

Cloud-Infrastrukturen

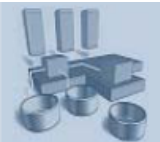
Seminar

Cloud Data Management WS09/10



Inhalt

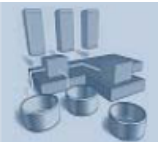
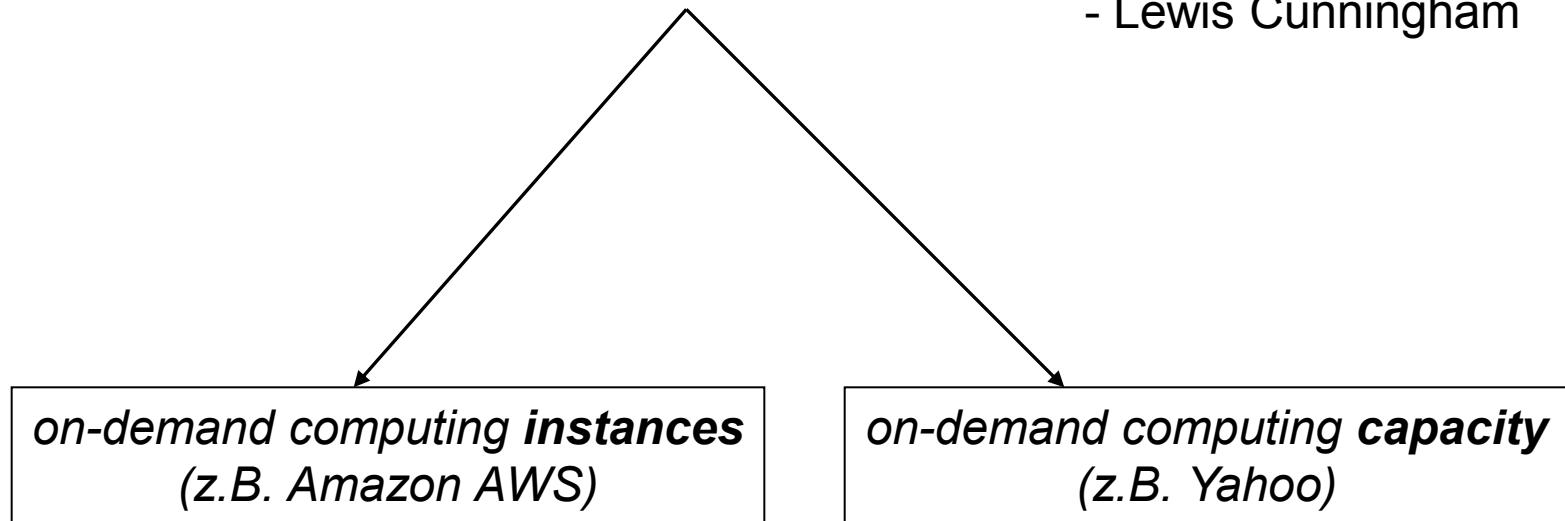
1. Allgemeines
2. Amazon EC2
3. Yahoo Cloud
4. Vergleich
5. Fazit
6. Literatur



