

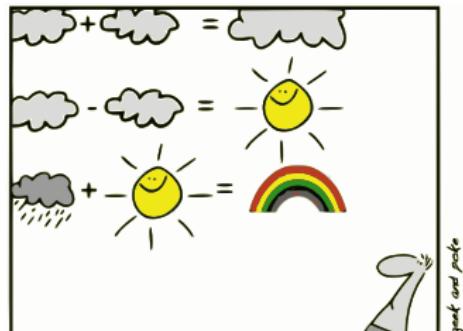
# Technische Aspekte des Cloud Computing

Dr. Christian Baun

wolkenrechnen@gmail.com

16.3.2012

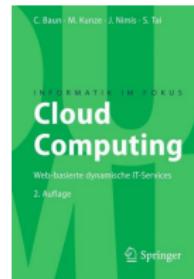
# Cloud Computing – Definitionen



SIMPLY EXPLAINED - PART 17:  
CLOUD COMPUTING

- Diverse Definitionen existieren
- Nicht alle sind hilfreich
- Gute Definitionen sind umfangreich

„Unter Ausnutzung virtualisierter Rechen- und Speicherressourcen und moderner Web-Technologien stellt Cloud Computing skalierbare, netzwerk-zentrierte, abstrahierte IT-Infrastrukturen, Plattformen und Anwendungen als on-demand Dienste zur Verfügung. Die Abrechnung dieser Dienste erfolgt nutzungsabhängig.“



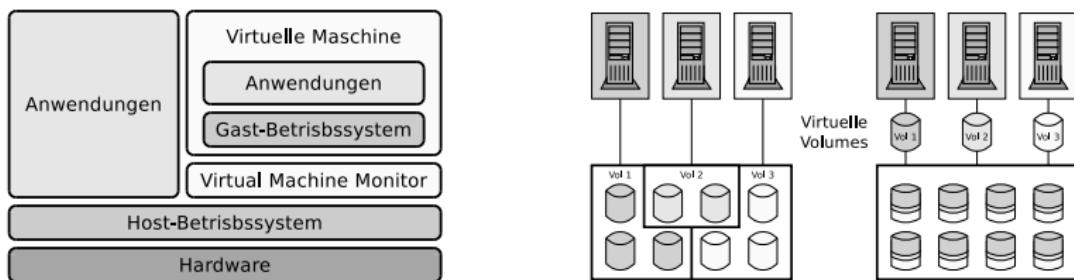
Alles klar?

*„Unter Ausnutzung virtualisierter Rechen- und Speicherressourcen und moderner Web-Technologien stellt Cloud Computing skalierbare, netzwerk-zentrierte, abstrahierte IT-Infrastrukturen, Plattformen und Anwendungen als on-demand Dienste zur Verfügung. Die Abrechnung dieser Dienste erfolgt nutzungsabhängig.“*

- **Teil 1:** Grundlegende Technologien – Fundament des Cloud Computing
  - **Virtualisierung**
  - **Web-Services**
- **Teil 2:** Dienste und deren Eigenschaften
  - **Infrastrukturdienste, Plattformdienste, Softwaredienste**
  - **skalierbar** ⇒ „elastisch“
  - **netzwerk-zentriert** ⇒ Dienste/Ressourcen über das Netz erreichbar
  - **abstrahiert** ⇒ unabhängig von der konkreten Hardware
  - **on-demand** ⇒ zeitnahe Erfüllung von Anforderungen bzw. Nachfragen
  - **verbrauchsabhängige Abrechnung**

# Grundlegende Technologien – Virtualisierung

- Erlaubt eine abstrakte, logische Sicht auf physische Ressourcen
  - Server, Speicher, Netzwerke
- Isoliert (versteckt) die physische Hardware
- Physische Hardware kann dank Virtualisierung...
  - ① transparent gemeinsam verwendet werden und
  - ② heterogene Hardwareressourcen können zu einem homogenen Ressourcenpool zusammengefügt werden

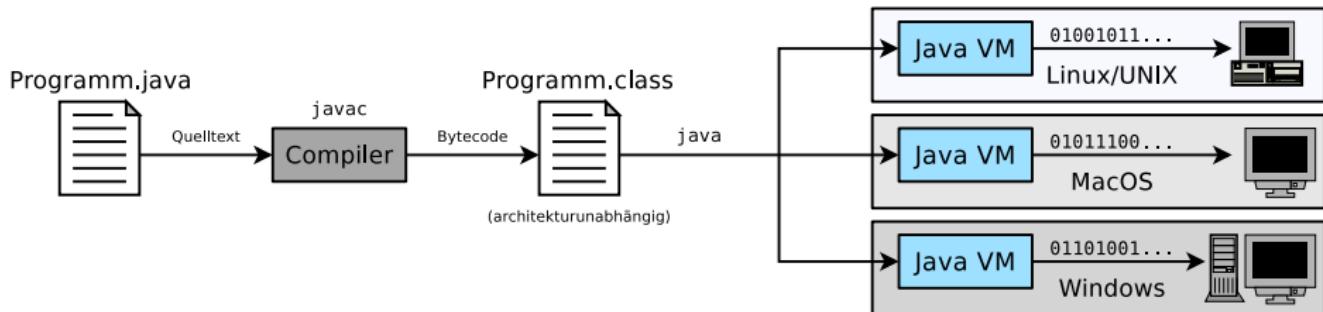


- Einige Vorteile
  - Serverkonsolidierung  $\implies$  Kostensenkung
  - Vereinfachte (kurzfristige) Bereitstellung
  - Flexibilität (Verschiedene Betriebssysteme auf einer Hardware)

- Virtualisierung ist kein neues Konzept
  - Einführung bereits vor ca. 40 Jahren bei Großrechnern
  - IBM stellte in den 1960er-Jahren die Virtual Machine Facility/370 vor
    - Auf dieser Plattform wurde Mehrbenutzerbetrieb realisiert, indem mehrere Einzelbenutzerbetriebinstanzen in virtuellen Maschinen ausführten wurden
    - Jede VM ist eine vollständige Nachbildung der physischen Hardware
- Es existieren unterschiedliche **Virtualisierungskonzepte**:
  - Partitionierung
  - Hardware-Emulation
  - **Anwendungsvirtualisierung**
  - **Vollständige Virtualisierung** (Virtueller Maschinen-Monitor)
  - **Paravirtualisierung** (Hypervisor)
  - Hardware-Virtualisierung
  - Betriebssystem-Virtualisierung bzw. Container bzw. Jails
  - **Speichervirtualisierung** (SAN)
  - **Netzwerkvirtualisierung** (VLAN)
  - ...

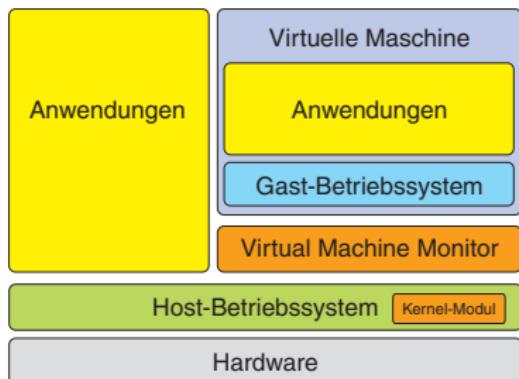
# Anwendungsvirtualisierung

- Anwendungen laufen in einer virtuellen Umgebung, die alle Komponenten bereitstellt, die die Anwendung benötigt
  - Die VM liegt zwischen der Anwendung und dem Betriebssystem
- Beispiel: Java Virtual Machine (JVM)
  - Teil der Java-Laufzeitumgebung (JRE)
  - Verantwortlich für die Ausführung des Java-Bytecodes
  - Ist für Java-Programme die Schnittstelle zum Betriebssystem
- Vorteil: Plattformunabhängigkeit
- Nachteil: Geringere Ausführungsgeschwindigkeit gegenüber nativer Programmausführung

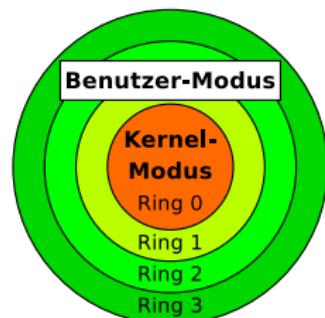


# Vollständige Virtualisierung (1)

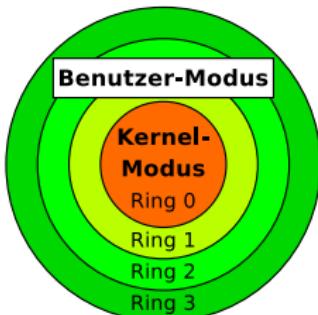
- Vollständige Virtualisierungslösungen bieten einer VM eine vollständige, virtuelle PC-Umgebung inklusive eigenem BIOS
  - Jedem Gastbetriebssystem steht ein eigener virtueller Rechner mit virtuellen Ressourcen wie CPU, Hauptspeicher, Laufwerken, Netzwerkkarten, usw. zur Verfügung
- Basiert auf einem **Virtueller Maschinen-Monitor** (VMM)
  - Der VMM heißt auch **Typ-2-Hypervisor**
  - Läuft *hosted* als Anwendung unter dem Host-Betriebssystem
- VMM verteilt die Hardwareressourcen des Rechners an die VMs
- Teilweise emuliert der VMM Hardware, die nicht für den gleichzeitigen Zugriff mehrerer Betriebssysteme ausgelegt ist
  - Beispiel: Netzwerkkarten
  - Emulation populärer Hardware vermeidet Treiberprobleme



- Virtualisierung in der x86-Architektur basiert auf dem Schutzkonzept der Ringe
- x86-kompatible CPUs enthalten 4 Privilegienstufen zum Speicherschutz, um die Stabilität und Sicherheit zu erhöhen
- Ein Prozess kann immer nur in einem Ring ausgeführt werden und sich nicht selbstständig aus diesem befreien
- Verbreitete Betriebssysteme nutzen nur Ring 0 und 3
- Ausnahme: OS/2 nutzt Ring 2 für Anwendungen, die auf Hardware und Eingabe-/Ausgabeschnittstellen zugreifen dürfen (z.B. Grafiktreiber)

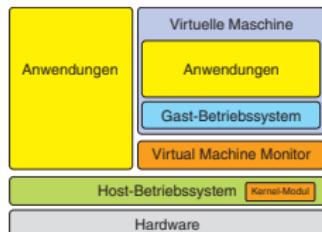
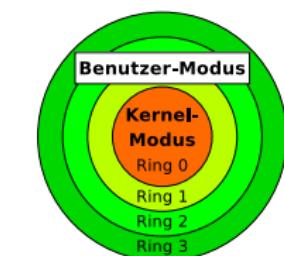


- Nur Prozesse in Ring 0 haben vollen Zugriff auf die Hardware und dürfen den vollständigen Befehlssatz der CPU nutzen
  - Ring 0 ist der Kernel-Bereich (Kernel-Space)
    - Hier läuft nur der Betriebssystemkern und zum Start des Betriebssystems nötige Hardwaretreiber
  - Ring 3 ist der Benutzerbereich (User-Space)
    - Hier laufen die Anwendungen
- Ruft ein Prozess in einem weniger privilegierten Ring eine privilegierte Operation auf, erzeugt die CPU eine Ausnahme (Exception)
  - Die Exception wird im benachbarten privilegierteren Ring abgefangen und dort behandelt
  - Ausnahmen, die nicht abgefangen werden können, verursachen eine allgemeine Schutzverletzung (General Protection Fault)
    - Der aufrufende Prozess stürzt ab
    - Handelt es sich bei dem Prozess um den Kernel, stürzt das System ab



# Vollständige Virtualisierung (2)

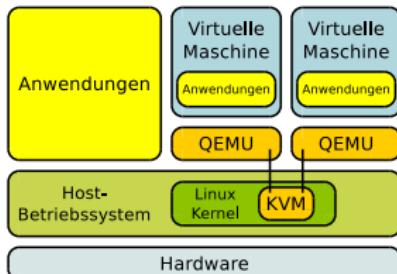
- Vollständige Virtualisierung nutzt die Tatsache, dass x86-Systeme nur 2 von 4 möglichen Privilegienstufen verwenden
  - Der VMM befindet sich in Ring 0 auf der Ebene des Betriebssystemkerns des Host-Betriebssystems und hat vollen Zugriff auf die Hardware
  - Die VMs befinden sich in einem der weniger privilegierten Ringe 1 oder 2
- Der VMM stellt für jede denkbare Ausnahme eine Behandlung zur Verfügung, die die privilegierten Operationen der Gastbetriebssysteme abfängt, interpretiert und ausführt
- Der VMM stellt sicher, dass die VMs nur über den Umweg des VMM Zugriff auf die Hardware erhalten
  - Kontrollierter Zugriff auf die gemeinsam genutzten Systemressourcen ist gewährleistet



- Vorteile:
  - Kaum Änderungen an Host- und Gast-Betriebssystemen erforderlich
  - Zugriff auf die wichtigsten Ressourcen wird nur durchgereicht  
    ⇒ Fast native Verarbeitungsgeschwindigkeit der Gast-Betriebssysteme
  - Jedes Gast-Betriebssystem hat seinen eigenen Kernel  
    ⇒ Hohe Flexibilität
- Nachteile:
  - Wechsel zwischen den Ringen erfordern einen Kontextwechsel  
    ⇒ Jeder Kontextwechsel verbraucht Rechenzeit
  - Fordert eine Anwendung im Gast-Betriebssystem die Ausführung eines privilegierten Befehls an, liefert der VMM eine Ersatzfunktion und diese weist die Ausführung des Befehls über die Kernel-API des Host-Betriebssystems an  
    ⇒ Geschwindigkeitseinbußen
- Beispiele: VMware Server/Workstation/Fusion, VirtualBox, KVM

# Kernel-based Virtual Machine (KVM)

- KVM ist als Modul direkt im Linux-Kernel integriert
  - KVM-Basismodul: `kvm.ko`
  - Hardwarespezifische Module: `kvm-intel.ko` und `kvm-amd.ko`
- Nach dem Laden der Module arbeitet der Kernel selbst als VMM
- KVM kann nur mit CPUs mit Hardwarevirtualisierung arbeiten
  - Dadurch braucht KVM weniger Quellcode als andere Lösungen
- Neben den Kernelmodulen enthält KVM den Emulator QEMU
  - CPU-Virtualisierung stellt der Prozessor bereit (Intel VT oder AMD-V)
  - Der Speicher wird durch KVM virtualisiert
  - E/A wird durch einen QEMU-Prozess pro Gastsystem virtualisiert



