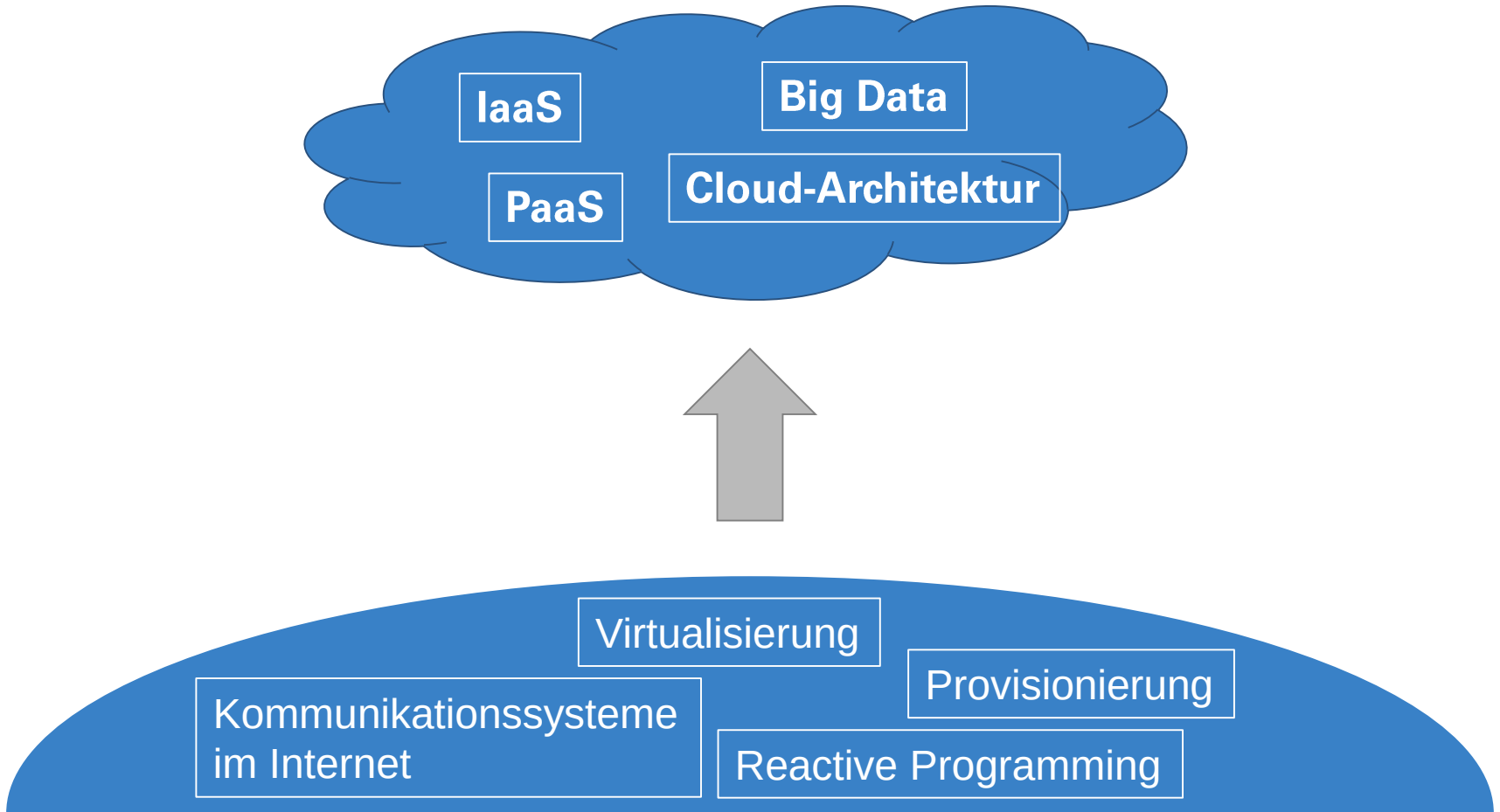
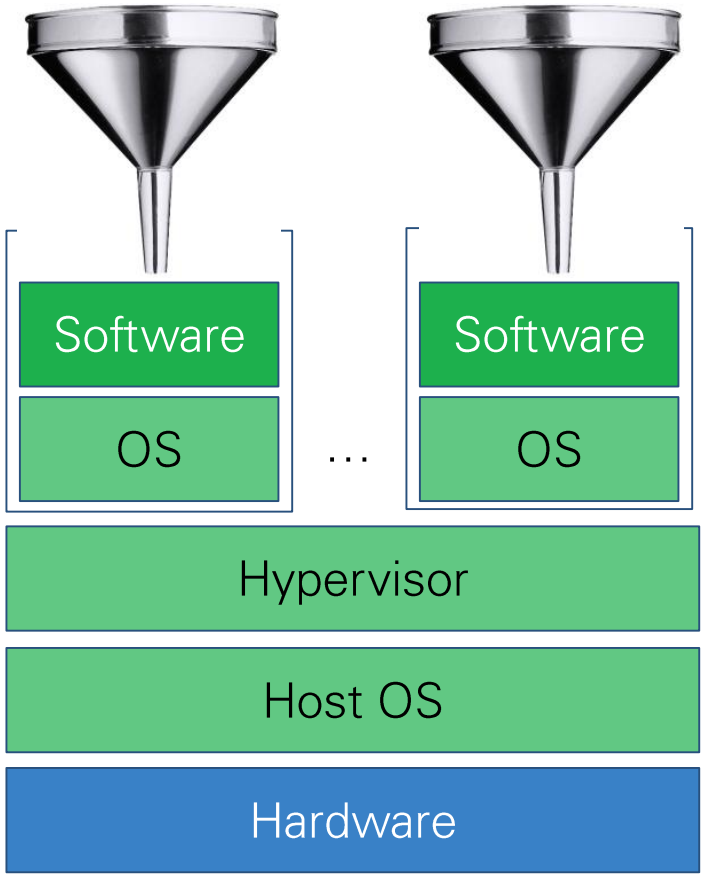
QAWARE
SOFTWARE ENGINEERING

