

Amazon Elastic Compute Cloud Hochschule Mannheim

Peter Gehrt

Fakultät für Informatik
Hochschule Mannheim
peter.gehrt@stud.hs-mannheim.de

20.11.2009

Agenda

- Einführung & Grundlagen
- Amazon EC2
- Services
- Performance
- Live-Demo
- Risiken und Nebenwirkungen
- Was macht die Konkurrenz?
- Zusammenfassung & Ausblick

Abstract

Neben Büchern, Elektronik und Kleidung bietet Amazon nun auch Server und Dienstleistungen an

- einfach
- performant
- automatisiert
- schnell replizierend
- hochverfügbar
- kostengünstig



Cloud-Computing

The interesting thing about cloud computing is that we've redefined cloud computing to include everything that we already do ... What is it? It's complete gibberish. It's insane. When is this idiocy going to stop?¹



Larry Ellison, CEO Oracle

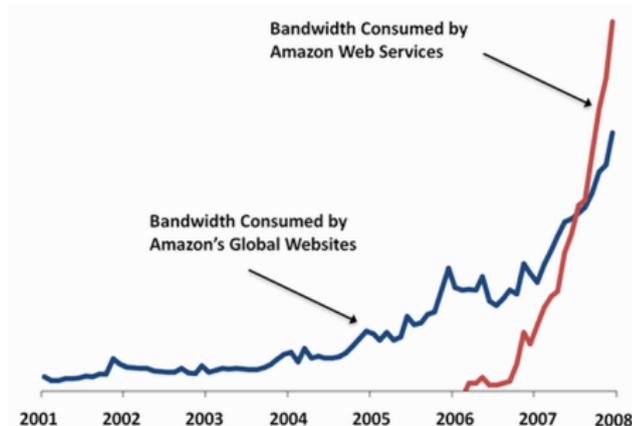
¹www.cloudave.com/link/oracle-cloudy-cloud-computing-plans-saas

Amazon Elastic Compute Cloud

- Bestandteil der Amazon Webservices und somit Anwendung von SOA und Cloud-Computing
- Private-Cloud: Konsolidierung durch XEN-Virtualisierung
- einfache API Schnittstellen für interne Zwecke
- Ab 2006 als Public-Cloud auf dem freien Markt verfügbar
- Abgrenzung zu PaaS und SaaS Angeboten

Entwicklung

- Benutzer haben Amazon Eigenbedarf längst überholt



- September 2008: 50.000 EC2 Instanzen gestartet
- Oktober 2008: 40.000 Server in den EC2 Rechenzentren
- Seit Bestehen über 15.5 Millionen EC2 Instanzen gestartet²

²www.jackofallclouds.com/2009/09/anatomy-of-an-amazon-ec2-resource-id/

Motivation

- Pay as you go
 - Umwandlung von einmaligen Investitionskosten in laufende Kosten
 - Einfache Planung, z.B. Datenbankserver für 81 Euro/Monat
- On-Demand Zugriff auf unendliche Ressourcen
 - Viele Instanzklassen und Betriebssysteme zur Auswahl
 - Unkomplizierte vollständige Kontrolle
- Elastische Skalierung
 - Herkömmliche Rechenzentren nur 5-20 % Auslastung³
 - Starten und Stoppen von Instanzen in wenigen Minuten, kein Risiko der Überprovisionierung
- Minimale eigene IT-Anforderungen & einfacher Zugang
 - Einfache Anbindung an andere Amazon Webservices
 - Regelmäßige Aktualisierung der Web 2.0 Management Console

³A Berkeley View of Cloud Computing

Beispiele

- Parallelisierung statt Batchbetrieb
 - Gleiche Kosten für 1 Server 100 Stunden oder
 - 100 Server für 1 Stunde!
- SmugMug (OpenID, EC2, S3): 20 Mitarbeiter und 10 Millionen Dollar Gewinn in 2006
- Animoto (EC2): Skalierung von 50 auf 400 Server in 3 Tagen - und wieder zurück
- The New York Times: Archivierung von 4 Millionen Artikeln in EC2, 36 Stunden und 260 US-Dollar