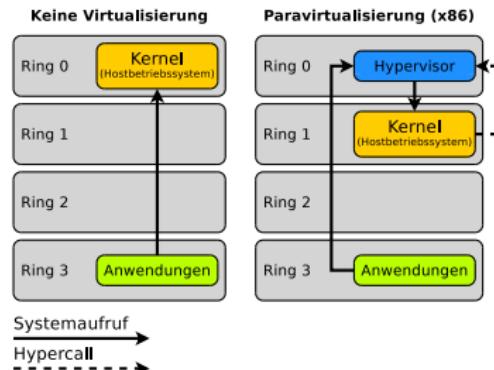
- Bei Paravirtualisierung wird keine Hardware virtualisiert oder emuliert
  - Gast-Betriebssystemen steht nur eine API zur Verfügung
- Virtuell gestartete Betriebssysteme verwenden eine abstrakte Verwaltungsschicht, den **Hypervisor**, um auf die physischen Ressourcen wie Speicher, Ein-/Ausgabegeräte und Netzwerkinterfaces zuzugreifen
  - Der Hypervisor ist ein minimales **Metabetriebssystem**, das die Hardwareressourcen unter den Gastsystemen verteilt, so wie ein Betriebssystem dieses unter den laufenden Prozessen tut
    - Der Hypervisor läuft *bare metal*
    - Der Hypervisor heißt auch **Typ-1-Hypervisor**
  - Ein Metabetriebssystem ermöglicht den unabhängigen Betrieb unterschiedlicher Anwendungen und Betriebssysteme auf einer CPU
- Das Host-Betriebssystem läuft nicht im privilegierten Ring 0, sondern im weniger privilegierten Ring 1
  - Ein Host-Betriebssystem ist wegen der Gerätetreiber zwingend nötig

# Paravirtualisierung (2)

- Das Betriebssystem läuft nicht mehr in Ring 0, sondern in Ring 1
  - Darum kann der Kernel keine privilegierten Anweisungen ausführen
  - Lösung: Der Hypervisor stellt **Hypercalls** zur Verfügung

- Hypercalls sind vergleichbar mit Systemaufrufen (System Calls)

- Die Interrupt-Nummern sind verschieden
- Fordert eine Anwendung die Ausführung eines Systemaufrufs an, wird eine Ersatzfunktion im Hypervisor aufgerufen
- Der Hypervisor weist die Ausführung des Systemaufrufs über die Kernel-API des Betriebssystems an

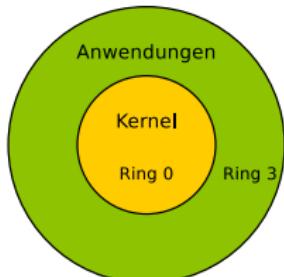


- Nachteil: Erweiterung des Kernels um die Hypercall-Funktionalität macht eine Modifikation der Betriebssysteme notwendig
- Vorteil: Abfangen und Prüfen aller Systemaufrufe durch den Hypervisor führt nur zu geringen Geschwindigkeitseinbußen
- Beispiele: Xen, Citrix Xenserver, Virtual Iron, VMware ESX Server

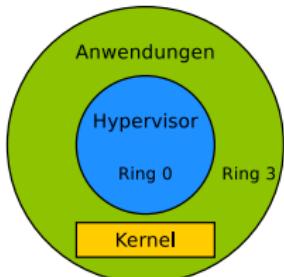
# Problem: x86-64-Architektur

- Die x86-64-Architektur (z.B. IA64) verzichtet auf die Ringe 1 und 2

Keine Virtualisierung



Paravirtualisierung (IA64)

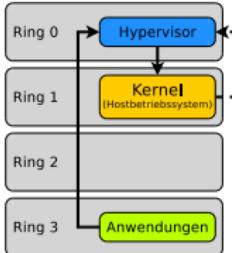


- Der Hypervisor befindet sich wie bei der x86-32-Architektur in Ring 0
- Der Kernel wird in Ring 3 zu den Anwendungen verschoben

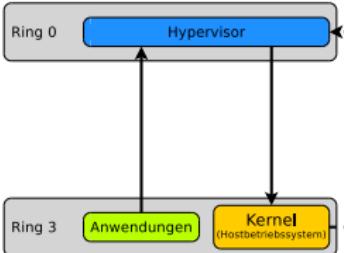
Keine Virtualisierung



Paravirtualisierung (x86)



Paravirtualisierung (IA64)

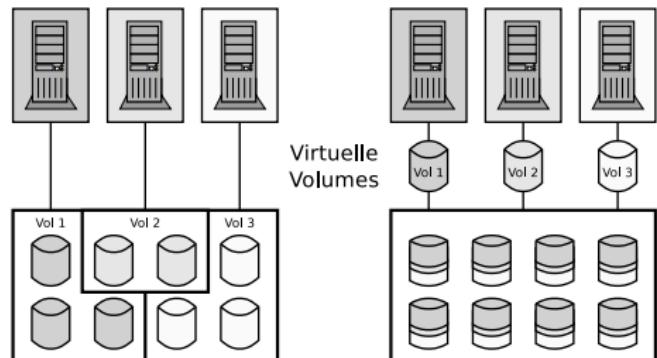


- Der Betrieb der Hardwaretreiber und Anwendungen in einem Ring ist tendenziell unsicher

- Paravirtualisierung erfordert angepasste Gastsysteme
  - Hypervisor läuft *bare metal* anstatt eines klassischen Betriebssystems
  - Typ-1-Hypervisor
  - Hypervisor läuft in Ring 0 und hat vollen Zugriff auf die Hardware
  - Beispiele: VMware ESX(i), Xen, Microsoft Hyper-V
- Bei Vollvirtualisierung kann man unveränderte Systeme betreiben
  - VMM läuft *hosted* als Anwendung unter einem klassischen Betriebssystem wie Linux oder Windows
  - VMM = Typ-2-Hypervisor
  - VMM läuft in Ring 3 auf der Ebene der Anwendungen
  - Beispiele: VMware Workstation, KVM, Oracle VirtualBox, Parallels

- Aktuelle Prozessoren von Intel und AMD enthalten Erweiterungen zur Hardware-Virtualisierung
  - Dank Hardware-Virtualisierung können unmodifizierte Betriebssysteme (z.B. Windows) mit Xen oder VMware ESX laufen
  - Die Lösungen von Intel und AMD sind ähnlich aber inkompatibel
- AMD erweitert seit Juni 2006 seine AMD64 CPUs um den sogenannten Secure-Virtual-Machine-Befehlssatz (**SVM**)
  - Die Lösung heißt **AMD-V** und war vorher als **Pacifica** bekannt
- Die Lösung von Intel heißt **VT-x** für IA32-CPUs und **VT-i** für Itanium
  - Intels Lösung lief vormals unter dem Stichwort **Vanderpool**
- Xen unterstützt ab Version 3 Hardware-Virtualisierung
- Auch Windows Server 2008 (Hyper-V) nutzt Hardwarevirtualisierung

- Kern der Neuerung ist eine Überarbeitung der Privilegienstruktur
  - Die neuen Befehle bei AMD und Intel bieten VMs eine Erweiterung zu den bereits beschriebenen Privilegienstufen Ring 0 und Ring 3
- Die Ringstruktur wurde durch eine Erweiterung von Ring 0 um eine Ebene, die neue Hypervisor-Schicht, ergänzt
  - Die Ebene wird als Root-Betriebsmodus oder Ring -1 bezeichnet
  - Der Hypervisor bzw. VMM läuft im Root-Betriebsmodus (Ring -1) und besitzt jederzeit die volle Kontrolle über die CPU und die Ressourcen, da damit ein höheres Privileg als Ring 0 implementiert ist
- VMs steht der gewohnte Zugriff auf Ring 0 zur Verfügung
  - Diese VMs heißen HVM (Hardware Virtual Machine)
- Vorteile:
  - Gastbetriebssysteme müssen nicht angepasst werden
    - Auch proprietäre Betriebssysteme (z.B. Windows) laufen als Gastsysteme
  - Der Kernel läuft nicht wie bei der Paravirtualisierung mit den Anwendungen auf einer Privilegienstufe



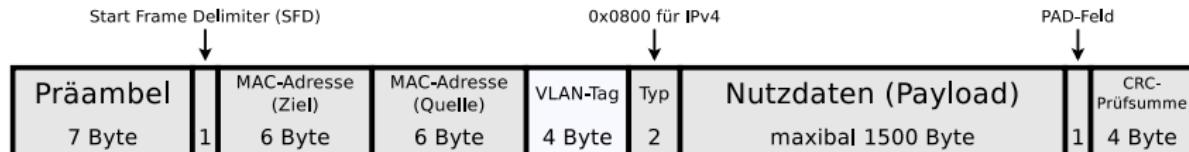
- Speicher wird in Form virtueller Laufwerke den Benutzern zur Verfügung gestellt
- Logischer Speicher wird vom physischen Speicher getrennt

- Vorteile:
  - Nutzer sind nicht an physische Grenzen der Speichermedien gebunden
  - Physischen Speicher umstrukturieren/erweitern stört die Nutzer nicht
  - Redundantes Vorhalten erfolgt transparent im Hintergrund
  - Besserer Auslastungsgrad, da der verfügbare physische Speicher effektiver auf die vorhandenen Benutzer aufgeteilt werden kann
- Nachteil: Professionelle Lösungen sind teuer
- Bekannte Anbieter: EMC, HP, IBM, LSI und SUN/Oracle

- Basiert auf virtuellen lokalen Netzen (**Virtual Local Area Networks**)
- Verteilt aufgestellte Geräte kann man via VLAN in einem einzigen logischen Netz zusammenfassen
  - Nützlich bei der Konzeption der IT-Infrastruktur verteilter Standorte
  - Mit VLANs trennt man physische Netze in logische Teilnetze
    - VLAN-fähige Switches leiten Pakete eines VLAN nicht in andere VLANs
    - Ein VLAN bildet ein nach außen isoliertes Netz über bestehende Netze
  - Zusammengehörende Systeme und Dienste können in einem eigenen Netz konsolidiert werden, um die übrigen Netze nicht zu beeinflussen
  - Ein VLAN bildet ein Netz über nicht vertrauenswürdige Netze und kann so helfen, verteilte Standorte in eine virtuelle Infrastruktur zu integrieren
- Nachteil: Steigender Aufwand für die Netzwerkadministration

- Es existieren unterschiedliche Typen von VLANs
- ① Der älteste Standard ist **statisches VLAN**
- Die Anschlüsse (Ports) eines Switches werden in mehrere logische Switches unterteilt
  - Jeder Anschluss ist fest einem einzigen VLAN zugeordnet oder verbindet unterschiedliche VLANs
  - Statische VLANs sind schlecht automatisierbar
- ② Aktuell ist das **paketbasierte, dynamische VLAN** nach IEEE 802.1Q
- Die Netzwerkpakete enthalten eine spezielle VLAN-Markierung (*Tag*)
  - Vor der Standardisierung 1998 durch das IEEE Konsortium existierten diverse proprietäre paketbasierte VLAN Lösungen
    - Cisco Inter-Switch Link (ISL)
    - Virtual LAN Trunk (VLT) von 3Com
  - Diese Lösungen wurden im herstellerübergreifenden Standard IEEE 802.1Q zusammengefasst
  - Dynamische VLANs können mit Hilfe von Skripten rein softwaremäßig erzeugt, verändert und wieder entfernt werden

# Ethernet mit VLAN-Markierung nach IEEE 802.1Q

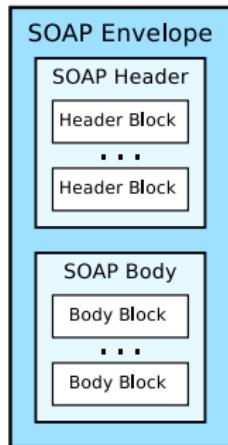


- Der IEEE 802.1Q-Standard definiert die VLAN-Markierung (*Tag*)
- Die eingefügte VLAN-Markierung umfasst 32 Bits
  - Die Protokoll-ID (16 Bits) hat immer den Wert 0x8100
  - 3 Bit repräsentieren die Priorität
    - Eine von 8 möglichen Prioritäten kann angegeben werden
    - Damit ist eine Priorisierung bestimmter Daten (z.B. VoIP) möglich
  - Das kanonische Format (1 Bit) bestimmt das höchstwertige Bit der MAC-Adressen (0 = Ethernet, 1 = Token Ring)
  - 12 Bits enthalten die ID des VLAN, zu dem das Netzwerkpacet gehört



- Verteilte Systeme integrieren häufig heterogene Ressourcen
  - Theoretische Ausbreitung dieser Ressourcen ist weltweit
- Nachteile von Verbindungen über weite Strecken im Vergleich zu LANs
  - Hohe Antwortzeiten
  - Geringe Datenübertragungskapazitäten
  - Potentiell unzuverlässige Verbindungen
- Web-Services ermöglichen **schwach gekoppelte, asynchrone** und **nachrichtenbasierte** Kommunikation auf Basis von HTTP und XML
- Die populärsten Anwendungsmöglichkeiten für Web Services sind
  - Entfernte Funktionsaufrufe (Remote Procedure Calls)
  - **SOAP** (früher: *Simple Object Access Protocol*)
  - **REST** (*REpresentational State Transfer* – Zustandsrepräsentationsübertragung)

- SOAP-Nachrichten basieren auf der Auszeichnungssprache XML
  - Meist werden SOAP-Nachrichten im Body einer HTTP POST-Anforderung an eine URL geschickt



```
<?xml version="1.0" encoding="UTF-8" ?>
<env:Envelope xmlns:env="http://www.w3.org/2001/09/soap-envelope">
  <env:Header>
    <n:alertcontrol xmlns:n="http://example.org/alertcontrol">
      <n:priority>1</n:priority>
      <n:expires>2001-06-22T14:00:00-05:00</n:expires>
    </n:alertcontrol>
  </env:Header>
  <env:Body>
    <m:alert xmlns:m="http://example.org/alert">
      <m:msg>Mary um 14 Uhr von der Schule abholen</m:msg>
    </m:alert>
  </env:Body>
</env:Envelope>
```

Quelle: Tanenbaum, van Steen: *Verteilte Systeme*. Pearson Studium (2008)

- Die Nachricht sendet einen Text an einen Web Service
- Die Nachricht hat eine bestimmte Priorität (1) und wird verworfen, wenn Sie nach 14 Uhr beim Web Service ankommt

- RESTful Web Services spricht man über die HTTP-Schnittstelle an
- Zustandslose Kommunikation
  - Jede HTTP-Nachricht enthält alle nötigen Informationen, um die Nachricht zu verstehen
  - Server hält keine Zustands- bzw. Sitzungsinformation über den Client vor
  - Jede Anfrage ist eine von anderen Anfragen unabhängige Transaktion
- 4 HTTP-Methoden genügen, um mit Ressourcen zu arbeiten

HTTP	CRUD-Aktionen	SQL	Beschreibung
PUT	Create	INSERT	Ressource erzeugen oder deren Inhalt ersetzen
GET	Read/Retrieve	SELECT	Ressource bzw. deren Repräsentation anfordern
POST	Update	UPDATE	Einer Ressource etwas hinzufügen
DELETE	Delete/Destory	DELETE	Ressource löschen

- Weitere sinnvolle HTTP-Methoden:
  - HEAD – fordert Metadaten zu einer Ressource an
  - OPTIONS – fragt an, welche Methoden auf einer Ressource möglich sind

„Unter Ausnutzung virtualisierter Rechen- und Speicherressourcen und moderner Web-Technologien stellt Cloud Computing skalierbare, netzwerk-zentrierte, abstrahierte **IT-Infrastrukturen, Plattformen und Anwendungen** als on-demand **Dienste** zur Verfügung. Die Abrechnung dieser Dienste erfolgt nutzungsabhängig.“

- Cloud Computing ist ein Sammelbegriff für verschiedene Dienste

Was versteht man unter **Diensten**?