# Kapitel 5: Infrastructure-as-a-Service

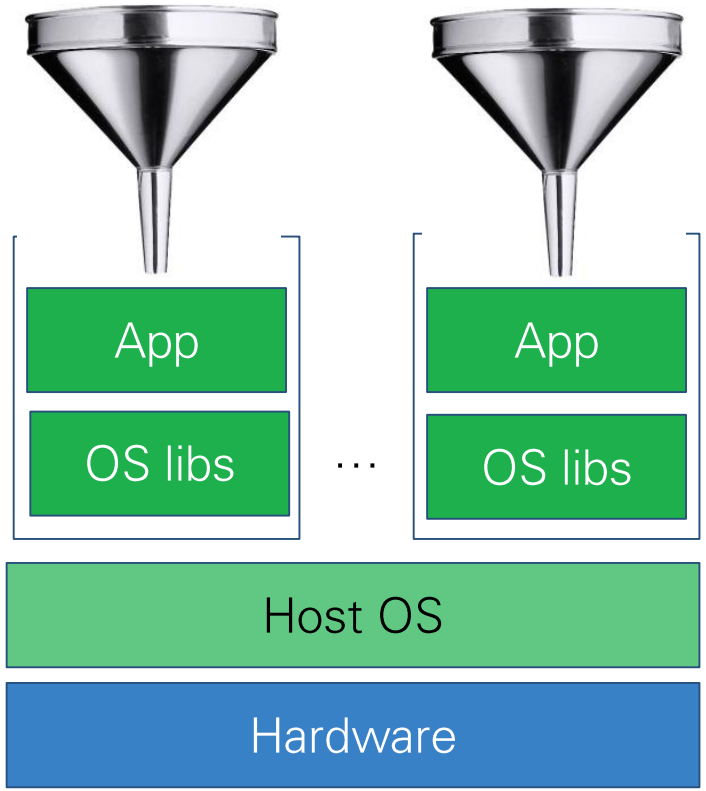
# Ab heute sind wir in der Cloud.



