrumentierung und Steuerung einer Cloud-Computing Plattform. Es können virtuelle Maschinen sowie Netzwerke erstellt werden. Derzeit werden verschiedene Virtualisierungstechnologien in Form von Treibern für KVM, Xen, LXC oder VMware bereitgestellt.

Neutron (ehemals Quantum) abstrahiert die Netzwerkschicht und erlaubt es Nutzern (virtuelle) Netzwerke und IP Adressen für OpenStack Installationen zu verwalten. Die Virtuellen Maschinen die von Nova verwaltet werden können über entsprechende Interfaces zu Netzwerken zugewiesen werden.

Mit der **Swift** Komponenten steht ein Speicher zur verteilten und redundanten Speicherung von Daten aller Art zur Verfügung. Die Organisation der Daten erfolgt in Containern in Form von Objekten die wiederum mit Metadaten versehen werden können.

Festplatten-Images werden in der **Glance** Komponenten registriert und organisiert. Hierbei stehen die gängigsten Image-Formate wie VHD, VDI oder AMI zur Verfügung. Die tatsächliche Speicherung der Images erfolgt in der Swift Komponente.

Die Erfassung der Ressourcennutzung einer OpenStack Installation ist die Aufgabe der **Ceilometer** Komponente. Hiermit kann für jede Ressource die exakte Nutzung ermittelt und gespeichert werden. Ähnlich der CloudWatch-Komponente von Amazon lassen sich über das Ceilometer diverse Metriken unterschiedlicher vorgestellter Komponenten erfassen.

In Abb. 7 wird dargestellt, wie die Ceilometer Komponente mit Hilfe von Agenten in der Lage ist, verschiedene Messwerte zu erlangen. Die Informationsgewinnung erfolgt entweder über den Oslo Notification Bus oder über Push- bzw. Pollagents. Die erfassten Messwerte werden in einer Datenbank gesammelt und können u.a. zur Aggregation historischer Werte genutzt werden.

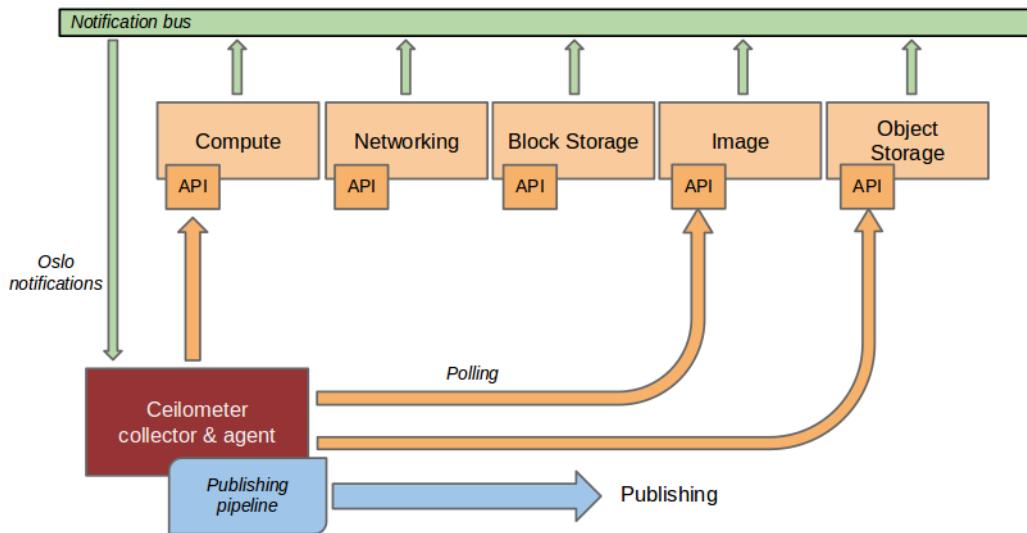


Abbildung 7: OpenStack Ceilometer

Mit Einführung der aktuellen Version besteht die Möglichkeit, Alarme zu definieren, die bei Überschreitung eines Grenzwertes für einen oder mehrere Messwerte auslösen. So ist es möglich ein Autoscaling auf Grundlage von Alarmen zu realisieren, welche die Ressourcennutzung beobachten und bei entsprechender Auslastung reagieren.