- Dienstleistung aus dem Bereich der Informationstechnologie (IT)
- Stellt ein Dienstanbieter (*Service Provider*) für einen oder mehrere Kunden (*Customer*) bereit
- Wird ähnlich einem Produkt angeboten
- Sollte über ein Service-Level-Agreement (SLA) definiert sein
- Erbringt ein externer Anbieter oder eine unternehmenseigene Abteilung
- Outsourcing = Auslagern von IT-Dienstleistungen

Wie unterscheidet man Cloud-Dienste **organisatorisch**?

- **Öffentliche Dienste** (Public Cloud)

- Anbieter und Kunden gehören unterschiedlichen Organisationen an  
    ⇒ Outsourcing
- Keine Kosten für Anschaffung, Betrieb und Wartung eigener Hardware
- Ressourcen sofort einsatzbereit und (fast) unbegrenzt verfügbar

- **Private Dienste** (Private Cloud)

- Anbieter und Benutzer gehören der gleichen Organisation an
- Kosten ähnlich einer nicht-Cloud-basierten Architektur

- **Hybride Dienste** (Hybrid Cloud)

- Öffentlich verfügbare und private Dienste werden gemeinsam verwendet
- Einsatzbeispiele:
  - Lastspitzen mit öffentlichen Diensten abfedern
  - Sicherheitskopien in öffentliche Dienste auslagern

Wie unterscheidet man Cloud-Dienste **funktional**?

- **Softwaredienste (SaaS)**

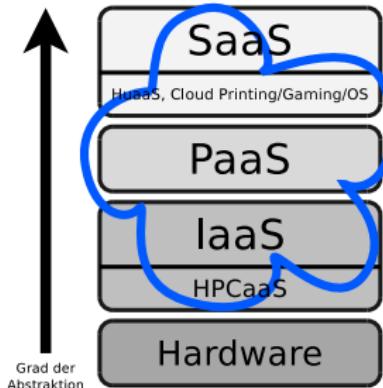
- Anbieter betreibt Webanwendungen
- Kunden brauchen nur einen Browser

- **Plattformdienste (PaaS)**

- Anbieter betreibt skalierbare Laufzeitumgebung(en)
- Kunden betreiben eigene Webanwendungen in der Infrastruktur des Dienstanbieters

- **Infrastrukturdienste (IaaS)**

- Anbieter betreibt physische Server
- Kunden betreiben VMs mit (fast) beliebigen Betriebssystemen und unveränderten Anwendungen
  - Kunden haben in ihren VMs Administratorenrechte und definieren die Firewall-Regeln

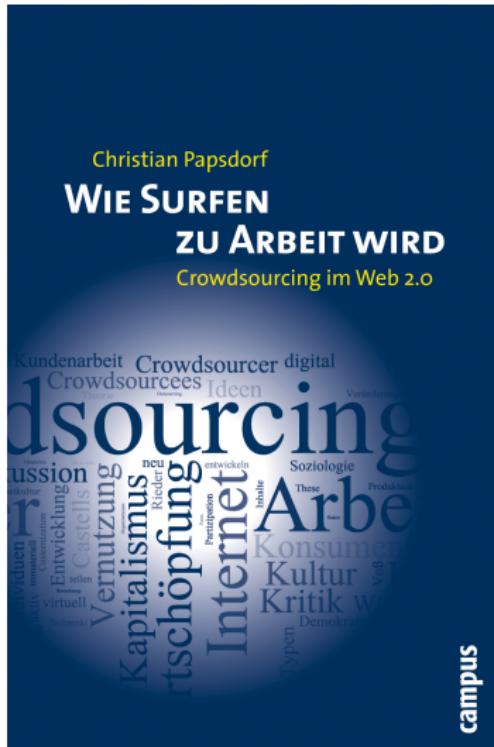




- (Freie) Lösungen zum Aufbau von Softwarediensten gibt es seit Jahren
  - Web-Server: Apache HTTP Server, nginx,...
  - Anwendungsserver für Webanwendungen: Apache Tomcat (Java), JBoss (Java), Zope (Python)
  - Skriptsprache für dynamische Webseiten: PHP
- Softwaredienste gibt es schon länger als den Begriff „Cloud Computing“
- Es existieren Untergruppen der Softwaredienste
  - Diese enthalten spezialisierte Dienste für bestimmte Anwendungen
    - Humans as a Service (HuaaS)
    - Cloud Printing
    - Cloud Gaming
    - Cloud-Betriebssysteme

- Prinzip des Crowdsourcing
- Menschliche Kreativität wird zu geringen Kosten oder als Spende Freiwilliger angeboten
- Interessant für...
  - Niedrig qualifizierte Tätigkeiten
  - Tätigkeiten, die ein Computer gar nicht, nur schlecht oder erst nach einer unverhältnismäßig hohen Entwicklungszeit erledigen könnte
- Einsatzmöglichkeiten sind u.a.
  - Bilderkennung
  - Subjektive (Produkt)-Bewertungen
  - Übersetzungen
  - (Produkt)-Zuordnungen zu (Produkt)-Kategorien
- Beispiele für öffentlich verfügbare HuaaS
  - Untersuchung von The Guardian zum britischen Spesenskandal 2009
  - GutenPlag, VroniPlag
  - Marktplatz für HuaaS: Amazon Mechanical Turk
- Im Bereich private Cloud spielt HuaaS keine Rolle

# Literaturempfehlung zu Crowdsourcing



- Christian Papsdorf. *Wie Surfen zu Arbeit wird*. Campus. 2009
- Konsumenten schreiben Bewertungen, entwickeln Ideen, erstellen Logos, usw.
- Diese wertschöpfenden Tätigkeiten sind von hoher ökonomischer Bedeutung
- Unternehmen nutzen die Internetkultur (Partizipation, Engagement, Selbstverwirklichung, ...), um die Konsumenten zumeist unentgeltlich für sich arbeiten zu lassen
- Warum machen die Konsumenten da freiwillig mit?

# Wettbewerb von Pril



- Am 1. April 2011 startete Henkel eine Crowdsourcing-Aktion
  - Trotz des Datums war es kein Scherz!
- Jeder konnte auf <http://mein.pril.de> Designvorschläge für die 600ml-Flasche entwerfen

- Zu gewinnen gab es Sachpreise
- Die beiden besten Designs sollten für kurze Zeit in den Handel kommen
- Über Facebook konnten die Benutzer für Ihren Favoriten stimmen
- Die Resonanz war groß – über 30.000 Vorschläge wurden eingereicht

# Wettbewerb von Pril – Ergebnis

- Nicht alle Etikettenvorschläge entsprachen der Erwartung von Henkel
- Zwei Vorschläge von Peter Breuer (ein professioneller Werbetexter) wurden schnell zu Favoriten
- Der Vorschlag *Hähnchen* war schnell auf Platz 1 und hatte mehrere tausend Stimmen Vorsprung
- Reaktion von Henkel: Regeländerung
  - Vorschläge mussten nun vorher von einer Jury bewertet und freigegeben werden
  - Erst danach konnte ein Vorschlag in die Abstimmung gehen
- Ergebnis: Empörung der Teilnehmer



# Wettbewerb von Pril – Manipulation des Ergebnis



- Es kam aber noch schlimmer
  - Empörte Kommentaren wurden von Henkel gelöscht
  - Die Stimmenzahl einiger Entwürfe wurde massiv von Henkel reduziert
  - Die Ergebnisse wurden laut Aussage von Henkel „bereinigt“
- Spätestens ab diesem Zeitpunkt ging die Sache durch die Presse  
⇒ PR-Gau (schlechte Öffentlichkeitsarbeit)

Jörg Breithut. **Virale Werbefallen – Pril schmeckt nach Hähnchen.** 12.04.2011  
<http://www.spiegel.de/netzwelt/web/0,1518,756532,00.html>

Es muss nicht unbedingt so enden...

Werde das neue

## Gesicht der OTTO Fanpage!

Anleitung    **Teilnehmer**    Mitmachen    Teilnahmebedingungen    Fee

Suche  sortieren nach  Zufall  Freunde finden

Stimme für deine Favoriten ab, indem du auf "Gefällt mir" klickst!

Brigitte Koblenz    Gefällt mir 16 Tsd

Klara Schweinfurt    Gefällt mir 11 Tsd

- Otto veranstaltete 2010 einen „Modelcontest“
    - Der Gewinner mit den meisten Stimmen wurde das neue Gesicht der Facebook-Fanseite
  - Der Gewinner gegen 48.488 Teilnehmer war „Der Brigitte“ (Sascha Mörs)
    - Ein 22-jähriger BWL-Student der FH Koblenz
  - Unglücklich war man bei Otto über die Entwicklung nicht
    - ca. 1,2 Millionen Stimmen wurde abgegeben
- ⇒ Gute Öffentlichkeitsarbeit

## Amazons mechanischer Türke

16.03.06 – Sam Williams

Schlagwörter: KI, Robotik



Amazon-Software-Chef Luis Felipe Cabrera hat auf der O'Reilly Emerging Technology Conference (ETech) in San Diego ein neues Projekt des Online-Händlers vorgestellt, bei dem Menschen künftig Computern helfen sollen.

Kernpunkt der Idee ist eine neue Web-Services-Plattform namens "Mechanical Turk" ("mechanischer Türke"). Die nutzt ein auktionsartiges System, um Aufgaben an Menschen zu delegieren, die für Computer zu komplex sind, wie die Erkennung des Unterschiedes zwischen menschlichen Gesichtern und Gegenständen oder die akkurate Mitschrift einer Tonaufzeichnung. Cabrera nennt die Plattform keck "künstliche künstliche Intelligenz", da diesmal ein Computer den Menschen bittet, ihm bei der Arbeit zu helfen – und nicht umgekehrt.

Um die Idee zu illustrieren, schilderte Cabrera einen Test, den die Amazon-Suchmaschine A9.com kürzlich durchgeführt hat. Dabei wurden Durchschnittsnutzer gebeten, so genannte "Human Intelligence Tasks" (HITs), zu erfüllen. Also Aufgaben, die menschliche Intelligenz benötigen, weil Computer hier normalerweise versagen – etwa ein Gebäude oder Geschäft aus einem Bild eines ganzen Gebäudeblocks herauszusuchen, um den Bildausschnitt dann mit einer Adresse zu verknüpfen.

Quelle: <http://www.heise.de/tr/artikel/Amazons-mechanischer-Tuerke-278255.html>

- Ermöglicht das Drucken quasi aus der Cloud heraus
- Verbreitung internetfähiger Geräte wie Netbooks, Touchpads und Mobiltelefone steigt
- Einrichtung eines lokalen Druckers ist schwierig
  - Fehlen geeigneter Druckertreiber
  - Teilweise sehr geringen Ressourcen in den Geräten
  - Vielfalt der Betriebssysteme (Mac OS, Android, Windows Mobile...)
- Lösung: Google Cloud Print  
<http://code.google.com/apis/cloudprint>
- HP bietet kompatible Drucker an
  - Über eine E-Mail-Adresse lassen sich die Geräte identifizieren und als Cloud-Drucker unter Chrome OS hinzufügen
- Nutzer sendet sein zu druckendes Dokument an den Dienst, legt die Druckeinstellungen fest und bekommt eine Rückmeldung über die erfolgreiche Ausführung des Auftrags

- **Google Cloud Print kompatible Netzwerkdrucker**

- Drucker ist beim Dienst angemeldet
- Druckjobs werden an einen Dienst gesendet
- Dienst bereitet den Druckauftrag auf und leitet ihn an den Drucker weiter

- **Legacy Drucker** (sind nicht zu Google Cloud Print kompatibel)

- Lokal angeschlossene Drucker (USB) oder Netzwerkdrucker
- Auf einem lokalen PC wird ein Proxy installiert
- Proxy registriert den Drucker und sendet Druckaufträge an den Dienst
- Aufbereitete Druckjobs werden über den Proxy an den Drucker gesendet
- Nachteil: Proxy-Rechner muss zum Drucken eingeschaltet sein



- Cloud Gaming-Dienste machen High-End Videospiele auf Low-End Geräten (ältere PCs, Fernseher, Mobiltelefone) verfügbar
- Videospiele werden in den Serverfarmen des Anbieters ausgeführt
- Ausschließlich die Darstellung erfolgt auf den lokalen Zielgeräten
- Darstellung wird als komprimierter Videostrom übertragen
- Benutzereingaben werden zum Anbieter gesendet und dort ausgewertet
- Nachteil: Die notwendige Kompression reduziert die optischen Qualität
- Problem: Netzwerklatenzen müssen gering sein, da die Benutzereingaben zu einem entfernten Server übertragen und ausgewertet werden
- Zeit zwischen Benutzereingabe und Ergebnis auf dem lokalen Anzeigegerät muss gering sein, um den Spielfluss nicht zu unterbrechen
- Positiver Nebeneffekt für die Anbieter: Keine Raubkopien möglich
- Anbieter: OnLive und Gaikai

- Web-Desktops werden auch als Cloud-Betriebssysteme bezeichnet
- Bekanntes Produkt aus dieser Gruppe: eyeOS  
<http://www.eyeos.org>
- Das Betriebssystem und alle installierten Anwendungen und Benutzerdaten befinden sich auf den Serverfarmen des Anbieters
- Benutzer benötigen lediglich einen Browser und Internetzugang
- Der Begriff Cloud-Betriebssystem ist in den meisten Fällen irreführend
- Auch für die Nutzung eines Cloud-Betriebssystems ist ein Rechner mit Browser und daher ein zugrunde liegendes Betriebssystem nötig
  - Das native Betriebssystem wird nicht ersetzt
    - Es werden nur die Benutzeranwendungen und Benutzerdaten ausgelagert

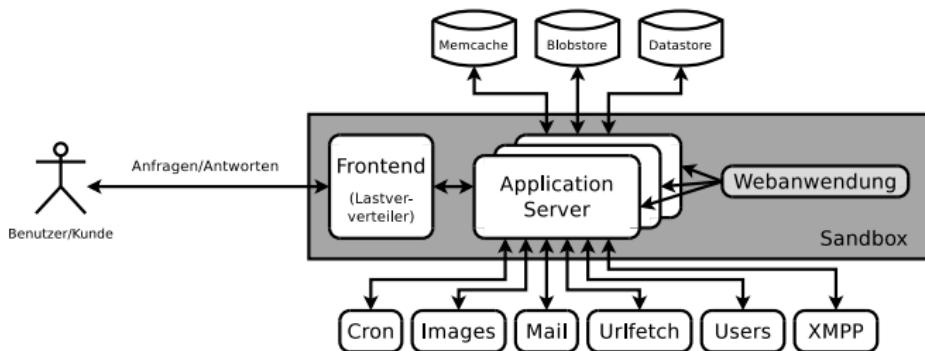
# Cloud Cooking – die Zukunft ?!