# Die letzte Vorlesung: Wie kommt Software auf das Blech?

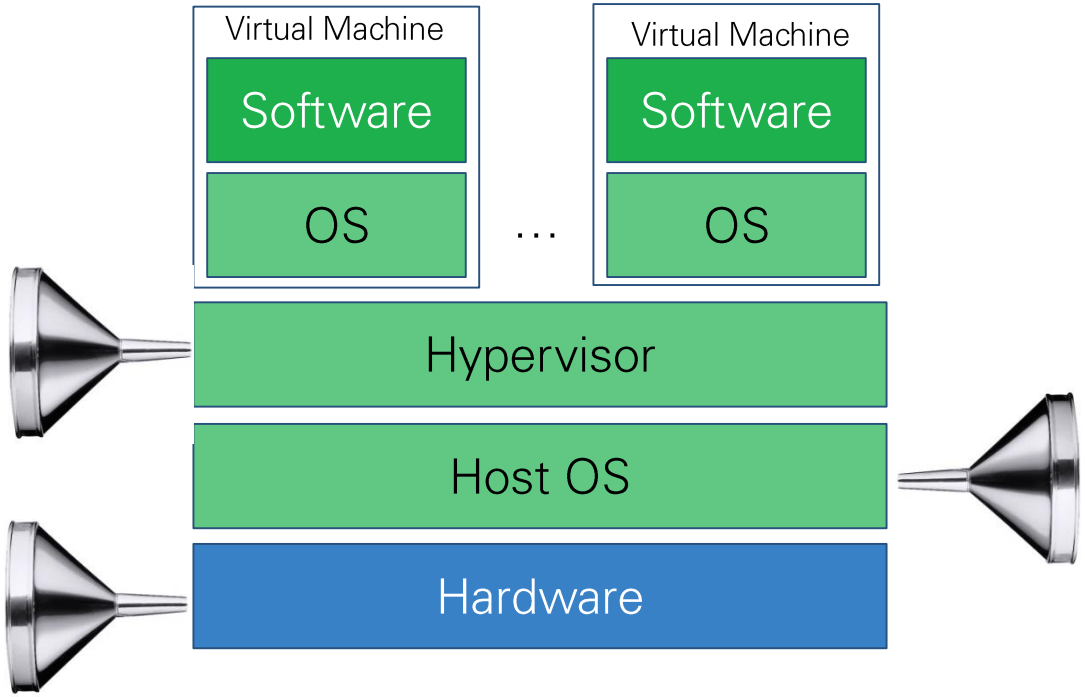



Hardware-Virtualisierung

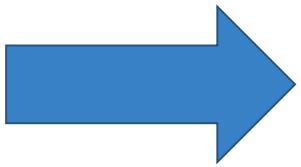


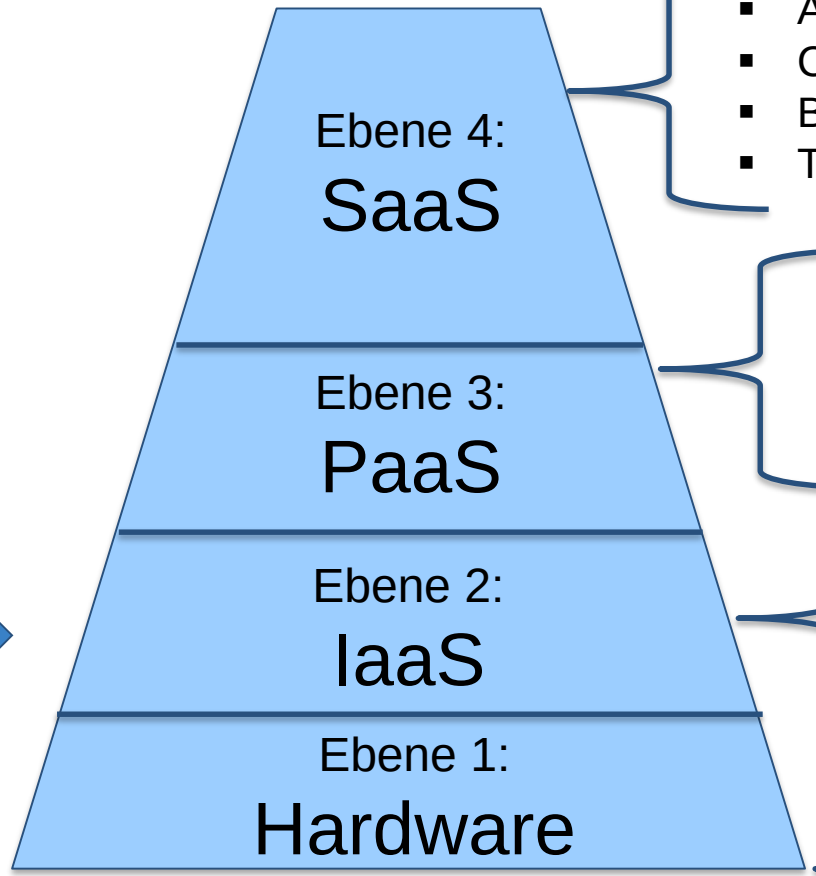

Betriebssystem-Virtualisierung

# Heute: Wie kommt Software an das Blech?



# Das Schichtenmodell des Cloud Computing: Vom Blech zur Anwendung.






- Anpassbare Softwaredienste
- CaaS: Component-as-a-Serivce (z.B. Google Charts)
- BaaS: Backend-as-a-Service
- Transparente Update

- Plattformdienste
- Programmierschnittstellen (APIs)
- Abstraktion der technischen Infrastruktur

- Elastizität
- Virtuelle Ressourcenpools
- Technische Infrastruktur:
  Maschinen, Server (DNS, DHCP, LB, NAS, …)