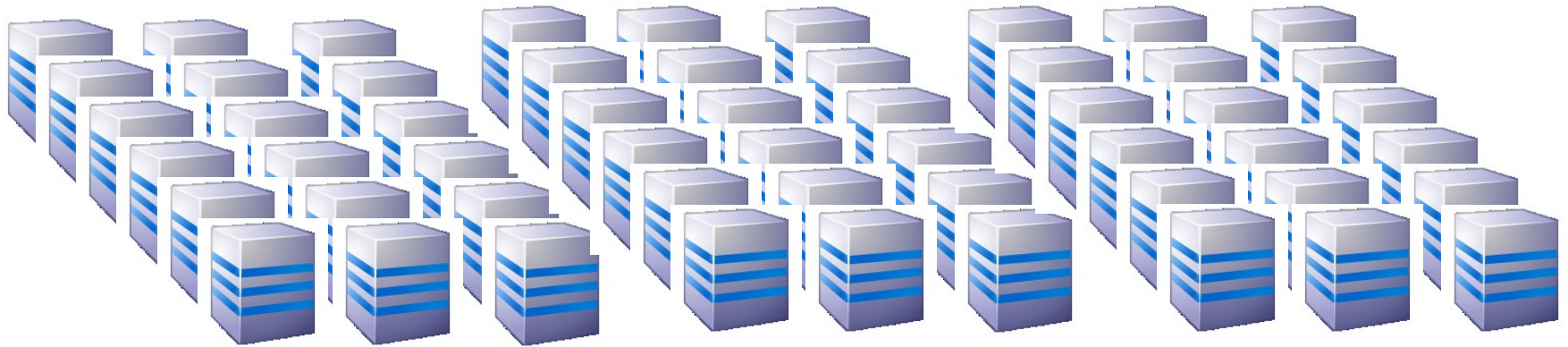
Definition

Cloud computing is using the internet to access someone else's software running on someone else's hardware in someone else's data center”

- Lewis Cunningham

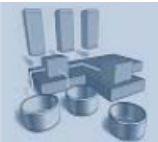


Allgemeiner Aufbau



Beide Architekturen bestehen aus lose gekoppelten Rechnern, ABER

1. On-Demand Computing Instances liefert **miet- oder kaufbare Instanzen** an Rechnerkapazität
2. On-Demand Computing Capacity bietet reine **Daten- oder Rechenleistung** zur Unterstützung rechenintensiver Algorithmen oder sehr großer Datenmengen



Amazon Webservice



Amazon Elastic Compute Cloud (Amazon EC2)

Amazon Simple Queue Service (Amazon SQS)

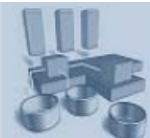
Amazon CloudFront

Amazon Elastic MapReduce

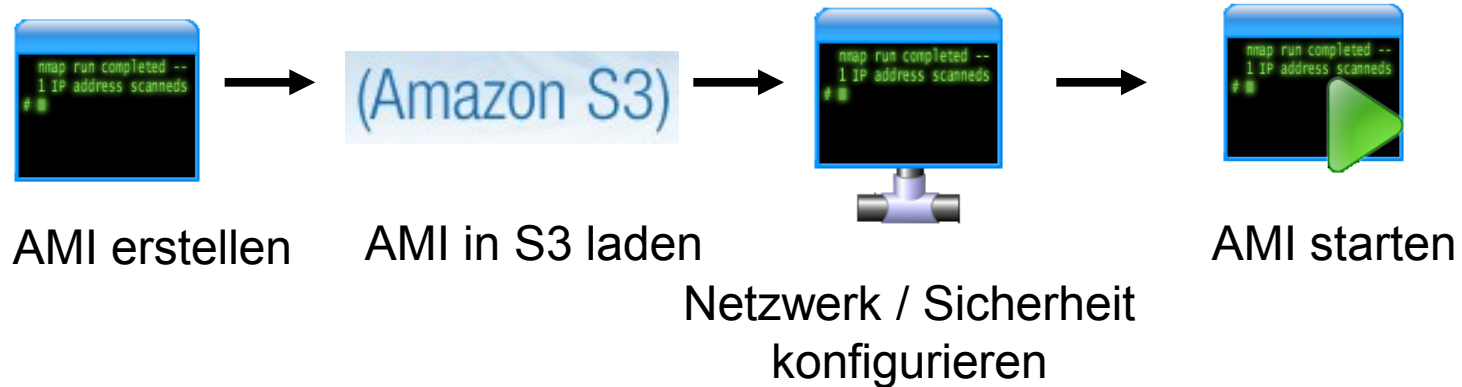
Amazon SimpleDB™

Amazon Simple Storage Service (Amazon S3)

und mehr....



- Reserved / on-demand **instances** mietbar
- Über webinterface sog. AMIs konfigurierbar
(Amazon machine image)



Instanz Typen

Standard Instances

Small (Default) 1.7 GB of memory, 1 EC2 CU 160 GB of instance storage, 32-bit platform

Large 7.5 GB of memory, 4 EC2 CU 850 GB of instance storage, 64-bit platform

Extra Large 15 GB of memory, 8 EC2 CU 1690 GB of instance storage, 64-bit platform