# Beispiel für einen Plattformdienst – App Engine



- Kunden können eigene Webanwendungen in Python 2.5.2/2.7, Java 6 und Go betreiben
- Skaliert automatisch nach Bedarf
- Anwendungen können verschiedene Infrastruktur- und Speicherdiensste nutzen



- Weitere öffentliche Plattformdienste: AWS Elastic Beanstalk, Engine Yard, Windows Azure Platform, Force.com

- **Authentifizierung**

- Authentifizierung/Authorisierung ist über Google-Accounts möglich
- Keine Implementierung einer eigenen Authentifizierungslösung notwendig

- **Datastore**

- Persistenter Speicher, realisiert als Key/Value-Datenbank
- Transaktionen sind atomar
- Definition, Abfrage und Manipulation von Daten erfolgt über eine eigene Sprache, die GQL (Google Query Language)
  - GQL hat große Ähnlichkeiten mit SQL (Structured Query Language)

- **Memcache**

- Hochperformanter temporärer Datenspeicher aus Hauptspeicher
- Sehr gute Zugriffszeiten
- Jeder Eintrag wird mit einem eindeutigen Schlüssel abgelegt
- Jeder Eintrag ist auf 1 MB beschränkt
- Es wird eine Verfallszeit in Sekunden angeben, wann der Eintrag aus dem Memcache entfernt werden soll
- Daten werden je nach Auslastung des Memcache früher wieder verdrängt

- **URL Fetch**

- Zugriff auf Inhalte im Internet
- Kommunikation über RESTful Web Services möglich
  - Unterstützte Methoden: GET, POST, PUT, DELETE und HEADER
  - Anfragen und Antworten sind auf jeweils 1 MB beschränkt

- **Mail**

- Emails versenden und empfangen über das GMail-Gateway
- Absender der E-Mail kann nicht frei gewählt werden
  - Als Absender erlaubt ist nur die E-Mail Adresse des eingeloggten Benutzers oder des Administrators der Anwendung

- **Bildmanipulation**

- Funktionen um Bilder zu drehen, spiegeln, beschneiden und zum Verändern der Bildgröße

- **Nachrichtenübermittlung via XMPP**

- Nachrichten können an jedes XMPP-kompatibles Nachrichtensystem (z.B. Google Talk) gesendet und von diesem empfangen werden

# Einsatz der GAE

- Jeder Nutzer kann bis zu 10 Anwendungen haben

The screenshot shows the Google App Engine dashboard under the account 'wolkenrechnen@googlemail.com'. The title bar indicates the URL is https://appengine.google.com/. The dashboard header includes the Google logo and the text 'Google app engine'. On the right, there are links for 'My Account', 'Help', and 'Sign out'. The main section is titled 'My Applications' and displays a table of six applications:

Application	Title	Account	Storage Scheme	Current Version
<a href="#">cloud-vorlesung</a>	Cloud Vorlesung		1	
<a href="#">cloudoctopus</a>	Octopus Cloud Storage System		2	
<a href="#">cray23kl</a>	cray23kl		2	
<a href="#">koalacloud</a>	Koala Cloud Manager		1	
<a href="#">neverland23kl</a>	Steht noch nicht fest		1	
<a href="#">wildestesten</a>	Hier werden Sachen ausprobiert		1	

Below the table, there is a 'Create Application' button and a message stating 'You have 10 applications remaining.' Navigation links at the bottom include '« Prev 20', '1-6 of 6', and 'Next 20 »'.

- Problem: Langsam werden die guten Namen knapp!
- Anwendungen sind erreichbar unter:  
`http://<name-der-anwendung>.appspot.com`
- Im Dashboard gibt es alle wichtigen Informationen zur Anwendung

# Dashboard

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

Google app engine https://appengine.google.com/dashboard?&app\_id=koalacloud

wolkenrechner@googlemail.com | My Account | Help | Sign out

koalacloud 1 My Applications

Main

- Dashboard**
- Instances
- Logs
- Versions
- Backends
- Cron Jobs
- Task Queues
- Quota Details

Data

- Datstore Indexes
- Datstore Viewer
- Datstore Statistics
- Blob Viewer
- Prospective Search
- Datstore Admin

Administration

- Application Settings**
- Permissions
- Blacklist
- Admin Logs

Billing

- Billing Settings
- Billing History

Resources

- Documentation
- FAQ
- Developer Forum
- Downloads
- System Status

Charts 

Number of Instances - Details

Average QPS	Average Latency	Average Memory
Unknown	Unknown ms	Unknown Mbytes

Billing Status: Free - [Settings](#)

Quotas reset every 24 hours. Next reset: 11 hrs 

Resource

Frontend Instance Hours	2%	0.5% of 28.00 Instance Hours
Backend Instance Hours	0%	0.00 of 9.00 Instance Hours
Datstore Stored Data	0%	0.00 of 1.00 GBbytes
Task Queue Stored Task Bytes	0%	0.00 of 0.49 GBbytes
Blobstore Stored Data	0%	0.00 of 5.00 GBbytes
Datstore Write Operations	0%	0.00 of 0.05 Million Ops
Datstore Read Operations	0%	0.00 of 0.05 Million Ops
Datstore Small Operations	0%	0.00 of 0.05 Million Ops
Outgoing Bandwidth	0%	0.00 of 1.00 GBbytes
Recipients Emailed	0%	0 of 100
Stances Sent	0%	0 of 10,000
Channels Created	0%	0 of 100

Current Load

URI	Requests last 15 hrs	Runtime MCycles last hr	Avg Latency last hr
/	18		
/instanzen	4		
/regionen	2		
/images	2		
/zonen	1		
/snapshots	1		
/walcon.ico	1		

Errors 

URI	Count	% Errors last 15 hrs

# Mehrere Versionen online haben

- Von jeder Anwendung können mehrere Versionen hochgeladen sein
- Jede Version ist direkt erreichbar  
`http://<version>.latest.<name-der-anwendung>.appspot.com`
- Gut zum Testen

Google app engine

wolkenrechnen@googlemail.com | My Account | Help | Sign out

koalacloud 1 < My Applications

Version	Default	Deployed	Delete
1  instances   python   api_version: 1	Yes	84 days, 2:58:05 ago by wolkenrechnen@googlemail.com	<input type="button" value="Delete"/>
2  instances   python   api_version: 1	No	640 days, 23:09:27 ago by	<input type="button" value="Delete"/>
3  instances   python   api_version: 1	No	599 days, 1:04:28 ago by wolkenrechnen@googlemail.com	<input type="button" value="Delete"/>
4  instances   python   api_version: 1	No	333 days, 9:00:33 ago by wolkenrechnen@googlemail.com	<input type="button" value="Delete"/>

Main

[Dashboard](#)

[Instances](#)

[Logs](#)

[Versions](#)

[Backends](#)

[Cron Jobs](#)

[Task Queues](#)

- Browser... (egal welcher)
- Google App Engine SDK für Python und/oder JAVA  
<http://code.google.com/appengine/downloads.html>
  - Verfügbar für Linux/UNIX, MaxOS-X und Windows
- Google Plugin for Eclipse
- Erweiterung PyDev für Eclipse (für Python-Anwendungen)
  - <http://code.google.com/eclipse/>

- GAE unterstützt nur Python 2.5.2

```
$ python --version  
Python 2.5.2
```

- SDK installieren

```
$ wget http://googleappengine.googlecode.com/files/  
    google_appengine_1.5.5.zip  
$ unzip google_appengine_1.5.5.zip  
$ echo -e "export PATH=\$PATH:~/google_appengine \n" > ~/.bashrc  
$ bash  
$ appcfg.py -h  
Usage: appcfg.py [options] <action>  
...
```

# Einschränkungen der GAE

- Nur wenige Programmiersprachen werden unterstützt
- Kommunikation mit anderen Web-Anwendungen oder Servern nur über URL Fetch, XMPP oder Email
  - Nur über die Ports 80, 443, 4443, 8080-8089, 8188, 8444 und 8990
  - Anfragen und Antworten dürfen maximal 1 MB groß sein
  - Es kann z.B. nicht über FTP oder SSH kommuniziert werden
- Emails mit Anhang dürfen maximal 1 MB groß sein
- Ein- und ausgehende XMPP-Nachrichten dürfen maximal 100 kB groß sein
- Bilder, die mit der GAE-API bearbeitet werden, dürfen maximal 1 MB groß sein
- Elemente im Memcache dürfen maximal 1 MB groß sein
- Java-Anwendungen dürfen nicht multithreaded sein
- Nicht alle Klassen der JRE Standard Edition sind verfügbar
- Nur lesender Zugriff auf das Dateisystem

# Kosten der GAE

## Billable Resource Unit Costs

**Important:** Currently, you do not see the instance classes you use in your billing line items. Instead, you will see the appropriate multiple of instance hours. For example, if you use an F4 instance for one hour, you do not see "F4" listed, but you will see billing for four instance hours at the F1 rate.

The cost for computing resources is as follows:

Resource	Unit	Unit cost
Outgoing Bandwidth	gigabytes	\$0.12
Frontend Instances (F1)	Instance hours	\$0.08
Frontend Instances (F2)	Instance hours	\$0.16
Frontend Instances (F4)	Instance hours	\$0.32
Discounted Instances	Instance hours	\$0.05
Backend Instances (B1 class)	Hourly per instance	\$0.08
Backend Instances (B2 class)	Hourly per instance	\$0.16
Backend Instances (B4 class)	Hourly per instance	\$0.32
Backend Instances (B8 class)	Hourly per instance	\$0.64
Stored Data (Blobstore)	gigabytes per month	\$0.13
Stored Data (Datastore)	gigabytes per month	\$0.24
Stored Data (Task Queue)	gigabytes per month	\$0.24
Channel	Channel opened	\$0.00001 (\$0.01/100 channels)
Recipients Emailed	email	\$0.0001
XMPP	XMPP stanzas	\$0.000001 (\$0.10/100,000 stanzas)

There are three types of low-level Datastore operations:

Operation	Cost
Write	\$0.10 per 100k operations
Read	\$0.07 per 100k operations
Small	\$0.01 per 100k operations

- Ziel: Gästebuch entwickeln
- Eckdaten:
  - Authentifizierungs-Informationen von Google werden sinnvoll genutzt
  - Daten werden im Datastore abgelegt
  - Einträge können gelöscht werden
  - Web-Seiten werden mit dem Web-Framework Django erstellt
- Das Beispiel orientiert sich stark an diesem Tutorial:  
**Developing and deploying an application on Google App Engine**  
<http://www.youtube.com/watch?gl=DE&hl=de&v=bfg0-LXGpTM>

# Konfigurationsdatei – app.yaml

- Für jede Anwendung muss eine Konfigurationsdatei app.yaml existieren

- application: Name der Anwendung
- version: Version der Anwendung
- runtime: Laufzeitumgebung
  - Hier kann nur python ausgewählt werden
  - Für JAVA-Anwendungen ist eine andere Konfigurationsdatei notwendig
- api\_version: API-Version der GAE
  - Aktuell ist Version 1
- handlers: Abbildung der URLs auf Skripte
  - In diesem Beispiel werden alle Anfragen an das Python-Skript main.py weitergeleitet

```
# cat app.yaml
application: cloud-vorlesung
version: 1
api_version: 1
runtime: python

handlers:
- url:.*
  script: main.py
```

- Für jede Anwendung sollte ein Verzeichnis gleichen Namens lokal existieren

```
mkdir cloud-vorlesung  
cd cloud-vorlesung
```

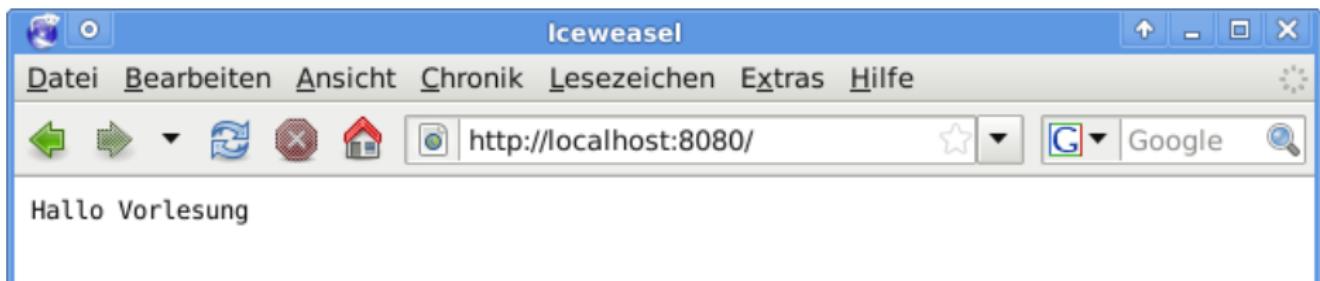
- Mini-Anwendung in `main.py` erstellen

```
$ cat main.py  
#!/usr/bin/env python  
  
print 'Hallo Vorlesung'
```

- Mini-Anwendung mit dem Entwicklungs-Server starten

```
$ dev_appserver.py cloud-vorlesung/  
INFO    2010-03-12 07:07:52,051 appengine_rpc.py:159] Server: appengine.google.com  
INFO    2010-03-12 07:07:52,068 appcfg.py:349] Checking for updates to the SDK.  
INFO    2010-03-12 07:07:52,435 appcfg.py:363] The SDK is up to date.  
INFO    2010-03-12 07:07:52,607 dev_appserver_main.py:399]  
Running application cloud-vorlesung on port 8080: http://localhost:8080
```

# Mini-Anwendung testen



- Der Entwicklungs-Server gibt Informationen aus, was passiert ist
- Es gab einen erfolgreichen Zugriff  $\Rightarrow$  HTML-Statuscode 200

```
$ dev_appserver.py cloud-vorlesung/
INFO    2010-03-12 07:07:52,051 appengine_rpc.py:159] Server: appengine.google.com
INFO    2010-03-12 07:07:52,068 appcfg.py:349] Checking for updates to the SDK.
INFO    2010-03-12 07:07:52,435 appcfg.py:363] The SDK is up to date.
INFO    2010-03-12 07:07:52,607 dev_appserver_main.py:399]
          Running application cloud-vorlesung on port 8080: http://localhost:8080
INFO    2010-03-12 07:07:56,724 dev_appserver.py:3246] "GET / HTTP/1.1" 200 -
INFO    2010-03-12 07:07:56,763 dev_appserver_index.py:205]
          Updating /home/baun/cloud-vorlesung/index.yaml
```

# Gästebuch – main.py

```
#!/usr/bin/env python
# -*- coding: iso-8859-15 -*-
import os, sys
import wsgiref.handlers
from google.appengine.ext import db
from google.appengine.ext import webapp
from google.appengine.ext.webapp import template

# Name des Datastore ist gaestebuch
# Hier wird die Struktur des Datastore festgelegt
class gaestebuch(db.Model):
    message = db.StringProperty(required=True)
    when = db.DateTimeProperty(auto_now_add=True)
    who = db.StringProperty()

class MyHandler(webapp.RequestHandler):
    def get(self):
        # Anfrage an den Datastore
        shouts = db.GqlQuery('SELECT * FROM gaestebuch ORDER BY when DESC')
        values = {'shouts': shouts}
        self.response.out.write(template.render('main.html', values))
    def post(self):
        shout = gaestebuch(message=self.request.get('message'), who=self.request.get('who'))
        # In den Datastore schreiben
        shout.put()
        self.redirect('/')
```

# Gästebuch – main.py

```
class Loeschen(webapp.RequestHandler):
    def post(self):
        alles_loeschen_query = gaestebuch.all(keys_only=True)
        alles_loeschen = alles_loeschen_query.fetch(300)
        db.delete(alles_loeschen)
        #self.response.out.write('posted!')
        self.redirect('/')

app = webapp.WSGIApplication([('/', MyHandler), ('/loeschen', Loeschen)], debug=True)

def main():
    wsgiref.handlers.CGIHandler().run(app)

# Die folgende Zeile bewirkt, dass das Programm als Modul
# und als eigenständiges Programm verwendet werden kann
if __name__ == "__main__":
    main()
```