- Rechner
- Netzwerk
- Storage

# Einführung: Infrastructure-as-a-Service

# Time2System im letzten Jahrhundert: > 1 Jahr.


http://de.wikipedia.org/wiki/Gro%C3%9Frechner

# Time2System in der Cloud-Ära: In Echtzeit.

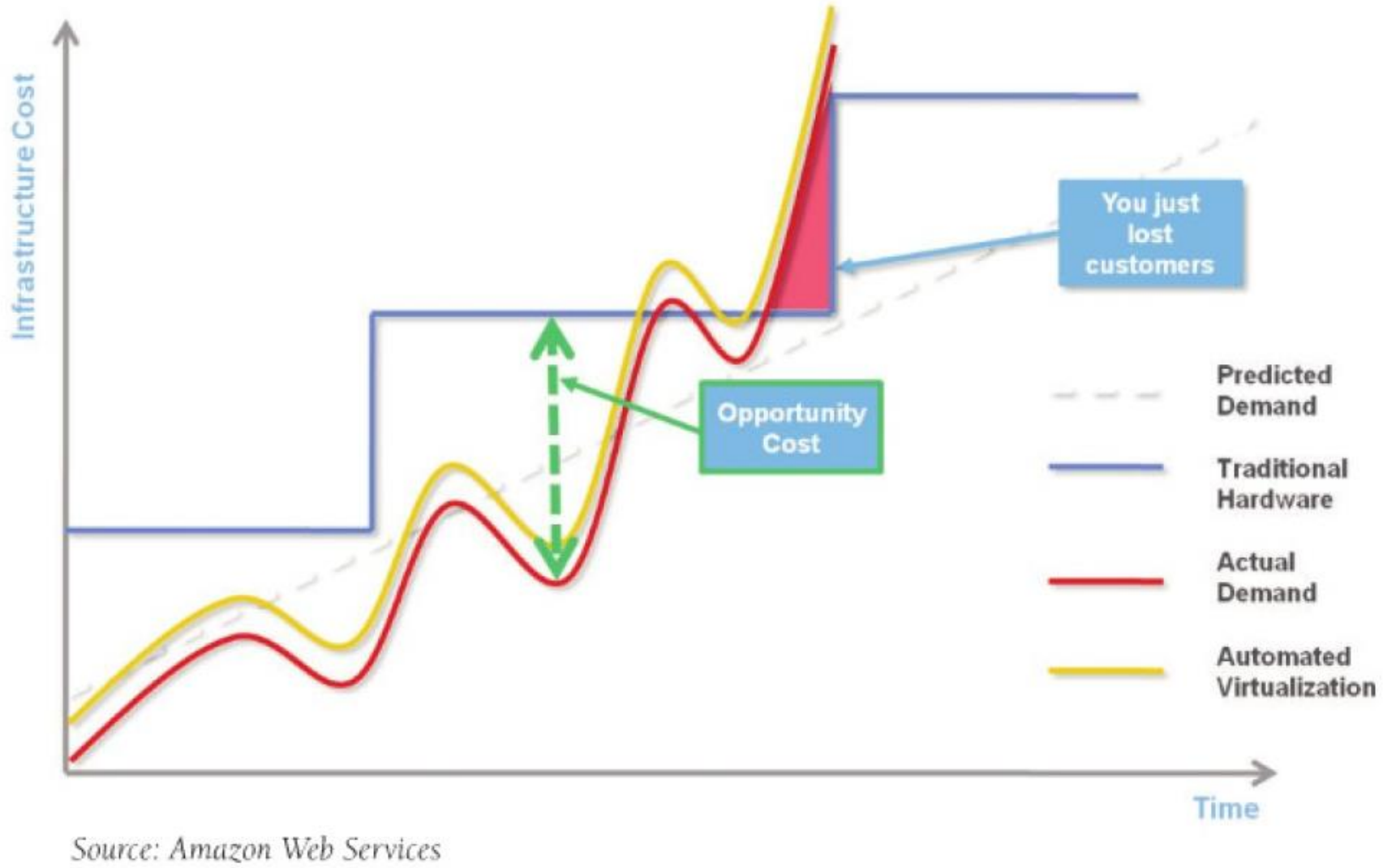
## Slashdot-Effekt

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 „geslashdottet“ (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 scherzhaft 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.

# Klassische Betriebsszenarien werden bei dynamischer Nachfrage teuer. Hohe Opportunitätskosten.




*Source: Amazon Web Services*

# Definition IaaS

Unter *IaaS* versteht man ein Geschäftsmodell, das entgegen dem klassischen Kaufen von Rechnerinfrastruktur vorsieht, diese je nach Bedarf anzumieten und freizugeben.

Eigenschaften einer IaaS-Cloud:

- **Ressourcen-Pools**: Verfügbarkeit von scheinbar unbegrenzten Ressourcen, die Anfragen verteilt verarbeiten.
- **Elastizität**: Dynamische Zuweisung von zusätzlichen Ressourcen bei Bedarf.
- **Pay-as-you-go Modell**: Abgerechnet werden nur verbrauchte Ressourcen.

Ressourcen-Typen in einer IaaS-Cloud:

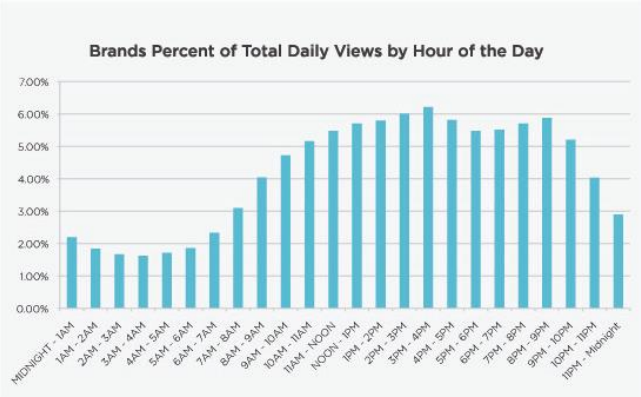
- **Rechenleistung**: Rechner-Knoten mit CPU, RAM und HD-Speicher.
- **Speicher**: Storage-Kapazitäten als Dateisystem-Mounts oder Datenbanken.
- **Netzwerk**: Netzwerk und Netzwerk-Dienste wie DNS, DHCP, VPN, CDN und Load Balancer.