Kwapi [43] ist ein Framework zu Erfassung des Energieverbrauchs durch unterschiedliche Messgeräte und die anschließende Einspeisung dieser Messwerte zurück in die OpenStack Installation. Die Vision von Kwapi ist es, den Energieverbrauch zu beobachten, als Weiterberechnungsgrundlage zu nutzen und effizientes Scheduling für OpenStack Installationen zu ermöglichen.

5. Untersuchung von Private- und Public-Cloud Anbieter

Derzeit unterstützt Kwapi populäre Energiemessinstrumente wie Wattsup oder IPMI. Diese Unterstützungen stehen in Form von Treibern zur Verfügung und sind leicht um neue Messgeräte erweiterbar. Die verfügbaren Treiber werden über einen Treiber Manager verwaltet, der u.a. regelt, in welchen Abständen Messungen vorgenommen werden. Diese Messungen werden durch die Treiber dekodiert und in Form von Nachrichten über einen Bus zur Verfügung gestellt.

Nachrichten können mithilfe dieses Buses von unterschiedlichen Plugins gelesen werden. Derzeit existiert ein Plugin zur Graph Visualisierung der Verbrauchswerte (RRD) (Abb. 8) sowie ein Plugin zur Bereitstellung einer API Schnittstelle für die OpenStack Ceilometer Komponente. Über diese Schnittstelle kann das Ceilometer Verbrauchsdaten in Watt (sowie kumulierter Werte in kWh) mit Hilfe eines Polling Agenten beziehen und somit die klassischen Auswertungen des Ceilometers um Energie Messwerte erweitern.



Abbildung 8: Kwapi RRD Visualisierung

Fazit

OpenStack ist ein wichtiges OpenSource Projekt bestehend aus vielen mächtigen modularen Komponenten. Diese Komponenten vereinen weitestgehend die Funktionalitäten der weiteren vorgestellten Cloud-Computing Anbieter. Im Gegensatz zu diesen Anbietern sind die Komponenten in diesem Fall jedoch erweiterbar und werden von einer großen Community gepflegt.

Unter der Prämisse möglicher Hardware Messungen im Bezug auf den Energieverbrauch (meist in einer on-site Cloud) steht mit dem Kwapi Framework eine hilfreiche und erweiterbare Komponente zur Verfügung. Durch die Alarming Funktion der Ceilometer Komponente ist es zudem möglich, auf einen möglichen hohen Energieverbrauch zu reagieren oder über einen Nova Scheduler Instanzen zu migrieren sowie ineffiziente Instanzen zu terminieren. Im untersuchten Umfeld stellt OpenStack somit den mächtigsten und umfanglichsten Ansatz für eine potentielle Steigerung der Energieeffizienz zur Verfügung.

6. Zusammenfassung und Ausblick

Ein effizientes Rechenzentrum (PUE), das lokale Klima sowie eine saubere Energiequelle (GPUE bzw. CUE) bilden den Referenzrahmen einer guten CO₂-Bilanz. Eine Kombination der in Kapitel 3 dargestellten Aspekte kann zusätzlich zu einer Steigerung der Energieeffizienz und in Folge dessen zu einer Minimierung des Energieverbrauchs bei der Nutzung einer Cloud-Computing Infrastruktur führen.