# Gästebuch – main.html

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd">
<html>
<head><title>Einfaches Gästebuch mit Google App Engine</title></head>
<body>
<h3>Einfaches Gästebuch mit Google App Engine</h3>

{% for shout in shouts %}
<div>
  {% if shout.who %}
    <b>{{shout.who}}</b>
  {% else %}
    <b>Anonymous</b>
  {% endif %}
  sagt:
  <b>{{shout.message}}</b>
</div>
{% endfor %}

<p>

<form action="" method="post" accept-charset="utf-8">
Name: <input type="text" size="20" name="who" value="" if="who">
Nachricht: <input type="text" size="20" name="message" value="" if="message">
<input type="submit" value="Absenden"></form>

<form action="loeschen" method="post" accept-charset="utf-8">
<table border="0" cellspacing="5" cellpadding="5">
  <tr><td align="center"><input type="submit" value="DB Löschen"></td></tr>
</table>
</form>

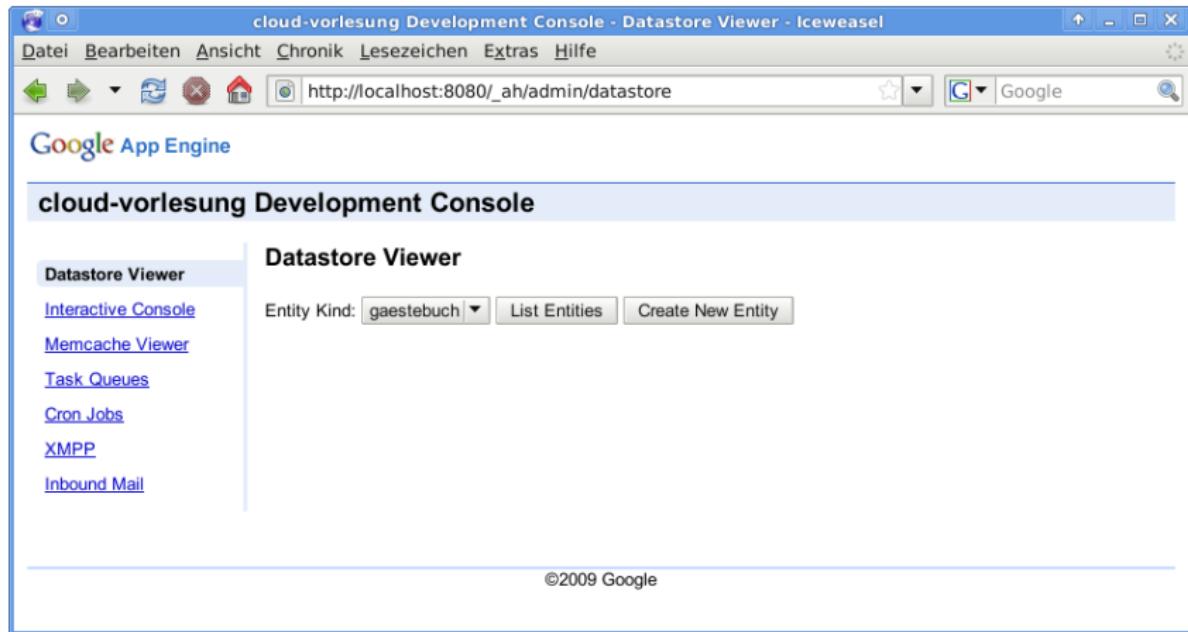
</body>
</html>
```

# Ergebnis



# Hilfreiches Werkzeug: Development Console (1)

- Im Entwicklungs-Server eingebaut
- Erreichbar unter `http://localhost:8080/_ah/admin/`



# Hilfreiches Werkzeug: Development Console (2)

The screenshot shows the Google App Engine Development Console's Datastore Viewer. The URL in the browser is `http://localhost:8080/_ah/admin/datastore?kind=gaestebuch`. The Datastore Viewer interface includes a sidebar with links like Datastore Viewer, Interactive Console, Memcache Viewer, Task Queues, Cron Jobs, XMPP, and Inbound Mail. The main area displays a table of guestbook entries with columns: Key, ID, Key Name, message, when, and who. There are three entries listed, all created by 'Christian' on March 12, 2010, at 07:57:31. The first entry says 'Hallo!! Vorlesung...', the second 'Ist doch ganz einfach', and the third 'Ich raff nix.' A 'Delete' button is at the bottom left, and a page number '1' is at the bottom right.

<input type="checkbox"/>	Key	ID	Key Name	message	when	who
<input type="checkbox"/>	<a href="#">ag9jbG91...</a>	146		Hallo!! Vorlesung...	2010-03-12 07:57:31	Christian
<input type="checkbox"/>	<a href="#">ag9jbG91...</a>	147		Ist doch ganz einfach	2010-03-12 07:57:38	Christian
<input type="checkbox"/>	<a href="#">ag9jbG91...</a>	148		Ich raff nix.	2010-03-12 07:57:56	

©2009 Google

- Datastore liegt lokal in `/tmp/dev_appserver.datastore`

# Anwendung mit appcfg.py aus dem SDK hochladen

```
$ echo <passwort> | appcfg.py --passin --email=cray@unix-ag.uni-kl.de update cloud-vorlesung/
Application: cloud-vorlesung; version: 1.
Server: appengine.google.com.
Scanning files on local disk.
Initiating update.
Cloning 3 application files.
Uploading 2 files and blobs.
Uploaded 2 files and blobs
Deploying new version.
Checking if new version is ready to serve.
Will check again in 1 seconds.
Checking if new version is ready to serve.
Will check again in 2 seconds.
Checking if new version is ready to serve.
Closing update: new version is ready to start serving.
Uploading index definitions.
```



# Logdaten mit appcfg.py holen

```
$ echo <passwort> | appcfg.py --passin --email=cray@unix-ag.uni-kl.de request_logs cloud-vorlesung/ logs.txt
Application: cloud-vorlesung; version: 1.
Server: appengine.google.com.
Downloading request logs for cloud-vorlesung 1.1.
Copying request logs to 'logs.txt'.
Copied 3 records.

$ cat logs.txt
87.178.89.156 - - [29/Apr/2010:06:01:20 -0700] "GET / HTTP/1.1" 200 540 - "Mozilla/5.0
(X11; U; Linux i686; de; rv:1.9.0.6) Gecko/2009020409 Iceweasel/3.0.14 (Debian-3.0.14-1),gzip(gfe)"
87.178.89.156 - - [29/Apr/2010:06:01:20 -0700] "GET /favicon.ico HTTP/1.1" 404 124 - "Mozilla/5.0
(X11; U; Linux i686; de; rv:1.9.0.6) Gecko/2009020409 Iceweasel/3.0.14 (Debian-3.0.14-1),gzip(gfe)"
```

- IP-Adresse des Clients
- Datum und Zeitpunkt der Anfrage
- HTML-Anfrage
- HTML-Statuscode (Antwort des Servers)
- Größe der Antwort in Bytes
- Client-Informationen
- ...

- Seit Januar 2011 bietet Amazon mit Elastic Beanstalk einen Plattformdienst, vergleichbar zur Google App Engine an
  - <http://aws.amazon.com/elasticbeanstalk>
- Java-Webanwendungen können in der AWS-Cloud betrieben werden
- Anwendungen werden in WAR-Dateien (Java Web Application Archive) verpackt und zu Elastic Beanstalk hochgeladen
- Im Hintergrund werden eine oder mehr EC2-Instanzen mit dem Apache Tomcat Applikation Server gestartet
- Innerhalb weniger Minuten ist die Webanwendung erreichbar  
<http://<name-der-anwendung>.elasticbeanstalk.com>

## Amazons Elastic Beanstalk, ein erster Praxistest

<http://www.heise.de/developer/artikel/Amazons-Elastic-Beanstalk-ein-erster-Praxistest-1185443.html>

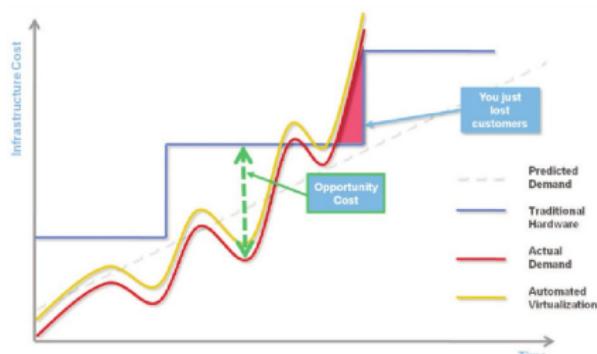
# Beispiel für Infrastrukturdienste – Amazon Web Services

- Sammlung öffentlicher Cloud-Dienste
- Bekannte Dienste der AWS sind u.a.



Elastic Compute Cloud (EC2)	⇒	Infrastrukturdienst für virtuelle Server
Simple Storage Service (S3)	⇒	Speicherdiens für Webobjekte
Elastic Block Store (EBS)	⇒	Speicherdiens für virtuelle Speichervolumen
Elastic Load Balancing (ELB)	⇒	Dienst für virtuelle Lastverteiler
CloudWatch	⇒	Dienst zum Überwachen von AWS-Ressourcen
Auto Scaling	⇒	Dienst zum Skalieren von EC2-Kapazitäten

Take the Risk Factor out of Capacity Planning

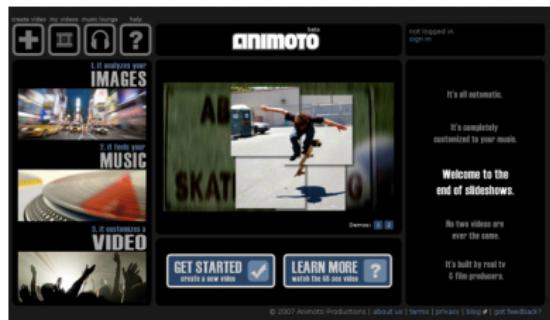


Source: Amazon Web Services

- Anwendungsbeispiele:
  - Rechenzentren virtuell realisieren
  - Elastische Infrastruktur aufbauen (z.B. für ein Startup)
  - Kurzfristig Ressourcen akquirieren
- Weitere öffentliche Infrastrukturdienste: Rackspace, GoGrid, IBM SmartCloud Enterprise

# Beispiel für AWS: Animoto – <http://animoto.com>

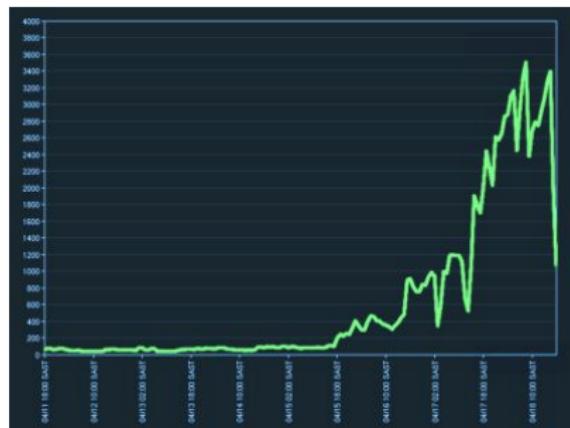
- Benutzer können Videos aus eigenen Bildern und Musik erstellen
- Die Bilder werden automatisch und passend mit der Musik unterlegt
- Die Software analysiert Bilder und Musik im generiert Videos im Stil eines Trailers bzw. MTV-Musikvideos
- Videos können direkt in YouTube eingestellt oder in diversen Formaten gespeichert werden



Quelle: <http://profy.com/2007/08/15/animoto-easy-video>

# Beispiel für AWS: Animoto (2)

- Benutzerzahlen waren von 2006 bis 2008 konstant (niedrig)
- Im April 2008 auf Facebook erwähnt
  - 750.000 neue Benutzer in 3 Tagen
  - Bis zu 25.000 Benutzer wollten in einer Stunde ein Video erstellen
  - Slashdot-Effekt!
  - Automatische Anpassung der Instanzen zum Rendern der Videos von 2 auf bis zu 450



# Slashdot-Effekt

The screenshot shows a Microsoft Internet Explorer window with the title bar "Slashdot-Effekt - Wikipedia - Iceweasel". The menu bar includes "Datei", "Bearbeiten", "Ansicht", "Chronik", "Lesezeichen", "Extras", and "Hilfe". The toolbar has icons for back, forward, search, and other functions. The address bar shows the URL "http://de.wikipedia.org/wiki/Slashdot-Effekt". The page content starts with the Wikipedia logo and the heading "Slashdot-Effekt". The text describes the Slashdot Effect as a sudden increase in traffic to a website after it is mentioned in a popular IT magazine like Slashdot or heise. It mentions that this can overwhelm servers and lead to "geslashdotted" pages. The text also notes that large websites with server farms usually don't have issues with this effect. Below the main text, there's a sidebar with a navigation menu and a "Contents" section.

Der sogenannte **Slashdot-Effekt** oder das **Slashdotting** tritt auf, wenn eine bisher wenig populäre **Website** von einem **IT-Online-Magazin** wie **Slashdot** oder **heise** aufgegriffen wird und so binnen Minuten ein erheblicher Benutzeransturm auf die Website beginnt. Dieser führt oft dazu, dass erheblicher **Traffic** verursacht wird und der **Server** vorübergehend einzelne Anfragen nicht mehr oder nur noch sehr langsam beantworten kann. Die Seite ist dann „geslashdotted“ (engl. *slashdotted*).

Große Websites, die von einer Server-Farm bedient werden, haben meistens keine Probleme mit dem erhöhten Traffic. Es sind vor allem kleinere Einzel-Server, die einem Slashdot-Effekt zum Opfer fallen. Manchmal wird der Slashdot-Effekt scherhaft mit einem **Distributed-Denial-of-Service**-Angriff verglichen.

Um den Ansturm auf die betroffenen Seiten zu reduzieren, werden von unabhängigen Seiten immer wieder **Mirrors** angeboten in der Hoffnung, dass die Leser auf die Mirrors anstelle der Originalseite zugreifen. Koordiniert werden solche Projekte von **Coral** und **MirrorDot**.

**Inhaltsverzeichnis [Verbergen]**

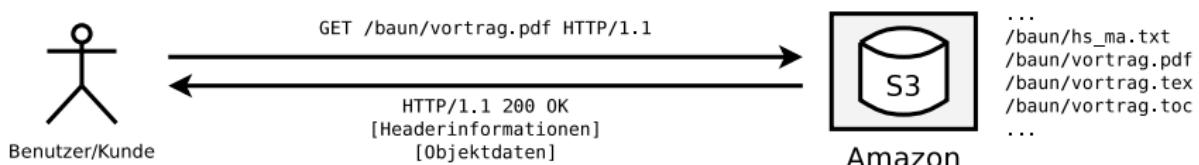
- 1 Andere Bezeichnungen
- 2 Beispiel
- 3 Weblinks
- 4 Einelnachweise

- Lineare Steigerung der Benutzer-/Zugriffszahlen ist unrealistisch
- Großes Problem für Start-ups mit eigenen Ressourcen!

- 2007: Die New York Times will PDF-Versionen der Artikel aus den Jahren 1851–1980 erstellen
  - Die Artikel 1851-1922 sollten komplett frei gegeben werden
- Die Artikel lagen in Form von 11 Millionen eingescannten Bildern vor
- Jeder Artikel musste aus verschiedenen TIFF-Dateien zusammengesetzt und skaliert werden
- 4 TB Daten wurden zuerst in S3 hochgeladen
- 100 EC2-Instanzen benötigten ca. 24 Stunden für die Berechnung
- Ergebnis: 1,5 TB PDF-Dateien in S3
- [http://open.blogs.nytimes.com/2007/11/01/  
self-service-prorated-super-computing-fun/](http://open.blogs.nytimes.com/2007/11/01/self-service-prorated-super-computing-fun/)
- [http://open.blogs.nytimes.com/2008/05/21/  
the-new-york-times-archives-amazon-web-services-timesmachine/](http://open.blogs.nytimes.com/2008/05/21/the-new-york-times-archives-amazon-web-services-timesmachine/)

# Amazon S3

- Simple Storage Service (S3)
- Daten werden als **Objekte** (1 Byte bis 5 TB) gespeichert
- Jedes Objekt ist einem **Bucket** zugeordnet
  - Jeder Bucket hat einen eindeutigen Namen und kann keinen weiteren Bucket enthalten
- Verzeichnisse sind nicht möglich
- Objekte sind online erreichbar
  - <http://s3.amazonaws.com/bucket/objekt>
- Zugriff auf Buckets und Objekte geschieht via REST oder SOAP



- Im Juli 2011 waren in S3 schon 449 Milliarden Objekte gespeichert

# Kosten von S3 (Speicherkapazität und Zugriffe)

## Speichergebühren

Region:	USA Ost (Virginia)	
	Standard-Speichermethode	Reduced Redundancy Storage
Erste 1 TB pro Monat	\$0,125 pro GB	\$0,093 pro GB
Nächste 49 TB pro Monat	\$0,110 pro GB	\$0,083 pro GB
Nächste 450 TB pro Monat	\$0,095 pro GB	\$0,073 pro GB
Nächste 500 TB pro Monat	\$0,090 pro GB	\$0,063 pro GB
Nächste 4.000 TB pro Monat	\$0,080 pro GB	\$0,053 pro GB
Über 5.000 TB pro Monat	\$0,055 pro GB	\$0,037 pro GB

## Anfragegebühren

Region:	USA Ost (Virginia)	
	Preise	
PUT, COPY, POST oder LIST-Anfragen	\$0,01 pro 1.000 Abfragen	
GET und alle anderen Anfragen †	\$0,01 pro 10.000 Anfragen	
† Keine Gebühr für Lösch-Anfragen		

- Stand: März 2012

# Kosten von S3 (Datentransfer)

## Anfragegebühren

Region:	USA Ost (Virginia)	Preise
PUT, COPY, POST oder LIST-Anfragen	\$0,01 pro 1.000 Abfragen	
GET und alle anderen Anfragen †	\$0,01 pro 10.000 Anfragen	
† Keine Gebühr für Lösch-Anfragen		

## Datenübertragungsgebühren

Region:	USA Ost (Virginia)	Preise
<b>Übertragung eingehender Daten</b>		
Übertragung aller eingehenden Daten	\$0,000 pro GB	
<b>Übertragung ausgehender Daten</b>		
Erstes 1 GB pro Monat	\$0,000 pro GB	
Bis zu 10 TB pro Monat	\$0,120 pro GB	
Nächste 40 TB pro Monat	\$0,090 pro GB	
Nächste 100 TB pro Monat	\$0,070 pro GB	
Nächste 350 TB pro Monat	\$0,050 pro GB	
Nächste 524 TB pro Monat	Kontakt	
Nächste 4 PB pro Monat	Kontakt	
Mehr als 5 PB pro Monat	Kontakt	

# AWS Import/Export

- Unterstützt den Transfer großer Datenmengen in oder aus der Cloud heraus

Available Internet Connection	Theoretical Min. Number of Days to Transfer 1TB at 80% Network Utilization	When to Consider AWS Import/Export?
T1 (1.544Mbps)	82 days	100GB or more
10Mbps	13 days	600GB or more
T3 (44.736Mbps)	3 days	2TB or more
100Mbps	1 to 2 days	5TB or more
1000Mbps	Less than 1 day	60TB or more

- Ein Datenspeicher (Festplatte) wird an Amazon gesendet und dann vor Ort in einen S3-Bucket kopiert
  - Dateisysteme: NTFS, ext2, ext3 und FAT32 mit maximal 2 TB
- Gebühr pro Datenspeicher: 80 Dollar
- Übertragung vor Ort kostet pro Stunde 2,49 Dollar
- <http://aws.amazon.com/importexport/>

## S3 mit s3cmd bedienen

- Für den Zugriff auf Amazon S3 von der Kommandozeile aus, gibt das einfach zu bedienende Werkzeug s3cmd
- <http://s3tools.org/s3cmd>

Konfiguration der Zugangsdaten	<code>s3cmd --configure</code>
Liste der eigenen Buckets	<code>s3cmd ls</code>
Bucket anlegen	<code>s3cmd mb s3://Bucket</code>
Objekt hochladen <code>s3cmd put LokaleDatei s3://Bucket/EntferntesObjekt</code>	
Inhalt eines Buckets auslesen	<code>s3cmd ls s3://Bucket</code>
Objekt herunterladen <code>s3cmd get s3://Bucket/EntferntesObjekt LokaleDatei</code>	
Objekt löschen <code>s3cmd del s3://Bucket/EntferntesObjekt</code>	
Löschen eines (leeren) Buckets	<code>s3cmd rb s3://Bucket</code>

# Firefox-Plugin S3Fox: <http://www.s3fox.net>

S3 Firefox Organizer - Iceweasel

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

chrome://s3fox/content/s3foxWindow.xul

Manage Accounts bauni Synchronize Folders AWS Import/Export Preferences

/home/bauni/Diverse\_Bilder/Sta Browse

File Name	File Size(KB)	Modified Time
54.jpg	211	04/18/2005 12:45 AM
55.jpg	231	04/18/2005 12:45 AM
56.jpg	216	04/18/2005 12:45 AM
57.jpg	255	04/18/2005 12:45 AM
58.jpg	270	04/18/2005 12:45 AM
59.jpg	261	04/18/2005 12:45 AM
60.jpg	250	04/18/2005 12:45 AM
61.jpg	244	04/18/2005 12:45 AM
62.jpg	255	04/18/2005 12:45 AM
63.jpg	272	04/18/2005 12:45 AM
64.jpg	251	04/18/2005 12:45 AM
65.jpg	244	04/18/2005 12:45 AM
66.jpg	247	04/18/2005 12:45 AM
67.jpg	255	04/18/2005 12:45 AM

/christianbaun/

File Name	File Size(KB)	File...
11.jpg	261	260723
10.jpg	278	277003
09.jpg	286	285504
08.jpg	277	276889
07.jpg	269	268698
06.jpg	281	280591
05.jpg	285	284468
04b.jpg	348	347084

Current Tasks:

Regular Transfer Synchronized Folders Transfer Log

✖ Clear ⏸ Pause ✓ Clear Completed ✖ Retry Failed Tasks

File Name	From	To	Type	Progress	Status
56.jpg	/home/bauni...	/christia...	Upload (1...	100%	Completed
57.jpg	/home/bauni...	/christia...	Upload (1...	100%	Completed
58.jpg	/home/bauni...	/christia...	Upload (3...	0%	Processing
59.jpg	/home/bauni...	/christia...	Upload (0...	0%	In Queue
60.jpg	/home/bauni...	/christia...	Upload (0...	0%	In Queue
61.jpg	/home/bauni...	/christia...	Upload (0...	0%	In Queue
62.jpg	/home/bauni...	/christia...	Upload (0...	0%	In Queue
63.jpg	/home/bauni...	/christia...	Upload (0...	0%	In Queue

# Mit der S3 API und boto arbeiten