Infrastruktur-Dienste einer IaaS-Cloud:

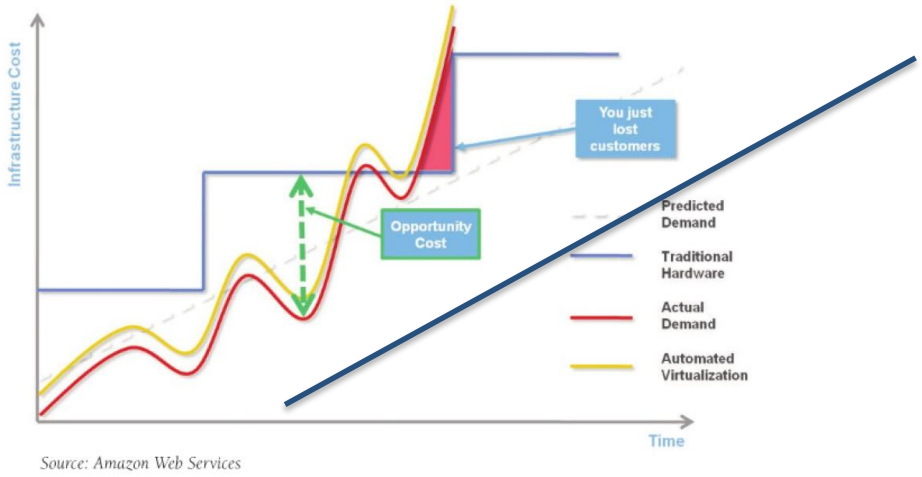
- **Monitoring**
- **Ressourcen-Management**

# Skalierbarkeit: Effekte

- **Tageszeitliche und saisonale Effekte**: Mittags-Peak, Prime-Time-Peak, Wochenend-Peak, Weihnachten, Valentinstag, Muttertag, … (vorhersehbare Belastungsspitzen)

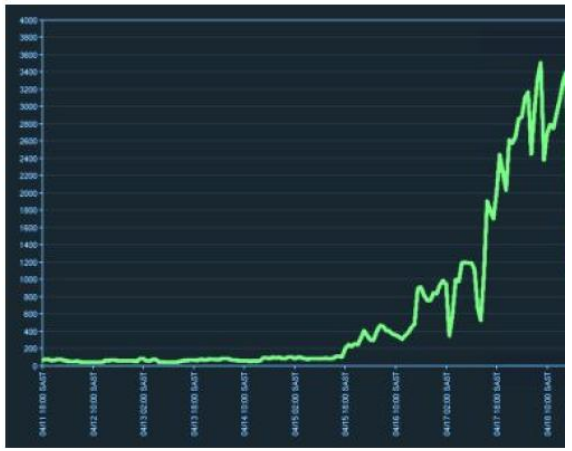



- **Kontinuierliches Wachstum**

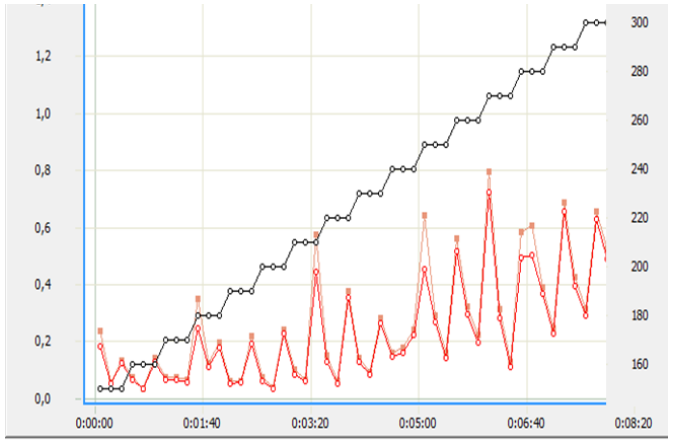



*Source: Amazon Web Services*

- **Sondereffekte**: z.B. Slashdot-Effekt (unvorhersehbare Belastungsspitzen)