Übersicht & API

- Gut dokumentierter WSDL-Webservice über SOAP/REST
- SDK und Bibliotheken für Java, C#, Python, PHP und Ruby mit über 130 Code-Snippets
- Command-Line-Interface API Tools
- Web 2.0 Management Console
- Firefox Plugin ElasticFox



Amazon Machine Image

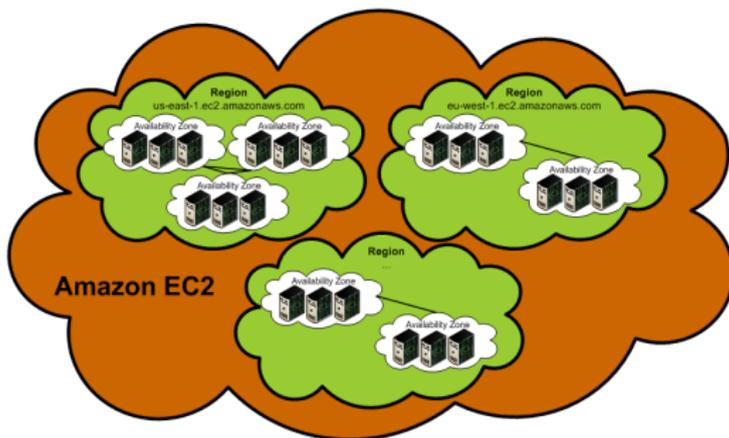
- AMI startet in wenigen Sekunden, danach voller Zugriff
- Betriebssysteme: Unix, Linux, Windows
- Herstellen und Starten eigener AMI möglich
- Neue Lizenzmodelle & Timesaver durch integrierte Software
 - zusätzliche Kosten pro Stunde (paid-AMIs)
 - IBM DB2, Oracle, Microsoft-SQL, Gigaspaces
- Login durch Zertifikat und asymmetrische Schlüssel
- Firewall-Konfiguration durch frei anpassbare Security Groups

Instanzklassen & Preismodell

Instanz	ECU	Ram/HDD(GB)	Cent/Stunde
Small	1	1.7 / 160	8.5
Large	4	7.5 / 850	34
Extralarge	8	15 / 1690	68
High Memory Double	13	34 / 850	120
High Memory Quadruple	26	68 / 1690	240
High CPU Medium	5	1.7 / 350	17
High CPU Extra Large	20	7 / 1690	68

- Elastic Compute Unit entspricht laut Amazon Intel Xeon 1.2 GHz
- Windowsinstanzen 15% Aufpreis
- Services und Netzwerktraffic werden gesondert abgerechnet

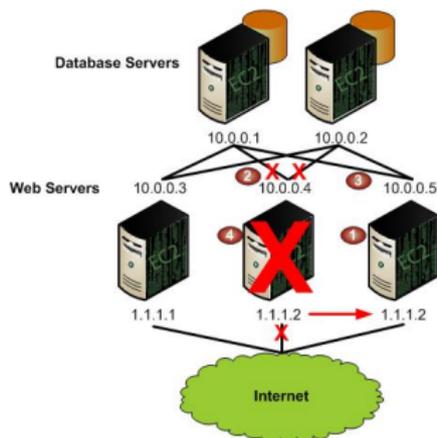
Regionen & Zonen



- Regionen in USA, Europa und ab 2010 auch Asien
 - Benutzernähe durch CloudFront
- Jede Zone besteht aus mehreren Availability Zonen
 - getrennte und vollredundante Rechenzentren
 - Verteilung der Anwendung durch Elastic Load Balancer

Elastic IP

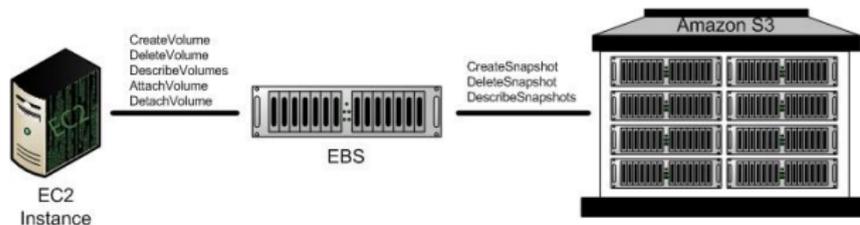
- Dynamische Zuweisung einer öffentlichen IP-Adresse zu verschiedenen Instanzen