```
#!/usr/bin/env python

from boto.s3.connection import S3Connection
# Verbindung zu S3 aufbauen
# Variable "conn" zeigt auf ein "S3Connection" Objekt
conn = S3Connection('<aws access key>', '<aws secret key>')

# Bucket "testbucket" erzeugen
conn.create_bucket('testbucket')

# Liste der Buckets erhalten und ausgeben
anfrage_buckets = conn.get_all_buckets()
print(anfrage_buckets)
```

- Image Hosting Service **SmugMug**
  - Nutzt S3 seit April 2006 um Bilder zu speichern
  - Im April 2008 hat SmugMug angegeben, durch S3 1 Million Dollar gespart zu haben
  - Berechnung:  
<http://don.blogs.smugmug.com/2006/11/10/amazon-s3-show-me-the-money/>
  - <http://www.smugmug.com>
- Online Backup **Jungle Disk**
  - <http://jungledisk.com>
- Online Backup **ElephantDrive**
  - <http://elephantdrive.com>
- Online Backup **ExEasy NetCDP**
  - <http://www.netcdp.com>
- Online Backup **Dropbox**
  - <http://www.dropbox.com>

# Erste NAS-Geräte nutzen schon S3 – Beispiel von HP

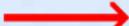


- Beispiel: HP MediaSmart Server EX485
- Benutzer kann festlegen welche Daten zum Backup in S3 gespeichert werden sollen

# Erste NAS-Geräte nutzen schon S3 – Beispiel von Qnap

## Netzwerkspeicher – technische Daten

Modell	TS-239 Pro II	TS-459 Pro
Hersteller/Anbieter	Qnap	Qnap
Web-Adresse	www.qnap.com	www.qnap.com
<b>Hardware und Lieferumfang</b>		
Firmware	3.2.2 (0128T)	3.2.2 (0128T)
Prozessor/RAM	Intel Atom D410 (1,66 GHz)/ 1 GByte DDR2	Intel Atom D510 (1,66 GHz)/ 1 GByte DDR2
LAN-Interface/Link Aggregation/Auto-failover/ Jumbo Frames	2× Gigabit-Ethernet/✓ /✓ /✓	2× Gigabit-Ethernet/✓ /✓ /✓
<b>Sharing-Funktionen</b>		
FTP/FTP verschlüsselt/abschaltbar	✓ /✓ /✓	✓ /✓ /✓
HTTP/HTTPS/abschaltbar	✓ /✓ /✓	✓ /✓ /✓
NFS/abschaltbar	✓ /✓	✓ /✓
AppleShare/abschaltbar	✓ /✓	✓ /✓
UPnP/abschaltbar	✓ /✓	✓ /✓
Medienserver per weitere Protokolle	UPnP-AV (TwonkyMedia), iTunes BitTorrent, Bonjour, IPv6, iSCSI, rsync, SNMP, SSH, telnet, WebDAV	UPnP-AV (TwonkyMedia), iTunes BitTorrent, Bonjour, IPv6, iSCSI, rsync, SNMP, SSH, telnet, WebDAV
Printserver/Protokolle	✓ /Windows-Share	✓ /Windows-Share
Besonderheiten	Unterstützung für DFS u. Amazon S3, IP-Kameras, MySQL, PHP	Unterstützung für DFS u. Amazon S3, IP-Kameras, MySQL, PHP



- „... Daten sichern die Qnap-NAS nicht nur über gängige Mechanismen wie rsync auf andere Server im Netz weg, sondern schicken sie auf Wunsch jetzt auch zeitgesteuert an den Cloud-Speicherdiest Amazon S3...“

Quelle: c't. Schnelle Gigabit-NAS für zu Hause und das Büro. 5/2010. S.114

# Verfügbarkeit von S3

- Amazon garantiert für S3 eine monatliche Verfügbarkeit von 99.9%

Verfügbarkeit	Downtime (HH:MM:SS)		
	pro Tag	pro Monat	pro Jahr
99.9%	00:01:26	00:43:49	08:45:56

- Wird die Verfügbarkeit unterschritten, erhält man eine Gutschrift
- <http://aws.amazon.com/s3-sla/>

Monthly Uptime Percentage	Service Credit Percentage
Equal to or greater than 99% but less than 99.9%	10%
less than 99%	25%

- Was bringt eine Gutschrift, wenn die eigenen, wichtigen Dienste und Daten nicht mehr erreichbar sind?
- Lösung: Dienste und Daten redundant verfügbar halten!
  - Mehrere öffentlich verfügbare (Public Cloud) Angebote nutzen
  - Private Cloud aufbauen (hybride Cloud)

- Speicherdiens für Webobjekte
- <http://code.google.com/apis/storage/>
- Schnittstelle ist kompatibel zu S3
- Funktionalität ist mit S3 (fast) identisch
- Für jedes Objekt und jeden Bucket können Zugriffsrechte festgelegt werden
  - Access Control List (ACL)
  - <http://bucket.commondatastorage.googleapis.com/object>
  - <http://commondatastorage.googleapis.com/bucket/object>
- Zugriff auf Buckets and Objekte geschieht via REST
- Kommandozeilenwerkzeug GSUtil und Softwaredienst (SaaS) Google Storage Manager werden bereitgestellt
- GSUtil kann mit Google Storage und S3 arbeiten
  - GSUtil basiert auf der Python-Bibliothek boto

# Google Cloud Storage – Buckets

- Buckets können auf Ressourcen in den USA oder in der EU liegen

## Pricing

If you exceed your free trial quota, your Google Cloud Storage account is billed according to the following tables:

### Storage

Monthly Usage	Price (per GB)
First 0 - 1TB	\$0.12
Next 9TB	\$0.105
Next 90TB	\$0.095
Next 400TB	\$0.085
Additional Storage	<a href="#">Contact us</a>

### Network

Monthly Usage	Network (Egress) - Americas and EMEA* (per GB)	Network (Egress) - Asia Pacific (per GB)	Network (Ingress)
0 - 1TB	\$0.12	\$0.21	Free
Next 9TB	\$0.11	\$0.18	
Next 90TB	\$0.08	\$0.15	
Additional Data Transfer	<a href="#">Contact us</a>		

### Requests

PUT, POST, GET bucket**, GET service** Requests (per 1,000 requests/month)	GET, HEAD Requests (per 10,000 requests/month)
\$0.01	\$0.01

\*Europe, the Middle East, and Africa (EMEA)

\*\*GET bucket requests are API calls to list objects and GET service requests are API calls to list buckets

- Stand: März 2012

- Speicherdiens für Webobjekte
- <http://www.hosteurope.de/produkte/Cloud-Storage>
- Schnittstelle ist kompatibel zu S3
- Funktionalität ist mit S3 (fast) identisch
- Für jedes Objekt und jeden Bucket können Zugriffsrechte festgelegt werden
  - Access Control List (ACL)
  - <http://bucket.cs.hosteurope.de/objekt>
- Zugriff auf Buckets and Objekte geschieht via REST
- Kein Hochladen von Objekten via POST
- Keine garantierter Verfügbarkeit
- Nicht jede Software unterstützt Host Europe Cloud Storage
  - Unterstützung u.a. durch s3cmd, Cyberduck und Cloudberry Explorer

- Mit EC2 kann der Benutzer Instanzen virtueller Server auf in den Serverfarmen von Amazon starten, nutzen und steuern
  - Unterstützte Betriebssysteme: Linux, Windows Server 2003/2008 und OpenSolaris
- Virtuelle Server werden aus Amazon Machine Images (AMI) erzeugt
  - Blaupause für das Anlegen eines neuen virtuellen Servers
  - Amazon stellt vorgefertigte Images zur Verfügung
  - Auch Drittanbieter wie z.B. IBM, Oracle und Sun stellen AMIs mit eigenen Software-Paketen zur Verfügung
  - Auch die Benutzer können eigene Images anfertigen
  - Es ist möglich, eigene AMIs zu veröffentlichen und über eine Produkt-ID zu vermarkten (Paid Instances)

- EC2 hat Standorte (**Regionen**), mit Ressourcen
  - US East  $\Rightarrow$  Virginia
  - US West  $\Rightarrow$  Kalifornien
  - US West 2  $\Rightarrow$  Oregon
  - EU East  $\Rightarrow$  Irland
  - Asia Pacific (Southeast)  $\Rightarrow$  Singapur
  - Asia Pacific (Northeast)  $\Rightarrow$  Tokio
  - South America  $\Rightarrow$  Sao Paulo
- Jeder Standort enthält Verfügbarkeitszonen (**Availability Zones**)
- Jede Verfügbarkeitszone ist ein in sich abgeschlossener Cluster
  - us-east-1a, us-east-1b, us-east-1c, us-east-1d
  - us-west-1a, us-west-1b
  - us-west-2a, us-west-2b
  - eu-west-1a, eu-west-1b
  - ap-southeast-1a, ap-southeast-1b
  - ap-northeast-1a
  - sa-east-1a, sa-east-1b

# EC2 Instanztypen

Instanztyp	Architektur	virt. Cores	ECU	RAM	Speicher	I/O Perf.
t1.micro	32 + 64 Bit	1	max. 2	613 MB	nur EBS	Low
m1.small	32 + 64 Bit	1	1	1,7 GB	160 GB	Moderate
c1.medium	32 + 64 Bit	2	5	1,7 GB	350 GB	Moderate
m1.medium	32 + 64 Bit	2	5	3,75 GB	410 GB	Moderate
m1.large	64 Bit	2	4	7,5 GB	850 GB	High
m1.xlarge	64 Bit	4	8	15 GB	1690 GB	High
m2.xlarge	64 Bit	2	6,5	17,1 GB	420 GB	Moderate
m2.2xlarge	64 Bit	4	13	34,2 GB	850 GB	High
m2.4xlarge	64 Bit	8	26	68,4 GB	1690 GB	High
c1.xlarge	64 Bit	8	20	7 GB	1690 GB	High
cc1.4xlarge	64 Bit	8	33,5	23 GB	1690 GB	10 Gbit/s
cg1.8xlarge	64 Bit	16	88	60,5 GB	3370 GB	10 Gbit/s
cg1.4xlarge	64 Bit	8	33,5	22 GB	1690 GB	10 Gbit/s

- 1 ECU (EC2 Compute Unit) ist äquivalent zu einer 2007er AMD Opteron oder Intel Xeon CPU mit 1,0 bis 1,2 GHz
- cg1.4xlarge enthält zwei NVIDIA Tesla „Fermi“ M2050 GPUs

- Benutzer braucht ein Schlüsselpaar für die Identifikation gegenüber seiner Instanzen
  - Einloggen ohne Passwort (Public-Key Verfahren)
  - Öffentliche Schlüssel wird in den Instanzen hinterlegt
  - Private Schlüssel bleibt auf der Arbeitsumgebung des Benutzers
  - Ein neues Schlüsselpaar kann angelegt oder auf ein bestehendes Schlüsselpaar zurückgegriffen werden
- Benutzer überlegt sich, welche Ports offen sein müssen
  - Je weniger Ports offen sind, desto besser (sicherer)
  - Benutzer legt eine Security Group für die neue Instanz an, in der die gewünschten Ports freigegeben sind
  - Eventuell kann auf eine bestehende Security Group zurückgegriffen werden
- Benutzer überlegt sich, welches Betriebssystem (AMI) und welcher Instanztyp passend für die eigenen Anforderungen ist
- Instanz starten mit den Parametern aus den obigen Vorüberlegungen

- Nach dem Startprozess erhält der entstandene virtuelle Server eine dynamisch zugewiesene öffentliche und eine private IP-Adresse
  - Unter der öffentlichen Adresse ist der Rechner aus dem Internet erreichbar
  - Unter der privaten Adresse ist er für andere Instanzen in der Amazon Cloud sichtbar
- Private und öffentliche Adressen werden dynamisch bei jedem Start einer Instanz neu vergeben
  - Daher ungeeignet für den dauerhaften Betrieb eines Servers
  - Server müssen immer mal wieder neu gestartet werden
  - Lösung: Elastic IPs
- Benutzer können Elastic IPs – einmal reserviert – immer wieder eigenen Servern zuweisen

- Bei der Beendigung einer Instanz gehen alle Änderungen verloren
- Wertvolle Daten müssen außerhalb der Instanz gespeichert werden
  - Speicherung großer Mengen schwach strukturierter Daten kann in S3 erfolgen
  - Block-basierten Zugriff auf ein Speichermedium bietet EBS

# Kosten von EC2: On-Demand Instances

Region: USA Ost (Virginia)		
	Linux/UNIX-Nutzung	Windows-Nutzung
<b>Standard On-Demand Instances</b>		
Small (Standard)	\$0,080 pro Stunde	\$0,115 pro Stunde
m	\$0,160 pro Stunde	\$0,230 pro Stunde
Large	\$0,320 pro Stunde	\$0,460 pro Stunde
Extra Large	\$0,640 pro Stunde	\$0,920 pro Stunde
<b>Micro On-Demand Instances</b>		
Micro	\$0,020 pro Stunde	\$0,030 pro Stunde
<b>Hi-Memory On-Demand Instances</b>		
Extra Large	\$0,450 pro Stunde	\$0,570 pro Stunde
Double Extra Large	\$0,900 pro Stunde	\$1,140 pro Stunde
Quadruple Extra Large	\$1,800 pro Stunde	\$2,280 pro Stunde
<b>Hi-CPU On-Demand Instances</b>		