- **Temporäre Plattformen**: Projekte, Tests, …



# Elastizitätsarten

**Nachfrageelastizität**: Die allokierten Ressourcen steigen / sinken mit der Nachfrage.

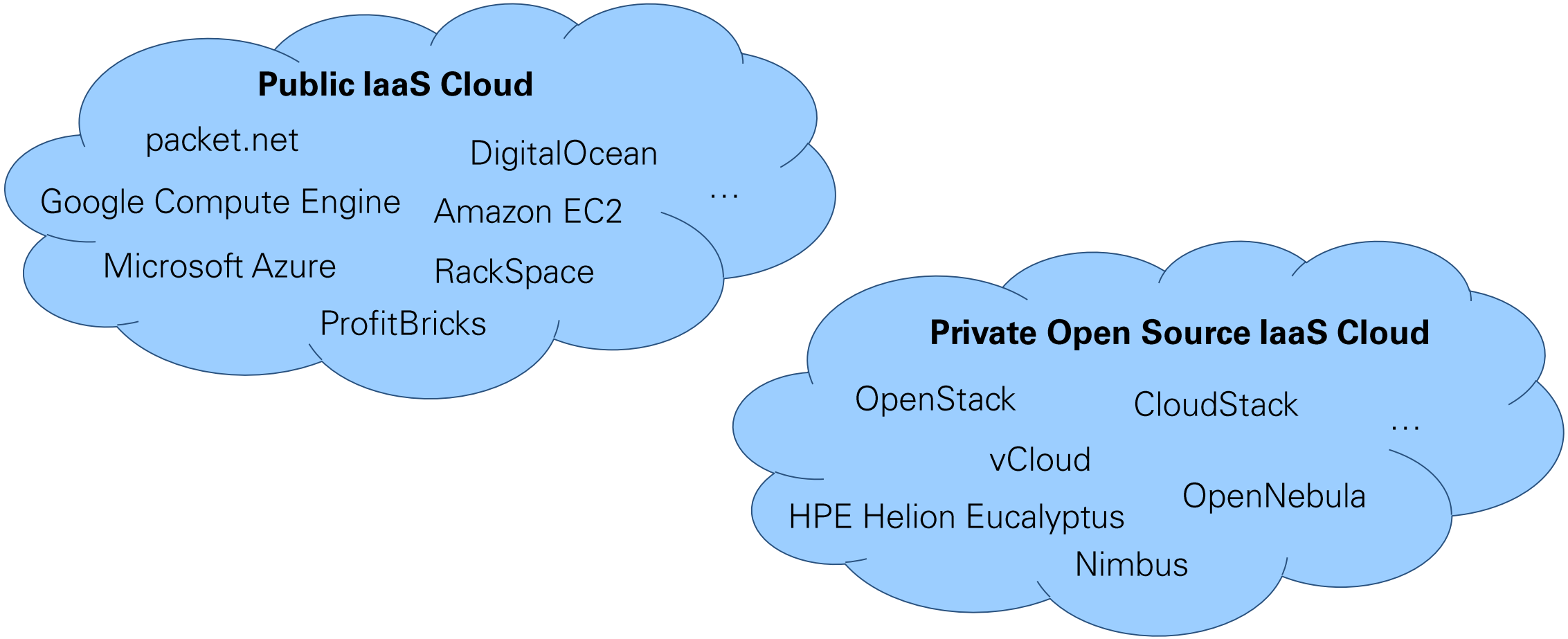
- Pseudo-Elastizität: Schneller Aufbau. Kurze Kündigungsfrist.
- Echtzeit-Elastizität: Allokation und Freigabe von Ressourcen innerhalb von Sekunden. Automatisierter Prozess mit manuellen Triggern oder nach Zeitplan.
- Selbstadaptive Elastizität: Automatische Allokation und Freigabe von Ressourcen in Echtzeit  auf Basis von Regeln und Metriken.

**Angebotselastizität**: Die allokierten Ressourcen steigen / sinken mit dem Angebot.

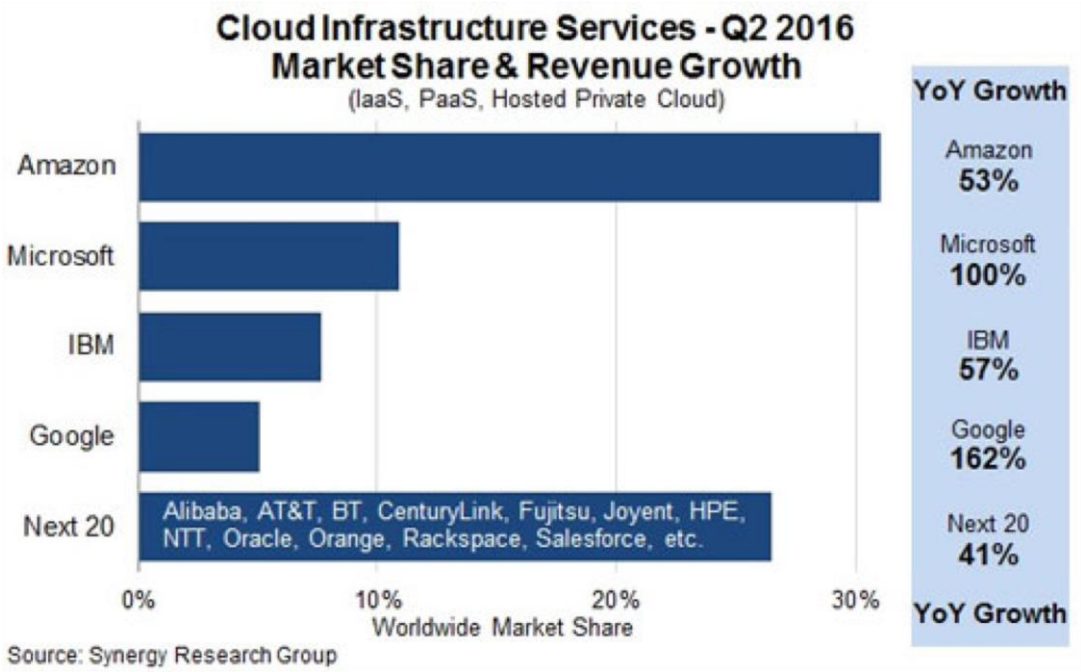
- Dies ist das typische Verhalten eines Grids: Alle verfügbaren Rechner werden allokiert.
- Es sind auch Varianten verfügbar, bei denen man für freie Ressourcen bieten kann.