High-Memory Instances

High-Memory Double Extra Large Instance

34.2 GB of memory, 13 EC2 CU, 850 GB of instance storage, 64-bit

High-Memory Quadruple Extra Large Instance

68.4 GB of memory, 26 EC2 CU, 1690 GB of instance storage, 64-bit

High-CPU Instances

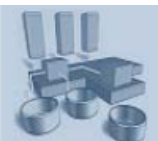
High-CPU Medium Instance

1.7 GB of memory, 5 EC2 CU, 350 GB of instance storage, 32-bit

High-CPU Extra Large Instance

7 GB of memory, 20 EC2 CU, 1690 GB of instance storage, 64-bit

EC2 Compute Unit => 1.0-1.2 GHz 2007 Opteron or 2007 Xeon processor



Preise Amazon EC2

Linux/UNIX	One-time Fee		
Standard Reserved Instances	1 yr Term	3 yr Term	Usage
Small (Default)	\$227.50	\$350	\$0.04 per hour
Large	\$910	\$1400	\$0.16 per hour
Extra Large	\$1820	\$2800	\$0.32 per hour
High-Memory Reserved Instances	1 yr Term	3 yr Term	Usage
Double Extra Large	\$3185	\$4900	\$0.56 per hour
Quadruple Extra Large	\$6370	\$9800	\$1.12 per hour
High-CPU Reserved Instances	1 yr Term	3 yr Term	Usage
Medium	\$455	\$700	\$0.08 per hour
Extra Large	\$1820	\$2800	\$0.32 per hour

Zuzügl. Transferkosten + Extra-Services...

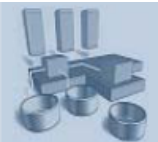


Amazon Simple Storage Service (Amazon S3)

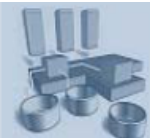
- Key/Value basierter Filehost-Service
- Daten werden in „amazonweit“ eindeutigen Buckets zusammengefasst
- Pro Bucket beliebig viele Dateien zw. 1B-5GB
- Host-Auswahl USA oder Europa möglich
- Zugriff über SOAP oder BitTorrent z.B.

- Beispiel:
 - <http://DB1.s3.amazonaws.com/db1.mdb>

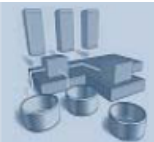
- Kosten:
 - \$0.170 per GB – next 50 TB / month of storage used
 - \$0.10 In / \$0.17-\$0.10 Out Transferkosten
 - \$0.12/1000 Transactions oder 10000 GET



- Einfaches verteiltes Datenbankmanagementsystem
- Einfachste Datenbankfunktionen
- Verzicht auf Schema, keine Joins, keine semantischen Integritätsbedingungen
- Dafür automatische Indizierung aller Daten
- Keine Administration & Wartung notwendig
- Kosten:
 - \$0.154 pro Amazon SimpleDB Machine Hour
 - 0.10\$ In / 0.17\$ - 0.10\$ Out Transferkosten



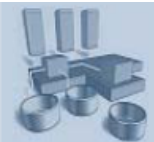
- Content Distribution Service
- Baut auf Amazon S3 auf und kopiert in S3 gespeicherte Daten zu geografisch unterschiedlichen Speicherknoten
- Ermöglicht niedrigere Latenzzeiten und höhere Datenübertragungsraten
- Kosten:
 - Ähnlich wie bei S3
 - Zuzüglich Kosten für die Verteilung der Originaldateien von Amazon S3