Medium	\$0,165 pro Stunde	\$0,285 pro Stunde
Extra Large	\$0,660 pro Stunde	\$1,140 pro Stunde
<b>Cluster Compute Instances</b>		
Quadruple Extra Large	\$1,300 pro Stunde	\$1,610 pro Stunde
Eight Extra Large	\$2,400 pro Stunde	\$2,970 pro Stunde
<b>Cluster GPU Instances</b>		
Quadruple Extra Large	\$2,100 pro Stunde	\$2,600 pro Stunde

- Eine Alternative zu On-Demand Instances sind **Reserved Instances**
- Noch preisgünstiger sind **Spot Instances**
- Es fallen auch je nach Bedarf Kosten an für Internet-Datenübertragung, elastische IPs, elastische Lastverteiler und CloudWatch zur Überwachung

# AWS Simple Monthly Calculator

 **amazon web services™ SIMPLE MONTHLY CALCULATOR**

NEW! - Effective July 1st 2011. Free Inbound Data Transfer, Lower Outbound Data Transfer and New Tiers and Amazon EC2 running Red Hat Enterprise Linux

FREE USAGE TIER: New Customers get free usage tier for first 12 months

Services Estimate of your Monthly Bill (\$ 186.28)

Choose region: US-East (Northern Vir. \*) Inbound Data Transfer is Free and Outbound Data Transfer is 1 GB free per region per month

Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides resizable compute capacity in the cloud. It is designed to make web-scale computing easier for developers. Amazon Elastic Block Store (EBS) provides persistent storage to Amazon EC2 Instances.

Add To Bill Clear Form

**Compute: Amazon EC2 On-Demand Instances:**

Instances	Description	Operating System	Instance Type	Usage	Detailed Monitoring
3	Web-Server	Linux/OpenSolaris	Micro	100 Hours/Month	<input checked="" type="checkbox"/>
2	Database	Linux/OpenSolaris	Micro	100 Hours/Month	<input checked="" type="checkbox"/>

**Compute: Amazon EC2 Reserved Instances:**

**Storage: Amazon EBS Volumes:**

Volumes	Description	Provisioned Storage	Average IOPS in volume	Snapshot Storage*
5	Images	10 GB-month	50	25 GB-month of Storage <input type="button" value="Edit"/>
10	Data	50 GB-month	10	0 GB-month of Storage <input type="button" value="Edit"/>

**Elastic IP:**  
Number of Elastic IPs:   
Elastic IP Non-attached Time:  Hours/Month  
Number of Elastic IP Remaps:  Times/Month \*

**Amazon EC2 Data Transfer:**

Data Transfer In:	<input type="text" value="5"/> GB/Month <input type="button" value="Edit"/>
Data Transfer Out:	<input type="text" value="60"/> GB/Month <input type="button" value="Edit"/>
Regional Data Transfer:	<input type="text" value="10"/> GB/Month <input type="button" value="Edit"/>
Public IP/Elastic IP Data Transfer:	<input type="text" value="35"/> GB/Month <input type="button" value="Edit"/>

Common Customer Samples  
Free Website on AWS  
AWS Elastic Beanstalk Default  
Marketing Web Site  
Web Application  
Media Application  
HPC Cluster  
Disaster Recovery and Backup  
European Web Application

Language: English

Reset All

- Amazon garantiert für EC2 eine jährliche Verfügbarkeit von 99.95%
- Maximale Downtime: ca. 263 Minuten (ca. 4,5 Stunden) pro Jahr
- Wenn die Verfügbarkeit unterschritten wird, bekommt der Benutzer eine Gutschrift von 10%
  - Aber was bringt eine Gutschrift über 10%, wenn die eigenen Dienste nicht mehr erreichbar sind?
- <http://aws.amazon.com/ec2-sla/>

## **Service Commitments and Service Credits**

If the Annual Uptime Percentage for a customer drops below 99.95% for the Service Year, that customer is eligible to receive a Service Credit equal to 10% of their bill (excluding one-time payments made for Reserved Instances) for the Eligible Credit Period. To file a claim, a customer does not have to wait 365 days from the day they started using the service or 365 days from their last successful claim. A customer can file a claim any time their Annual Uptime Percentage over the trailing 365 days drops below 99.95%.

- Kommandozeilenprogramme
  - Amazon EC2 Getting Started Guide  
<http://docs.amazonaws.com/AWSEC2/latest/GettingStartedGuide/>
  - Amazon EC2 Quick Reference Card  
<http://awsdocs.s3.amazonaws.com/EC2/latest/ec2-qrc.pdf>
- Firefox-Erweiterungen (Elasticfox und Hybridfox)
  - <http://s3.amazonaws.com/ec2-downloads/elasticfox.xpi>
  - <http://code.google.com/p/hybridfox/>
- Webanwendungen/SaaS (AWS Management Console, Ylastic, KOALA)
  - <http://aws.amazon.com/console/>
  - <http://ylastic.com>
  - <http://koalacloud.appspot.com>

# AWS Management Console (Instanzen)

The screenshot shows the AWS Management Console interface for managing EC2 instances. The top navigation bar includes links for File, Edit, View, History, Bookmarks, Favorites, Help, and the AWS logo. The main content area is titled "My Instances" under the "Amazon EC2" tab. A sidebar on the left provides navigation links for EC2 Dashboard, Instances (selected), Spot Requests, AMIs, Bundle Tasks, Elastic Block Store, Volumes, Snapshots, Networking & Security, Elastic IPs, Security Groups, Key Pairs, and Load Balancers.

The main table displays 9 instances:

Instance	AMI ID	Root Device Type	Type	Status	Lifecycle	Public DNS	Security Groups
i-e5509aae	ami-f0f61599	instance-store	m1.large	running	normal	ec2-204-236-242-21.compute-1.amazonaws.com	default
i-fb509aa0	ami-f0f61599	instance-store	m1.large	running	normal	ec2-204-236-202-136.compute-1.amazonaws.com	default
i-95509aa2	ami-19a34270	instance-store	c1.medium	running	normal	ec2-174-129-122-119.compute-1.amazonaws.com	default
i-c3509aa8	ami-dcf615b5	instance-store	m1.small	running	normal	ec2-184-73-13-39.compute-1.amazonaws.com	default
i-af509ac4	ami-e348af8a	instance-store	m1.small	running	normal	ec2-204-236-250-62.compute-1.amazonaws.com	default
i-ad509ac6	ami-e348af8a	instance-store	m1.small	running	normal	ec2-75-101-249-132.compute-1.amazonaws.com	default
i-b3509ad8	ami-f0f61599	instance-store	m1.xlarge	running	normal	ec2-184-73-25-106.compute-1.amazonaws.com	default
i-f5509aa9e	ami-19a34270	instance-store	c1.medium	terminated	normal		default
i-cb509aa0	ami-19a34270	instance-store	c1.medium	terminated	normal		default

At the bottom, a summary indicates "2 EC2 Instances selected" and lists them as "EC2 Instances: i-cb509aa0, i-f5509aa9e". The "Description" tab is active, showing the instance IDs i-cb509aa0 and i-f5509aa9e.

© 2008 - 2009, Amazon Web Services LLC or its affiliates. All rights reserved. | Feedback | Support | Privacy Policy | Terms of Use | An [amazon.com](#) company

# AWS Management Console (AMIs)

The screenshot shows the AWS Management Console interface for launching a new instance. The main window is titled "Request Instances Wizard" and is currently on the "CHOOSE AN AMI" tab. A search bar at the top right contains the text "debian". Below it, a table lists various AMIs, all of which are of type "Debian". Each row includes an "AMI ID", "Root Device", "Manifest", "Platform", and a "Select" button with a dropdown arrow.

AMI ID	Root Device	Manifest	Platform	Action
ami-01d03c68	instance-store	debianzone-us-east-1/ami/lenny/postgresql/20100210/32/le	Debian	Select
ami-0256b16b	instance-store	alestic/debian-6.0-squeeze-desktop-20090419.manifest.xml	Debian	Select
ami-045db96d	instance-store	alestic-64/debian-4.0-etch-base-64-20081222.manifest.xml	Debian	Select
ami-055db96c	instance-store	alestic-64/debian-5.0-lenny-base-64-20081222.manifest.xml	Debian	Select
ami-099e7a60	instance-store	alestic/debian-4.0-etch-base-20080803.manifest.xml	Debian	Select
ami-0af51663	instance-store	alestic/debian-5.0-lenny-desktop-20091011.manifest.xml	Debian	Select
ami-0b5eba62	instance-store	alestic/debian-5.0-lenny-desktop-20081222.manifest.xml	Debian	Select
ami-0c638165	ebs	369929726641/EB-Debian-lenny-64-EBS-2009120901	Debian	Select
ami-0d5eba64	instance-store	alestic/debian-4.0-etch-desktop-20081222.manifest.xml	Debian	Select
ami-0ed43067	instance-store	alestic-64/debian-4.0-etch-base-64-20080922.manifest.xml	Debian	Select
ami-0fd03c66	instance-store	debianzone-us-east-1/ami/lenny/mysql/20100210/32/lenny-	Debian	Select
ami-0fd43066	instance-store	alestic-64/debian-5.0-lenny-base-64-20080922.manifest.xml	Debian	Select

Navigation sidebar on the left:

- Amazon EC2
- Navigation
  - Region: US East (N. Virginia)
  - EC2 Dashboard
  - Instances
    - Instances
    - Spot Requests
  - Images
    - AMIs
    - Bundle Tasks
  - Elastic Block Store
    - Volumes
    - Snapshots
  - Networking & Security
    - Elastic IPs
    - Security Groups
    - Key Pairs
    - Load Balancing

Bottom footer:

© 2008 - 2009, Amazon Web Services LLC or its affiliates. All rights reserved. | Feedback | Support | Privacy Policy | Terms of Use | An company

# AWS Management Console (Elastic IPs)

The screenshot shows the AWS Management Console interface for managing Elastic IPs. The top navigation bar includes links for File, Edit, View, History, Bookmarks, Favorites, Help, and a search bar for Google. The main menu bar has options like Datei, Bearbeiten, Ansicht, Chronik, Lesezeichen, Extras, and Hilfe.

The left sidebar navigation menu includes:

- Amazon EC2** (selected)
- Amazon Elastic MapReduce
- Amazon CloudFront
- Navigation**
  - Region:** US East
  - EC2 Dashboard**
  - INSTANCES**
    - Instances**
    - Spot Requests**
  - IMAGES**
    - AMIs**
    - Bundle Tasks**
  - ELASTIC BLOCK STORE**
    - Volumes**
    - Snapshots**
  - NETWORKING & SECURITY**
    - Elastic IPs**
    - Security Groups**
    - Key Pairs**
    - Load Balancers**

Addresses	
<a href="#">Allocate New Address</a>   <a href="#">Release Address</a>   <a href="#">Associate</a>   <a href="#">Disassociate</a>	
Viewing: All Addresses	
Address	Instance ID
<input type="checkbox"/> 184.73.225.185	
<input checked="" type="checkbox"/> 184.73.226.126	i-91569cfa
<input type="checkbox"/> 184.73.226.168	
<input type="checkbox"/> 184.73.226.171	i-93569cf8

1 Elastic IP selected	
<b>Address:</b>	184.73.226.126
<b>Instance ID:</b>	i-91569cfa

© 2008 - 2009, Amazon Web Services LLC or its affiliates. All right reserved. | [Feedback](#) | [Support](#) | [Privacy Policy](#) | [Terms of Use](#) | An company

# Mit der EC2 API und boto arbeiten

- Einfacher Zugriff auf die EC2 API mit boto und Python  
<http://code.google.com/p/boto/>