**Einkommenselastizität**: Die allokierten Ressourcen steigen / sinken mit dem Einkommen bzw. dem Budget.

# Es gibt vielerlei Anbieter für Public und Private IaaS Clouds.



# Der momentane IaaS Markt.


2016 Magic Quadrant for Cloud Infrastructure as a Service, Worldwide, Gartner
https://aws.amazon.com/de/resources/gartner-2015-mq-learn-more

# Es gibt eine Reihe an gängigen Kriterien bei der Auswahl einer passenden IaaS-Cloud.

- Unterstützte Cloud-Varianten (Private Cloud, Public Cloud, Hybrid Cloud, …)
- Zuverlässigkeit / Verfügbarkeit
- Sicherheit und Datenschutz
- Vorhersagbare und stabile Performance
- Preismodell: Fixe und flexible Kosten
- Skalierbarkeit: Grenzen, Automatismen und Reaktionszeiten
- Lock-In der Daten und Anwendungen: Offene APIs
- Haftung
- Support

# Ein Service Level Agreement (SLA) ist ein Vertrag mit Zuverlässigkeitszusagen für Ressourcen und Dienste.

**Verfügbarkeitsklassen:**

| Availability % | Downtime per Year | Downtime per Month | Downtime per Week |
|---|---|---|---|
| 99.9% (three nines) | 8.76 hours | 43.2 minutes | 10.1 minutes |
| 99.95% | 4.38 hours | 21.56 minutes | 5.04 minutes |
| 99.99% (four nines) | 52.6 minutes | 4.32 minutes | 1.01 minutes |
| 99.999% (five nines) | 5.26 minutes | 25.9 seconds | 6.05 seconds |
| 99.9999% (six nines) | 31.5 seconds | 2.59 seconds | .0605 seconds |

Beispiel: Amazon S3 (Storage)

**Service Commitment**

AWS will use commercially reasonable efforts to make Amazon S3 available with a Monthly Uptime Percentage (defined below) of at least 99.9% during any monthly billing cycle (the "Service Commitment"). In the event Amazon S3 does not meet the Service Commitment, you will be eligible to receive a Service Credit as described below.

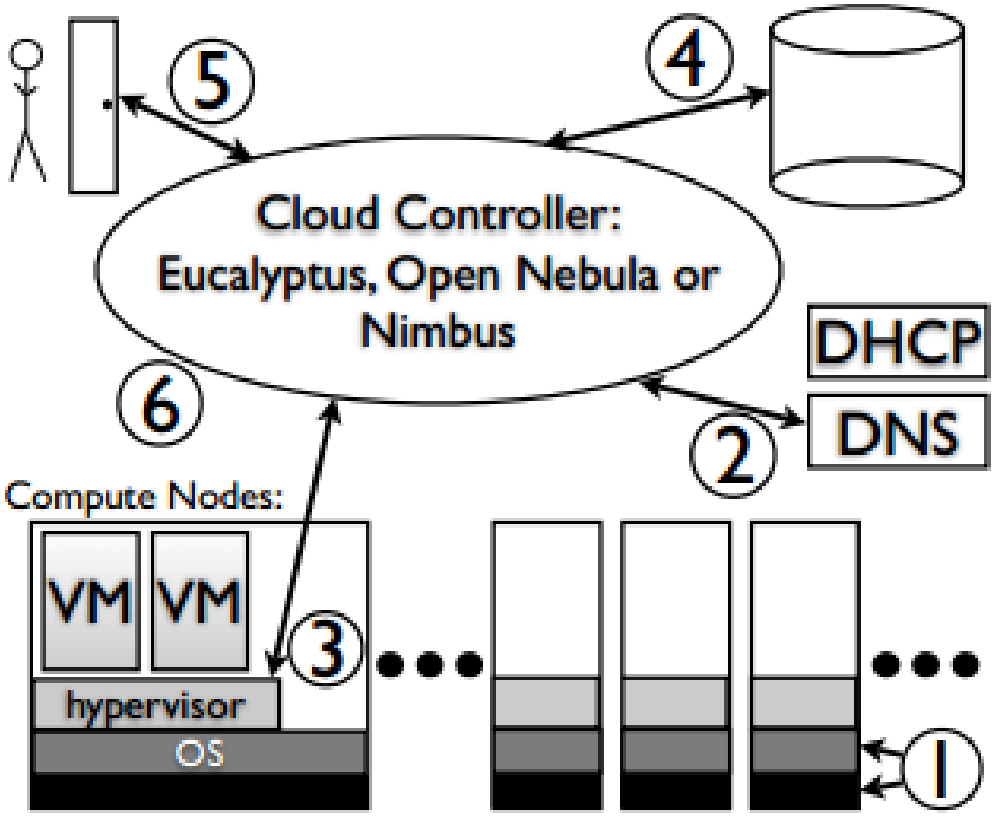
| Monthly Uptime Percentage | Service Credit Percentage |
|---|---|
| Equal to or greater than 99% but less than 99.9% | 10% |
| less than 99% | 25% |

# Aspekte der Sicherheit in einer IaaS-Cloud.

- Vertraulichkeit der Daten und Datenkommunikation: Datenverschlüsselung, VPNs
- Nachvollziehbarkeit der Daten: Einhaltung nationaler Gesetze (z.B. EU-Datenschutzbestimmung, US Patriot Act) durch geographische Datenhaltung
- Firewalls und starke Authentifizierungsverfahren
- Backup der VMs, Storages und Datenbanken
- Zertifizierungen: ISO 27001, TÜV IT
- Siehe auch Sopot Memorandum: http://datenschutz-berlin.de/content/nachrichten/datenschutznachrichten/%2027-april-2012