- Servicetransparenz bei Ausfall oder Wartung eines Servers: Umzug der EIP auf neuen Server
- IPv4 Knappheit: EIP verursacht nur Kosten, wenn beantragt aber nicht zugewiesen

Storage

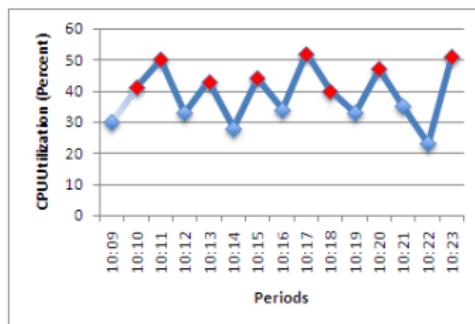
- Interner Speicher einer Instanz ist flüchtig und nach dem Herunterfahren nichtmehr verfügbar
- Persistenz durch Elastic Block Storage (EBS) oder S3
 - Betriebssystem oder andere CD/DVD Images anbinden
 - Eigene AMI herstellen und abspeichern
 - Zentrale Lagerung der Unternehmensdaten



- Public Data Sets (PDS) für Naturwissenschaften, Klimadaten oder Enzyklopädien

CloudWatch & AutoScale

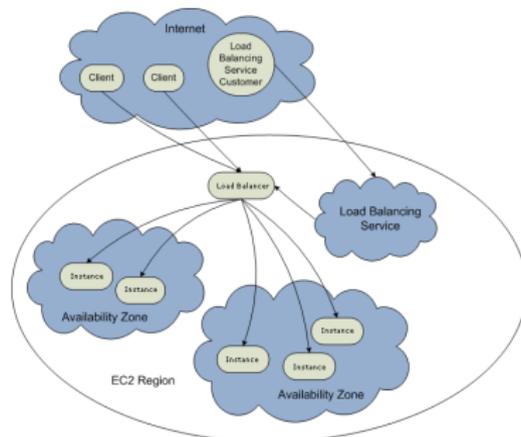
- CloudWatch:
 - Überwachung von Instanzparametern und Berechnung von Metriken
 - CloudWatch Diagramme in Management Console



- AutoScale skaliert die Instanzen aufgrund der CloudWatch-Metriken
 - Skalierung durch Erhöhung der Instanzanzahl, Konfiguration neuer Instanzen durch Startparameter
 - Triggerbasiert nach Zeitraum und Dauer des Durchbruchs
- Proaktives Scaling durch Skripte nach Zeitplan

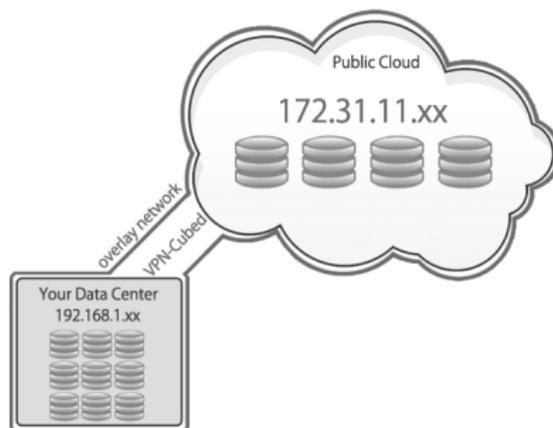
Elastic Load Balancer

- Überwacht verteilte Anwendung durch CloudWatch-Metriken
- Vermeidung von Spitzenlast auf einer Instanz
 - Starten von neuen Instanzen mit AutoScale
- Verteilt Last gleichmäßig auf alle Zonen in einer Region



Virtual Private Cloud

- VPN-Anbindung in das Unternehmensnetzwerk
- Transparenter Zugriff auf die EC2 Server
- Trennung Datacenter und Anwendungsserver
- Probleme:
 - komplexe IPSEC-Konfiguration
 - Netzwerktraffic im Auge behalten



