

Ausgewählte Kapitel der Systemsoftware: Cloud Computing

Tobias Distler, Timo Hönig

Friedrich-Alexander-Universität Erlangen-Nürnberg
Lehrstuhl Informatik 4 (Verteilte Systeme und Betriebssysteme)
www4.informatik.uni-erlangen.de

Sommersemester 2010



Grundkonzepte

- **Virtualisierung**
 - Erfunden in den 60ern (IBM)
 - Populär seit den 90ern
 - Xen
 - VMWare
 - ...
 - Mehrere (heterogene) Betriebssysteme parallel auf einem Rechner
 - Schlüsseltechnologie für die Cloud
- Bedeutung für die Cloud: **Technologie zur Serverkonsolidierung**
 - Geringere Kosten für Rechenzentren
 - Beispiel [1]: "a major engineering services company"
 - Ohne Virtualisierung: 2-3 % Auslastung, Größe: > 3.000 m²
 - Mit Virtualisierung: 80 % Auslastung, Größe: < 100 m²
 - Andererseits: Freie Kapazitäten werden für andere nutzbar



Was ist die Cloud?

- Die Meinungen gehen auseinander...
 - Paradigmenwechsel
 - "Probably more than anything we've seen in IT since the invention of timesharing or the introduction of the PC, cloud computing represents a **paradigm shift** in the delivery architecture of information services."* (aus [1])
- oder
 - Alter Wein in neuen Schläuchen?
- Es existiert keine einheitliche Definition
- Gemeinsame Basis aller Definitionsversuche (nach [4])
 - Virtualisierung
 - Data-Center
 - Verteiltes Rechnen
 - Utility-Grid



Grundkonzepte

- **Data-Center**
 - Strukturwandel bei Rechenzentren
 - Typisches Data-Center während der „Dot-Com“-Ära
 - Einige wenige sehr teure Hochleistungsrechner
 - Eng vermaschtes Netzwerk
 - Google's Ansatz
 - Commodity-Hardware in enormer Stückzahl
 - Vergleichbar dünnes, dafür weitspannendes Netzwerk
- Bedeutung: **Fast unbegrenzte Speicher- und Rechenkapazität**
 - Fallende Preise für Ressourcenverbrauch
 - Beispiel (ebenfalls aus [1]):
 - Eigenes Rechenzentrum: 3,75 \$ pro Gigabyte Speicher
 - Amazon: 0,10-0,15 \$ pro Gigabyte Speicher



■ Verteiltes Rechnen

- Große verteilte Projekte sind keine Erfindung der letzten Jahre
 - SETI@home
 - Folding@home
 - ...
- Seit 2006: Open-Source-Projekt Hadoop (Apache Foundation)
 - MapReduce: effizientes paralleles Rechnen
 - HDFS: verteiltes Dateisystem
 - ...

■ Bedeutung: **Framework zur Entwicklung weitverteilter Projekte**

- Mittel, um das Potenzial der Cloud auch tatsächlich zu nutzen
- Beispiel [3]: Facebook
 - Hive (Erweiterung von Hadoop)
 - Analysen auf 700 TB pro Tag



■ Utility-Grid

- Prinzip: Inhaber von Rechenzentren verkaufen Rechenzeit
- Vor der Cloud: Mainframe-Rechner und Supercomputer
 - Universitäten
 - Andere Forschungseinrichtungen
- In der Cloud: Rechenzentren von Firmen mit freien Kapazitäten
 - Google (AppEngine)
 - Amazon (EC2)
 - Microsoft (Windows Azure)

■ Bedeutung: **Abrechnung nach effektivem Ressourcenverbrauch**

- Keine Notwendigkeit mehr für den maximalen Bedarfsfall vorzusorgen
- Bessere Korrelation zwischen Kosten und wirtschaftlichem Erfolg



Probleme und offene Fragen

■ Sicherheit und Verfügbarkeit

- Was wenn die Cloud-Infrastruktur ausfällt?
- Beispiel (Februar 2008) [2]: 3-stündiger Ausfall von Amazon EC2

■ Vertraulichkeit

- Erhöhter Grad an Vertrauen notwendig
- Verschlüsselung spielt (zur Zeit) noch keine Rolle

■ Problem des "Vendor lock-in"

- Wie einfach ist es den Anbieter zu wechseln?
- Es existieren keine einheitlichen Standards

■ Versteckte Kosten

- Mehrkosten für Netzwerktransfervolumen
- Zusätzliche Latenzen



Themen

■ Virtualisierung

- Xen
- KVM
- Hyper-V

■ Infrastructure as a service (IaaS)

- Amazon EC2
- Eucalyptus

■ Spezialisierte Ausführungsunterstützung im Kontext von IaaS

- Specialized Execution Environments
- Libra

■ Verteilte Dateisysteme für Cloud-Anwendungen

- Google File System
- Cassandra (Facebook)

■ Skalierbarer und zuverlässiger Zugriff auf feingranulare Daten: Dynamo (Amazon)



- Programmierunterstützung im Kontext von Cloud Computing
 - MapReduce, Disco
 - BOOM
- Status Cloud Computing
 - Are Clouds Ready for Large Distributed Applications?
 - The Future of Cloud Computing
- Zuverlässige Konfiguration von Cloud-Systemen: Yahoo ZooKeeper
- Vertrauenswürdigkeit und Sicherheit in Cloud-Infrastrukturen
 - Towards Trusted Cloud Computing
 - A Privacy Manager for Cloud Computing
- Skalierbare Byzantinische Fehlertoleranz: UpRight
- Aktuelle Forschung am Lehrstuhl 4
 - TLOUDS
 - ReFIT



- Ausarbeitung
 - (Mindestens) 6 Seiten, ACM-Style
 - **Wichtig: Bezug zur Cloud herstellen**
 - Erste Version spätestens zwei Wochen vor dem Vortrag
 - Finale Version eine Woche vor dem Vortrag
- Vortrag
 - 40 Minuten + anschließende Diskussion
 - Erste Version der Vortragsfolien eine Woche vor dem Vortrag
 - Vorbereitung: Vortrag (mindestens einmal) zur Probe halten
- Aktive Teilnahme
 - Vorbereitung anhand der Ausarbeitung
 - Anwesenheit
 - Beteiligung an der Diskussion



-  Mache Creeger.
Cloud computing: An overview.
ACM Queue, 7(5):3–4, 2009.
-  Judith Myerson.
Cloud computing versus grid computing.
IBM Web Development Technical Library, 2009.
-  Ashish Thusoo, Raghotham Murthy, Joydeep Sen Sarma, Zheng Shao, Namit Jain, Prasad Chakka, Suresh Anthony, Hao Liu, and Ning Zhang.
Hive – a petabyte scale data warehousing using Hadoop.
In *Proc. of the 26th Int. Conference on Data Engineering Conference*, 2010.
-  Aaron Weiss.
Computing in the clouds.
netWorker, 11(4):16–25, 2007.