Im Rahmen dieser Fachstudie wurden verschiedene Private- und Public-Cloud Anbieter auf Möglichkeiten zur Steigerung der Energieeffizienz hin untersucht. Die ermittelten technischen und nicht-technischen Aspekte sprechen deutlich für ein hohes Potential beim Einsatz von Cloud-Computing gegenüber herkömmlichen Rechenzentren. Während das technische Fundament für energieeffizientes Handeln besteht, mangelt es jedoch an Transparenz bzgl. der Energieverbrauchswerte, um tatsächlich Entscheidungen auf Grundlage dieser treffen zu können. Demzufolge lassen sich neben den rein auf der Technik des Cloud-Computings basierenden Vorzügen derzeit keine fundierten Aussagen über die Energieeffizienz der einzelnen Anbieter ableiten. Eine Ausnahme hiervon stellt der Anbieter GreenQloud dar, der sich explizit in dieser Nische positioniert hat. Wie in Kapitel 4 angemerkt, birgt eine Veröffentlichung der Energieverbrauchswerte für die Cloud-Computing Anbieter zum aktuellen Zeitpunkt jedoch mehr Risiken als Vorteile.

Einen interessanten Ansatz, der sich zwar hauptsächlich mit der Vertragsgestaltung zwischen Cloud-Computing Anbietern und Nutzern beschäftigt, liefert die Arbeit „*Pay as You Go in the Cloud: One Watt at a Time*“ [44]. Derzeit werden die Fixkosten für den Betrieb des Rechenzentrums zusammen mit den variablen Energiekosten für den Betrieb von Instanzen auf Grundlage von standardisierten Nutzungszeiten auf den Kunden umgelegt. Eine ressourcenintensive Anwendung führt zu einem erhöhten Energieverbrauch und damit zu höheren Kosten für die Cloud-Computing Anbieter. Eine erhöhte Auslastung der Rechenressourcen kann demgemäß zu einer Minimierung des Gewinns der Cloud-Computing Anbieter führen, da die Kalkulationen darauf basieren dass Ressourcen ungenutzt bleiben. Eine Abrechnung in Instanz-Stunden berücksichtigt diese potentiellen Mehrkosten für die Anbieter jedoch nicht.

6. Zusammenfassung und Ausblick

Das alternativ vorgeschlagene Modell sieht eine Bezahlung auf Grundlage von tatsächlich verbrauchter Energie (in Wattstunden) statt (potentiell ungenutzter) Instanz-Stunden vor. Diese Art der Abrechnung setzt eine Energietransparenz der einzelnen Anbieter voraus. Die Preisgestaltung als primäres Entscheidungsmerkmal kann somit in den Hintergrund gedrängt werden und gleichermaßen zu einer Sensibilisierung der Nutzer hinsichtlich der Energieeffizienz und der eingesetzten Energiequellen beitragen.

Gleichzeitig kann eine Abrechnung nach Wattstunden zu einem konstanteren Gewinn für die Cloud-Computing Anbieter führen, da die oben aufgezeigten potentiellen Mehrkosten durch den Nutzer getragen werden. Zudem wird ein Anreiz geschaffen, durch die Bereitstellung einer energieeffizienten Infrastruktur, die Konkurrenzfähigkeit gegenüber anderen Anbietern zu steigern. In Summe könnten also sowohl die Nutzer als auch die Anbieter von dieser Abrechnungsmethode profitieren.

Abschließend soll nicht unerwähnt bleiben dass der Einsatz von moderner Software bzw. die Nutzung effizienter Algorithmen einen nicht unerheblichen Beitrag zur Energiebilanz leisten kann. Selbst auf hocheffizienter Hardware führen unnötige Speicherzugriffe in der Anwendung zu einer Erhöhung des Energiebedarfs.