4

⁴www.cohesiveft.com/vpn4ec2/

Sicherheitsaspekte

- Ausfallsicherheit
 - Kein Single Point of Failure: Vollredundanz durch Regionen & Availability Zonen, hohe Verfügbarkeit und Fehlerresistenz
 - Optimierung durch Elastic Load Balancer und Elastic IP
 - Menschliche Fehlerfaktoren durch Fachpersonal und professionelles Equipment reduziert (10.000 statt 100 Server pro Admin)
 - Seit Oktober 2008 klar definiertes Service Level Agreement (SLA)
- Datensicherheit
 - Zugriff durch Zertifikat und asymmetrische Keypairs
 - Trennung der Benutzer durch XEN Technologie
- Netzwerksicherheit
 - Firewall im XEN Hypervisor, Security Groups frei anpassbar
 - DoS-Angriffe durch Hochskalierung aussitzen
 - Amazon mehr Ressourcen für Sicherheit als normales Unternehmen

Grundlagen

- Illusion einer unendlichen Performance durch hohen Durchsatz und Rechengeschwindigkeit
 - Amazon hohe Leistungsdichte und HighEnd-Hardware
 - Optimierung durch CloudFront, Elastic Load Balancer und AutoScale
- Eigene Messungen
 - Netzwerkdurchsatz (iperf)
 - Ladezeiten (ping)
 - Rechenleistung bei 32 MB Primzahlen (wprime ⁵)
 - I/O-Durchsatz der Festplatte und Arbeitsspeicher (hdparm und kcbench)

⁵www.wprime.net

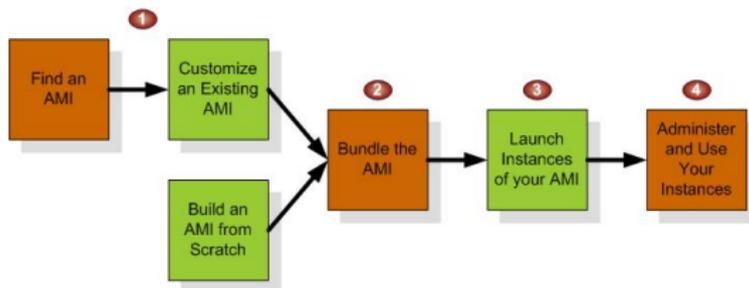
Untersuchung

- Small: 1x Xeon @ 2.66 GHz mit 1.7 GB Arbeitsspeicher
- Quadruple: 8x Xeon Quadcore @ 2.66 GHz mit 70 GB Arbeitsspeicher

Test	Small	Quadruple	Bemerkung
Primzahlen (s)	170	7	(Core i7 @ 4GHz: 6s)
Kernelkompilierung (s)	604	33	
Ping (ms)	15 - 40		(intern 0.5)
Netzwerk extern (Mbit/s)	20	90	(Engpaß)
Netzwerk intern (Mbit/s)	350	930	
I/O Durchsatz (MB/s)	300	860	

→ Im Vergleich sehr hohe Performance

Ablauf



Datensicherheit

- Daten stellen wichtigsten Teil eines Unternehmens dar, externe Speicherung führt zu rechtlichen Bedenken
- Fremde Unternehmen im Ausland andere geltende Gesetze
- Abspeichern von Logindaten in einer eigenen AMI?
- Lock-In von Daten und Anwendungen
- Beispiel:
 - Microsoft Sidekick USA (1 Millionen Geräte) verlor 2009 alle Adressen, Kalendereinträge, Fotos und Notizen ihrer Benutzer
 - Vertragsstrafen der Amazon SLA?
- Lösungen:
 - Verschlüsselung der Übertragung und Datenspeicherung
 - Einheitliche API (IBM)
 - Aufteilung der Ressourcen in Regionen oder Availability Zonen