```
#!/usr/bin/env python
```

```
from boto.ec2.connection import EC2Connection
# Verbindung zu EC2 aufbauen
# Variable "conn" zeigt auf ein "EC2Connection" Objekt
conn = EC2Connection('<aws access key>', '<aws secret key>')

# Liste aller Regionen erhalten und ausgeben
liste_regionen = conn.get_all_regions()
print liste_regionen

# Liste aller Zonen erhalten und ausgeben
liste_zonen = conn.get_all_zones()
print liste_zonen
```

# Mit der EC2 API und boto arbeiten (Fortsetzung)

```
# Liste aller Sicherheitsgruppen erhalten und ausgeben
liste_gruppen = conn.get_all_security_groups()
print liste_gruppen

# Liste aller Schluesselpaare erhalten und ausgeben
liste_keys = conn.get_all_key_pairs()
print liste_keys

# Instanz(en) starten
reservation = conn.run_instances('ami-e348af8a',
                                  min_count=2,
                                  key_name='geheim',
                                  instance_type='m1.small')

# Liste aller Instanzen erhalten und ausgeben
liste_instanzen = conn.get_all_instances()
print liste_instanzen
```

- EBS ist ein Teil von EC2
- Innerhalb jeder Verfügbarkeitszone können Benutzer EBS-Volumes erzeugen
  - Größe: 1 GB bis 1 TB
- Ein EBS-Volume ist persistenter Speicher
- Ein neues Volume ist wie ein unformatiertes Block-Gerät
- Ein Volume kann immer an genau eine Instanz angehängt werden
  - Dafür müssen sich Volume und Instanz zu innerhalb der gleichen Verfügbarkeitszone befinden
- Ein Volume kann ein beliebiges Dateisystem enthalten
- Ein Volume ist von der Bedienung her exakt wie ein USB-Stick
- Merke: EBS ist Speicher für Menschen und S3 für Anwendungen
- Volume-Snapshots können erzeugt (und in S3 gespeichert) werden

# AWS Management Console (EBS-Volumes)

The screenshot shows the AWS Management Console interface for managing EBS volumes. The left sidebar navigation includes sections for EC2 Dashboard, Instances, AMIs, Bundle Tasks, Volumes, Snapshots, Elastic IPs, Security Groups, Key Pairs, and Load Balancers. The main content area is titled "EBS Volumes" and displays a table of six volumes. The table columns are Volume ID, Capacity, Snapshot, Created, Zone, Status, and Attachment Information. The volumes listed are:

Volume ID	Capacity	Snapshot	Created	Zone	Status	Attachment Information
vol-4d5a8a24	10 GiB	—	2010-03-03 14:24 GMT+0100	us-east-1c	in-use	i-93569cf8:/dev/sdf (attached)
vol-435a8a2a	1 GiB	—	2010-03-03 14:24 GMT+0100	us-east-1d	available	
vol-415a8a28	15 GiB	—	2010-03-03 14:24 GMT+0100	us-east-1c	in-use	i-91569cfa:/dev/sdi (attached)
vol-755a8a1c	5 GiB	—	2010-03-03 14:24 GMT+0100	us-east-1a	available	
vol-4b5a8a22	2 GiB	—	2010-03-03 14:24 GMT+0100	us-east-1b	available	
vol-03ef246a	1 GiB	—	2010-02-23 10:40 GMT+0100	us-east-1a	available	

Below the table, a section titled "1 Elastic Block Store Volume selected" provides detailed information for the selected volume (Volume ID: vol-415a8a28). The details are:

Volume ID:	vol-415a8a28
Capacity:	15 GiB
Snapshot:	
Status:	in-use
Attachment:	i-91569cfa:/dev/sdi (attached)
Created:	2010-03-03 14:24 GMT+0100

At the bottom of the page, there is a footer with links to Feedback, Support, Privacy Policy, Terms of Use, and a note that it is an Amazon.com company.

# Mit der EBS und boto arbeiten

```
#!/usr/bin/env python

from boto.ec2.connection import EC2Connection
# Verbindung zu EC2 aufbauen
# Variable "conn" zeigt auf ein "EC2Connection" Objekt
conn = EC2Connection('<aws access key>', '<aws secret key>')

# Volume (1 GB) in Region "us-east-1a" erzeugen und ID ausgeben
volume = conn.create_volume(1, 'us-east-1a')
print volume.id

# Volume "vol-1e0f0677" entfernen
conn.delete_volume('vol-1e0f0677')
```

- Volume an Instanz anhängen  $\Rightarrow$  attach\_volume()
- Volume von Instanz lösen  $\Rightarrow$  detach\_volume()

# Amazon Elastic Load Balancing (ELB)

- ELB ist ein Teil von EC2
- Innerhalb jeder Verfügbarkeitszone können Benutzer elastische Lastverteiler erzeugen
- Jedem Lastverteiler weist der Benutzer einen Pool an Instanzen zu
- Ein elastischer Lastverteiler verteilt automatisch die eingehenden Anfragen über die EC2-Instanzen seines Pools
- Ein ELB ermittelt fehlerhafte Instanzen innerhalb seines Pools und leitet die Anfragen automatisch an fehlerfreie Instanzen im Pool weiter

**Load Balancer Name:**

**Listener Configuration:**

Common Applications	Protocol	Load Balancer Port	EC2 Instance Port	Actions
Apache HTTP Server	HTTP	80	80	<button>Remove</button>
Secure HTTP Server	HTTPS (Secure HTTP)	443	80	<button>Remove</button>
Custom...	TCP	21	21	<button>Save</button>

- Hauptunterschiede zwischen den Diensten untereinander:
  - Funktionalität
  - Verfügbarkeit bzw. Qualität
  - Preis
  - **Schnittstelle**  $\iff$  Wird häufig nicht erst genug genommen

Wie arbeiten die Benutzer/Kunden mit den Cloud-Diensten?

- Interaktion mit Diensten
  - SaaS  $\implies$  Browser
  - PaaS, IaaS  $\implies$  Protokollorientierte Programmierschnittstelle (API)
    - Werkzeuge der Dienst- und Drittanbieter verwenden die **jeweilige API**
- Gefahren
  - Datenschutz und Datensicherheit
  - Parental Computing
  - Lock-in (!!!)

- Auswahl eines Betriebssystems oder einer Programmiersprache hat Auswirkungen u.a. auf:
  - Betriebssystem: Verwendbare Anwendungen, Sicherheit,...
  - Programmiersprache: Portabilität, verfügbare Bibliotheken,...
- Auswahl einer Schnittstelle hat u.U. noch gravierendere Auswirkungen

## Ein Gedankenspiel

- Wechsel des Energieversorgers  $\Rightarrow$  Auswirkungen auf meine Geräte?
- Wechsel des Gasversorgers  $\Rightarrow$  Auswirkungen auf meine Heizung?
- Wechsel des Telefonanbieters  $\Rightarrow$  Auswirkungen auf mein Telefon?
- Wechsel des Cloud-Dienstanbieters  $\Rightarrow$  Auswirkungen auf meine Software bzw. mein Unternehmen?

- Entscheidet sich ein Kunde für einen öffentlich verfügbaren Dienst (PaaS oder IaaS), entscheidet er sich auch für eine Schnittstelle
- Gefahr des **Lock-in**
  - Abhängigkeit zwischen Dienstnutzer und -anbieter
- Denkbare Szenarien: Preiserhöhung, Änderung des Dienstangebots (Funktionalität), Insolvenz des Anbieters, . . .
- Wechsel des Anbieters nur bei gleichzeitigem **Verlust** der Infrastruktur (**Dienste**) und eventuell sogar der **Daten**
  - Auswirkungen für Kunden (insbesondere Unternehmen) u.U. fatal
- Verwendet man einen Dienst langfristig, **investiert** man in diesen
  - Dienste werden *veredelt*
  - Das eigene Geschäftsmodell wird darauf ausgerichtet
  - Mitarbeiter werden geschult



## Dropbox

- Beispiel: Dropbox
- 2007 gegründeter Webdienst
- Stellt ein Netzwerk-Dateisystem für die Synchronisation von Dateien zwischen verschiedenen Rechnern und Benutzern bereit
- Verwendet zur Datenspeicherung Amazon S3
  - Geschäftsmodell: Einen Cloud-Dienst veredeln

- Was passiert mit Dropbox, wenn S3 die Preise verdoppelt oder auf einmal nicht mehr existiert?
- Was wären die Auswirkungen für die Kunden von Dropbox und S3?
- Was kann man gegen die Gefahr des Lock-in tun?

- **Wettbewerber**

- Bieten öffentliche Dienste mit gleicher Funktionalität und Schnittstelle an
- Beispiele
  - Für S3 API: Google Storage, Host Europe Cloud Storage

- **(Freie) Lösungen**

- Aufbau privater Dienste mit gleicher Funktionalität und Schnittstelle
  - Beispiele
    - Für EC2 API: Eucalyptus, Nimbus, OpenNebula, CloudStack, OpenStack
    - Für S3 API: Walrus (Eucalyptus), Cumulus (Nimbus), Swift (OpenStack)
    - Für GAE API: AppScale, typhoonAE
  - Idealerweise kann man damit hybride Clouds realisieren
- 
- Existieren Wettbewerber und/oder (freie) Lösungen mit kompatibler Schnittstelle, eröffnen sich viele Möglichkeiten

- Benutzer interagieren direkt mit Infrastruktur- und Speicherdienssten
  - Dienste sind meist Insellösungen
- Werkzeuge der Anbieter unterstützen meist nur wenige Dienste

The screenshot shows the KoalaCloud web interface. At the top, there's a navigation bar with 'Logout' and language selection (German and English). Below it, a dropdown menu says 'Amazon EC2 (US East)' and 'Region wechseln'. A status bar indicates 'Aktive Region: Amazon (us-east-1)'. A dropdown menu for 'Funktionalität/Dienst auswählen' is open. The main area displays two instances:

	Status	Type	Reservation	Root	Besitzer	Image	Kernel	Ramdisk	Zone	Gruppe
Instance 1	running	m1.large	r-5a489337	instance-store	178412210831	ami-68ac7801	aki-427d952b	None	us-east-1a	default
Instance 2	running	i-1e1b0f28d			ec2-184-73-1-108.compute-1.amazonaws.com	ip-10-124-170-46.ec2.internal				

- Wünschenswert: **Marktplatzportal** ⇒ KOALA
  - Dienst, der öffentliche und private Dienste verschiedener Anbieter unter einer GUI integriert
- Kein Werkzeug unterstützt mehr Cloud-Dienste
- Entwickelt für die GAE
  - Läuft auch im privaten Kontext
- Dienst: <http://koalacloud.appspot.com>
- <http://code.google.com/p/koalacloud/>
- Herausforderungen bei der Entwicklung:
  - Fähigkeiten der (privaten) Dienste
  - Qualität der Dokumentationen

- Dienst, der S3-kompatible Dienste zu einem RAID-1 zusammenfügt



- Vorteile:
  - Höhere Verfügbarkeit der Daten
  - Unabhängigkeit von einzelnen Anbietern
- Entwickelt für die GAE

- Kopiert Objekte zu den genutzten Diensten und kontrolliert deren Synchronität anhand der MD5-Prüfsummen
- Dienst: <http://clodoctopus.appspot.com>
- Projektseite: <http://code.google.com/p/octopuscloud/>
- Herausforderung bei der Entwicklung:
  - Verhalten der Dienste nicht zu 100% identisch

- Anbieter betreiben virtualisierte Ressourcen als Dienste und ermöglichen den Zugriff über das Netz via Web Services
- Fundament des Cloud Computing
  - Virtualisierung zur gemeinsamen und effizienten Ressourcennutzung
  - Web-Services zur Kommunikation zwischen Kunden und Diensten sowie zwischen den Cloud-Diensten selbst
- Organisatorische Unterscheidung der Dienste
  - Public Cloud, Private Cloud, Hybrid Cloud
- Funktionale Unterscheidung der Dienste
  - SaaS, PaaS, IaaS
- Gefahr des Lock-in
  - Wettbewerber und Lösungen für private Dienste vermeiden diese Gefahr
  - Möglichkeit zum Daten-Export sollte immer geprüft werden
    - Portabilität der Daten beachten

10 TB Daten sollen aus einer Cloud exportiert werden

- Wie lange dauert die Übertragung via DSL mit 16.000 Kbit/s?

# Aufgabe zur Portabilität von Daten

Daten in der Cloud (10 TB) 10.000.000.000.000 Byte

Bandbreite des DSL (16.000 Kbit/s) 16.000.000 Bit/s

Bandbreite des DSL in Byte/s 2.000.000 Byte/s

$$10.000.000.000.000 \text{ Byte} / 2.000.000 \text{ Byte/s} = 5.000.000 \text{ s}$$

$$\text{Dauer der Datenübertragung [s]} = 5.000.000 : 60$$

$$\text{Dauer der Datenübertragung [min]} = 83.333,\overline{3} : 60$$

$$\text{Dauer der Datenübertragung [h]} = 1.388,\overline{8} : 24$$

$$\text{Dauer der Datenübertragung [d]} \approx 57,87$$

⇒ ca. 57 Tage, 20 Stunden, 53 Minuten