Literaturverzeichnis

- [1] Mell, P. & Grace, T.: The NIST Definition of Cloud Computing, NIST Special Publication 800-145, Gaithersburg, September 2011. Die deutsche Übersetzung kann den Ausführungen des BSI entnommen werden:
https://www.bsi.bund.de/DE/Themen/CloudComputing/Grundlagen/Grundlagen_node.html
- [2] Liu, F.; Tong, J.; Mao, J.; Bohn, R.; Messina, J.; Badger, L. & Leaf, D.; NIST Cloud Computing Reference Architecture, NIST Special Publication 500-292, Gaithersburg, September 2011
- [3] Ibrahim, S.: Towards Pay-As-You-Consume Cloud Computing, Washington DC, 2011
- [4] Irrek, W. & Thomas, S.: Definition Energieeffizienz, Wuppertal, 2008
http://wupperinst.org/uploads/tx_wupperinst/energieeffizienz_definition.pdf
- [5] APC - Vorbeugende Wartungsstrategie für Rechenzentren
http://www.apcmedia.com/salestools/DBOY-73VLPT/DBOY-73VLPT_R0_DE.pdf
- [6] Cloud Computing and Sustainability: The Environmental Benefits of Moving to the Cloud, Accenture, 2010
http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture_Sustainability_Cloud_Computing_TheEnvironmentalBenefitsofMoving_totheCloud.pdf
- [7] Google's Green PPAs: What, How, and Why
http://static.googleusercontent.com/external_content/untrusted_dlcp/www.google.com/en/us/green/pdfs/renewable-energy.pdf
- [8] Green Grid Data Center Power Efficiency Metrics: PUE and DCIE, 2008
http://www.thegreengrid.org/~media/WhitePapers/White_Paper_6_-PUE_and_DCIE_Eff_Metrics_30_December_2008.pdf?lang=en
- [9] Bitkom - Wie messe ich den PUE richtig? Leitfaden, 2011
http://www.bitkom.org/files/documents/Leitfaden_PUE_Messung.pdf
- [10] 2013 Uptime Institute Data Center Industry Survey
<http://uptimeinstitute.com/2013-survey-results>
- [11] GPUE – Green Power Usage Effectiveness
<http://blog.greenqloud.com/greenpowerusageeffectiveness-gpue/>
- [12] Sovacool, B. K.: A Critical Evaluation of Nuclear Power and Renewable Electricity in Asia, Journal of Contemporary Asia, Vol. 40, No. 3, August 2010, p. 386.

- [13] Carbon Usage Effectiveness (CUE): A Green Grid Data Center Sustainability Metric, 2010
<http://www.thegreengrid.org/~/media/WhitePapers/CarbonUsageEffectivenessWhitePaper20101202.ashx>
- [14] Bartalos, P.; Blake, M.; Remy, S.: Green Web Services: Models for Energy-Aware Web Services and Applications, 2011, Irvine, CA.
- [15] Kansal, A.; Zhao, F.; Liu, J.; Kothari, N.; Bhattacharya, A.: Virtual machine power metering and provisioning. In 1st ACM symposium on Cloud computing, 2010
- [16] Verge, J.: Data Center Knowledge - How Big is AWS? Netcraft Finds 158,000 Servers
<http://www.datacenterknowledge.com/archives/2013/06/04/how-big-is-aws-new-netcraft-numbers-show-insight/>
- [17] Mastodon C Live carbon ranking
<https://www.mastodonc.com/dashboard>
- [18] amee - Free financial & environmental business risk insights
<http://www.amee.com>
- [19] AWS Global Infrastructure
<http://aws.amazon.com/de/about-aws/globalinfrastructure/>
- [20] AWS and Sustainable Energy
<http://aws.amazon.com/de/about-aws/sustainable-energy/>
- [21] Amazon's Cloud Goes Modular in Oregon
<http://www.datacenterknowledge.com/archives/2011/03/28/amazons-cloud-goes-modular-in-oregon/>
- [22] Amazon EC2 – Häufig gestellte Fragen
<http://aws.amazon.com/de/ec2/faqs/>
- [23] Google Developers – Häufig gestellte Fragen
<https://developers.google.com/compute/docs/faq?hl=de>
- [24] Google Rechenzentren - Standorte von Rechenzentren
<http://www.google.com/about/datacenters/inside/locations/index.html>
- [25] Google Compute Engine Documentation
<https://developers.google.com/compute/docs/>
- [26] Heise Newsticker - Google nimmt Rechenzentrum im finnischen Hamina in Betrieb
<http://www.heise.de/newsticker/meldung/Google-nimmt-Rechenzentrum-im-finnischen-Hamina-in-Betrieb-1341059.html>