Netzwerk

- Geteilte Ressourcen mit anderen Benutzern
 - I/O-Engpässe
 - Spam Sperrlisten
- Engpass der Datenübertragung
10 TB in S3 speichern dauert 45 Tage und kostet 1000 US-Dollar
- Lösung:
 - Durch alternative Protokolle (Aspera) Verdreifachung der Übertragungsrate
 - Jim Gray: Keine Datenübertragung kann mit einer LKW Ladung voller Festplatten mithalten⁶
 - Durch Versand von Festplatten (FedEx): 2-5 Tage und nur 400 Dollar
 - Alternative Datenbestände der Public Data Sets verwenden

⁶<http://queue.acm.org/detail.cfm?id=864078>

Programmierung

- Allgemeine Risiken bei der parallelen Programmierung wie Synchronisation und Datenschutz
- Unzureichende Lizenzmodelle von fremden Softwareanbietern
- Testen und Debuggen der eigenen Anwendungen
- Lösung:
 - Umdenken bei Lizenzmodellen
 - Amazon baut Testbarkeit durch Attribut in Netzwerk und Instanzen ein

Herkömmliche Anbieter

- Vergleich des c't Magazins⁷ von 5 Hostinganbietern:
 - 15 % der Antwortzeiten über 300 ms
 - 4 Ausfälle, im schlechtesten Fall 16!
- EC2 im gleichen Zeitraum:
 - Eine kurze Störung der API, nach 37 Minuten behoben
 - Serverfunktionalität war nicht gestört
 - Antwortzeiten durch CloudFront deutlich besser

⁷Ausgabe 23/2007, Heise Verlag

Cloud-Computing Anbieter

- GoGrid: freiere Zuweisung der Ressourcen wie CPU oder Arbeitsspeicher aber keine eigenen AMIs
- Mosso: benutzt Cloud-Computing für wartungsfreie Hostingpakete
- IBM BlueCloud: nicht auf dem freien Markt verfügbar, entwickelt einheitliche API Schnittstelle
- Microsoft Azure: .NET Framework Umgebung, Komplettpaket von Dell
- HP: IT-Industrialisierung durch HP-Pods
 - Cloud-Computing Industriecontainer mit 10.000 CPU-Kernen oder 9 PB Speicher
 - In Zukunft auch HP Dynamic Cloud Services für die Öffentlichkeit

Was wissen wir nun?

- Geringe Einstiegshürde durch einfache Bedienbarkeit und umfangreiche Dokumentation
- Performante, kostengünstige, vollautomatisierte und hochverfügbare Lösungen
- Transparente Anbindung an eigene IT-Infrastruktur
- Kritische Punkte wie Datensicherheit und Übertragungsraten beachten
- Rechtzeitig eigene Erfahrungen sammeln

Was bringt die Zukunft?

- Anbieterübergreifende einheitliche API
- Umfangreiche Management Console
- Fertige Lösungen wie MapReduce und Plattform / Software as a Service-Angebote
- Auswirkungen auf Lizenzmodelle der Softwareanbieter
- Innovationen der Hardwarehersteller

Vielen Dank für Eure Aufmerksamkeit!

Quellen

- Bilder und Dokumentation von Amazon Webservices unter <http://aws.amazon.com/>
- Reese, George. Cloud Application Architectures. O'Reilly.2009.
- Above the Clouds: A Berkeley View of Cloud Computing
<http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.pdf>
- Larry Ellison im Videointerview über Cloud-Computing.
<http://www.techfieber.de/2009/10/02/video-larry-lastert-warum-oracle-ceo-larry-ellison-cloud-computing-hasst/>
- Weitere EC2 Beispiele <http://aws.amazon.com/solutions/case-studies/>
- New York Times EC2 Projekt <http://open.blogs.nytimes.com/2008/05/21/the-new-york-times-archives-amazon-web-services-timesmachine/>
- Firefox Plugin ElasticFox <http://sourceforge.net/projects/elasticfox/>
- Untersuchung der Java Performance http://pvs.informatik.uni-heidelberg.de/Teaching/CLCP-09/CLCP_SS2009_Christian_Schwarz_EC2.pdf
- Aspera
<http://gigaom.com/2009/09/10/aspera-bypasses-bandwidth-bottlenecks-to-ec2/>
- Vergleich Cloud-Computing Anbieter <http://www.infoq.com/articles/cloud-comparison>