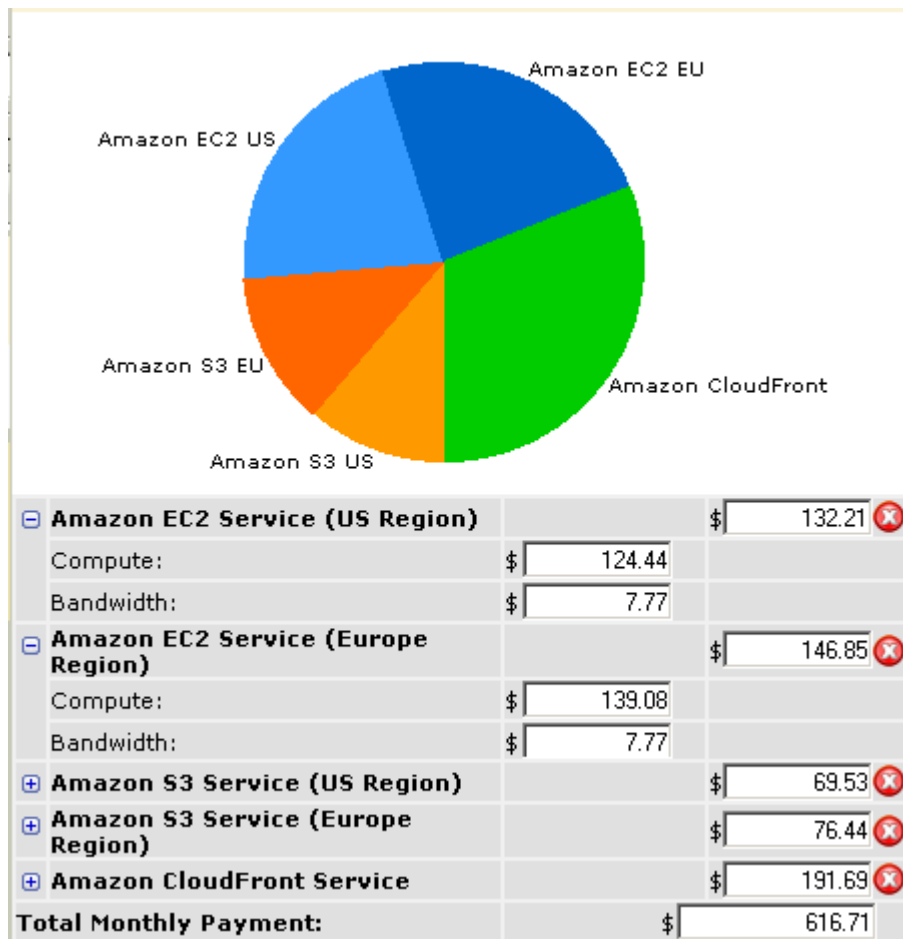
- Messaging-Service
- Erstellung eigener Message-Queues möglich
- Datenvolumen bis zu 8KB pro Message
- Messages werden 4 Tage gespeichert
- Können simultan abgerufen und gespeichert werden
- Einfaches Sperren beim Verarbeiten möglich
- Kosten:
 - \$0.01 pro 10000 Requests
 - \$0.10 In / \$0.17-\$0.10 Out Transferkosten
 - Zw. EC2 & SQS kostenlos



- MapReduce-Service mit Hilfe vom Hadoop-Framework
- Kombiniert Amazon EC2 und Amazon S3
- Teilt rechen- oder datenintensive Aufgaben in Teilaufgaben und fasst deren Ergebnisse dann wieder zusammen
- Kosten
 - Small EC2 Unit \$0.095/h + \$0.015/h MapReduce



Überblick Gesamtpreis Small System Webserver



Instances	Operating System	Type
2	Linux/OpenSolaris	Small

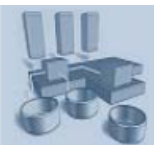
Storage:	230 GB/Month
Data Transfer In:	10 GB/Month
Data Transfer Out:	200 GB/Month
PUT/COPY/POST/LIST Requests:	2000 Requests
GET and Other Requests:	1000 Requests

Monthly Volume:	1000 GB/Month
Average Object Size:	30 KB

Edge Location Traffic Distribution:

United States	30 %
Europe	30 %
Hong Kong	25 %
Japan	15 %





Allgemeiner Aufbau der Yahoo Cloud

Edge content services

Edge caching

Edge routing

Messaging

Message Broker

Batch processing

Hadoop

HDFS

Operational Storage

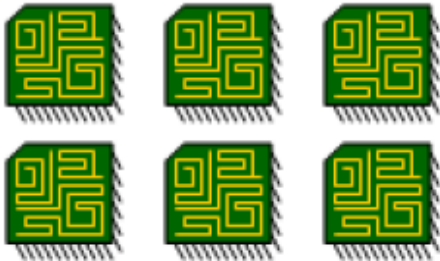
MOBStor

Sherpa

Provisioning

Provisioning service

VMs

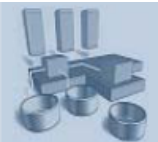


	A	B	C	D
1	5	A	T	
2	6	A	F	
3	2	B	T	



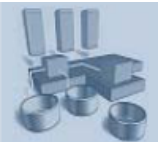
Edge Services

- Routing & Caching
- Vergleichbar mit Amazons CloudFront
- Geografisch orientierte Zwischenspeicherung von Inhalten
- Routing-Optimierung