- [27] Google Rechenzentren - Effizienz: So machen wir das
<http://www.google.com/about/datacenters/efficiency/internal/index.html#measuring-efficiency>
- [28] The ISO 14000 family of International Standards
http://www.iso.org/iso/theiso14000family_2009.pdf
- [29] Auto Scaling on the Google Cloud Platform
<https://cloud.google.com/resources/articles/auto-scaling-on-the-google-cloud-platform>
- [30] GitHub - Orchestrator and status publisher tools
<https://github.com/GoogleCloudPlatform/solutions-google-compute-engine-orchestrator>
- [31] Microsoft Windows Azure Datenschutz
<https://www.windowsazure.com/en-us/support/trust-center/privacy/>
- [32] Microsoft Citizenship Report 2013
<http://download.microsoft.com/download/2/5/9/2597728D-72EE-4FDC-BD93-814AD436ABDA/FY13%20Report%20FINAL%20Oct%202013.pdf>
- [33] MSDN - The Autoscaling Application Block
[http://msdn.microsoft.com/en-us/library/hh680892\(v=PandP.50\).aspx](http://msdn.microsoft.com/en-us/library/hh680892(v=PandP.50).aspx)
- [34] Windows Azure Documentation - Load Balancing Virtual Machines
<http://www.windowsazure.com/en-us/manage/windows/common-tasks/how-to-load-balance-virtual-machines/>
- [35] Greenpeace International: How Clean is Your Cloud?,
The Netherlands, April 2012
<http://www.greenpeace.org/international/Global/international/publications/climate/2012/iCoal/HowCleanisYourCloud.pdf>
- [36] 2012 Energy Statistics in Island
http://www.os.is/gogn/os-onnur-rit/orkutolur_2012-enska.pdf
- [37] Verne Global – Why Verne Global?
<http://www.verneglobal.com/why-verne-global/available-power>
- [38] Advania Environment
<http://www.advania.com/datacentres/environment/>
- [39] Verne Global & Colt Industries Show a Zero Carbon Data Center
<http://www.zdnet.com/blog/forrester/verne-global-and-colt-industries-show-a-zero-carbon-data-center/818>
- [40] GreenQloud – Go Green
<http://greenqloud.com/startup-gogreen/>
- [41] Verne Global Energy Calculator
http://verneglobal.com/resources/energy_calculator

- [42] VMware vCloud Architecting a vCloud
<http://www.vmware.com/files/pdf/VMware-Architecting-vCloud-WP.pdf>
- [43] Github – Kwapi
<https://github.com/stackforge/kwapi>
- [44] Teramoto, K.; Huang, H. H.: Pay as You go in the Cloud: One Watt at a Time, Salt Lake City, UT, November 2012

Alle URLs wurden zuletzt am 25.10.2013 abgerufen und geprüft.

Erklärung

Ich versichere, diese Arbeit selbstständig verfasst zu haben. Ich habe keine anderen als die angegebenen Quellen benutzt und alle wörtlich oder sinngemäß aus anderen Werken übernommene Aussagen als solche gekennzeichnet. Weder diese Arbeit noch wesentliche Teile daraus waren bisher Gegenstand eines anderen Prüfungsverfahrens. Ich habe diese Arbeit bisher weder teilweise noch vollständig veröffentlicht. Das elektronische Exemplar stimmt mit allen eingereichten Exemplaren überein.

Ort, Datum, Unterschrift

Ort, Datum, Unterschrift