Messaging

- Nachrichtensystem für Kommunikation zwischen Cloudkomponenten
- Beispiel: Updatebenachrichtigung für Edge Services (Update oder Fehlerfall)



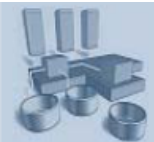
Hadoop

- System für Parallelverarbeitung großer Datenmengen oder komplexer Algorithmen
- Hat bei Yahoo! Zugriff auf Sherpa und MObsStor
- Eingesetzt für semantisch optimierte Werbung, Spamreduzierung oder allgemeine Suchalgorithmen bei Yahoo! Search
- Yahoo! WebMap → Hadoop auf 10.000 Core-Linux-System mit 300TB komp. Daten
- Ergebnis: 33% Leistungssteigerung



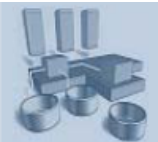
MObStor

- Virtualisierung der Hardware von SAN Systemen
- Bereitsstellung von Datenspeicher mit einfacher Zugriffsschnittstelle
- Zugriff auf Daten via URL
- Unterstützt Streaming und Caching von Daten
- Eingesetzt bei Yahoo! Mail, Bilder, Maps, Werbung



Sherpa (PNUTS)

- Einfaches relationales Datenbanksystem
- Reguläre Tabellen, Datensätze und Attribute
- Auch „blobs“ möglich, kann beliebige Inhalte und Strukturen beinhalten
- Einfache Anfrageoperationen
- Komplexere Anfragen, JOINS aus Performancegründen nicht unterstützt
- Zugriff via Hadoop möglich



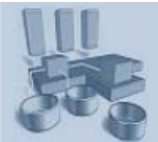
Vergleich

Amazon AWS	Yahoo! Cloud	Funktion
Amazon EC2	VMs (Provisioning Service)	Bereitstellen einzelner virtueller Rechner
Amazon S3	MObStor	Simpler Filehosting-Service, mit Vorteil des schnellen Zugriffs und großer Dateien
Amazon RDS / Simple DB	Sherpa	Datenbanksystem
Amazon Elastic Map Reduce	Hadoop + HDFS	Parallelverarbeitung komplexer Prozesse
Amazon SQS	Message Broker	Simpler Nachrichtenaustausch zwischen Instanzen
Amazon Cloud Front	Edge Caching / Edge Routing	Geografisch orientiertes Cachen von Daten zur Optimierung von Zugriffszeiten



Fazit

Ähnlichkeiten	Unterschiede
Infrastrukturebene	On Demand Instances / Capacity
Modulare Architektur	Betriebszweck



Literatur

- [AWS EC2]: <http://aws.amazon.com/ec2/>, aufgerufen 27.12.2009.
- [AWS SEC]: <http://aws.amazon.com/security>, aufgerufen 27.12.2009.
- [BRANTNER 2008]: M. Brantner, D. Florescu et al., *Building a database on S3.*, SIGMOD 2008.
- [Cooper 2008]: B. F. Cooper, R. Ramakrishnan, U. Srivastava, et al., *PNUTS: Yahoo!'s hosted data serving platform.* VLDB, 1(2):1277–1288, 2008.
- [COOPER 2009]: B. Cooper, E. Baldeschwieler et al., *Building a Cloud for Yahoo!*, IEEE DE Bulletin, 2009.
- [GROSSMAN, GU 2009]: Robert L. Grossman, Yunhong Gu, *On the Varieties of Clouds for Data Intensive Computing*, <http://sites.computer.org/debull/A09mar/grossman.pdf>, aufgerufen 27.12.2009.
- [HADOOP]: <http://hadoop.apache.org/>, aufgerufen 27.12.2009.
- [YAHOO DEV]: <http://developer.yahoo.net/blog/archives/2009/07/mobstor.html>, aufgerufen 27.12.2009.



Vielen Dank für ihre Aufmerksamkeit!

Fragen?