# Architektur einer IaaS-Cloud

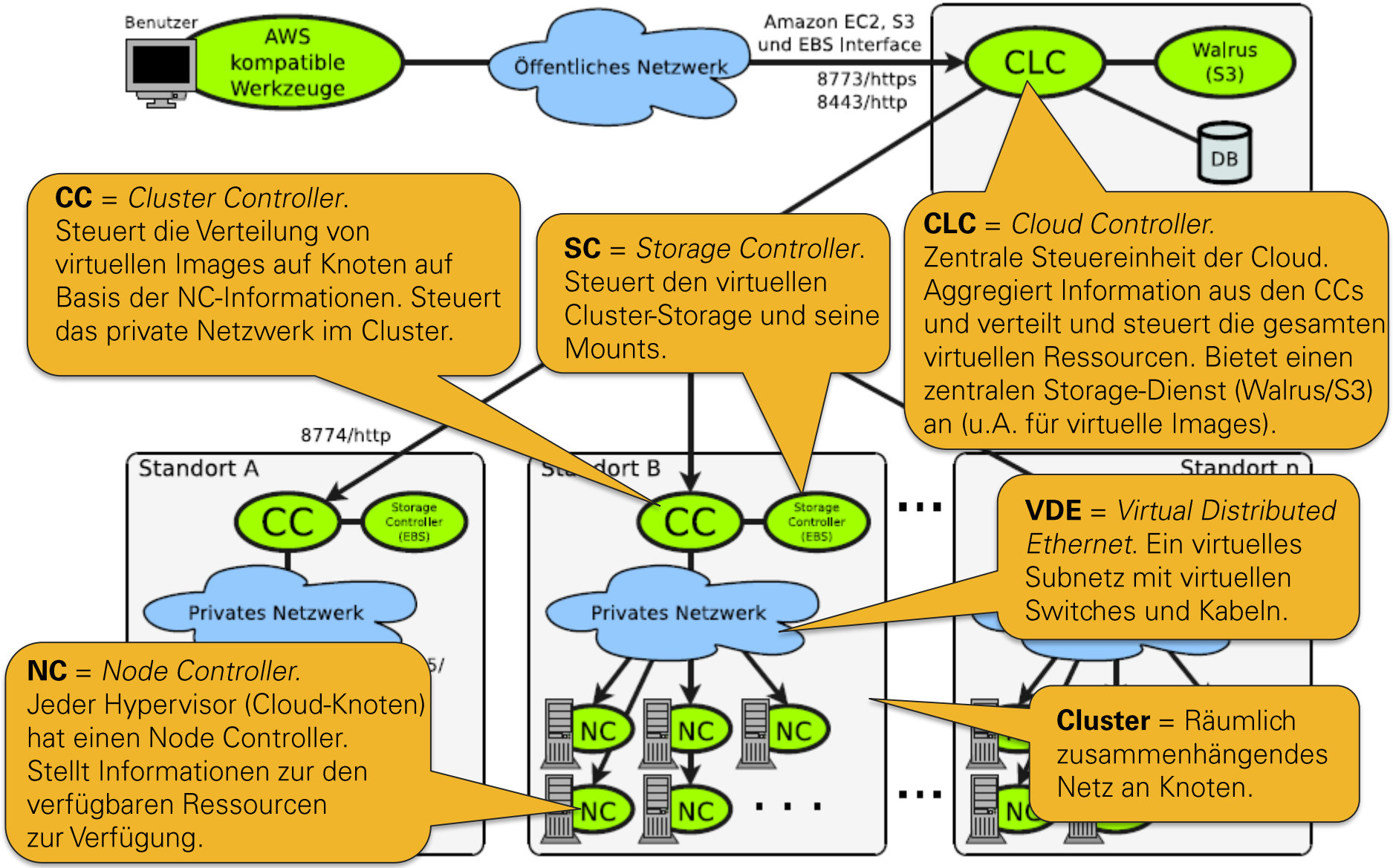
# Eine IaaS-Referenzarchitektur.




1. Hardware und Betriebssystem
2. Virtuelles Netzwerk und Netzwerkdienste
3. Virtualisierung
4. Datenspeicher und Image-Verwaltung
5. Managementschnittstelle für Administratoren und Benutzer
6. Cloud Controller für das mandantenspezifische Management der Cloud-Ressourcen

Peter Sempolinski and Douglas Thain,
**"A Comparison and Critique of Eucalyptus, OpenNebula and Nimbus",**
IEEE International Conference on Cloud Computing Technology and Science, 2010.

# Der interne Aufbau einer IaaS-Cloud am Beispiel Eucalyptus.




**CC** = *Cluster Controller.* Steuert die Verteilung von virtuellen Images auf Knoten auf Basis der NC-Informationen. Steuert das private Netzwerk im Cluster.

**SC** = *Storage Controller.* Steuert den virtuellen Cluster-Storage und seine Mounts.

**CLC** = *Cloud Controller.* Zentrale Steuereinheit der Cloud. Aggregiert Information aus den CCs und verteilt und steuert die gesamten virtuellen Ressourcen. Bietet einen zentralen Storage-Dienst (Walrus/S3) an (u.A. für virtuelle Images).

**VDE** = *Virtual Distributed Ethernet.* Ein virtuelles Subnetz mit virtuellen Switches und Kabeln.

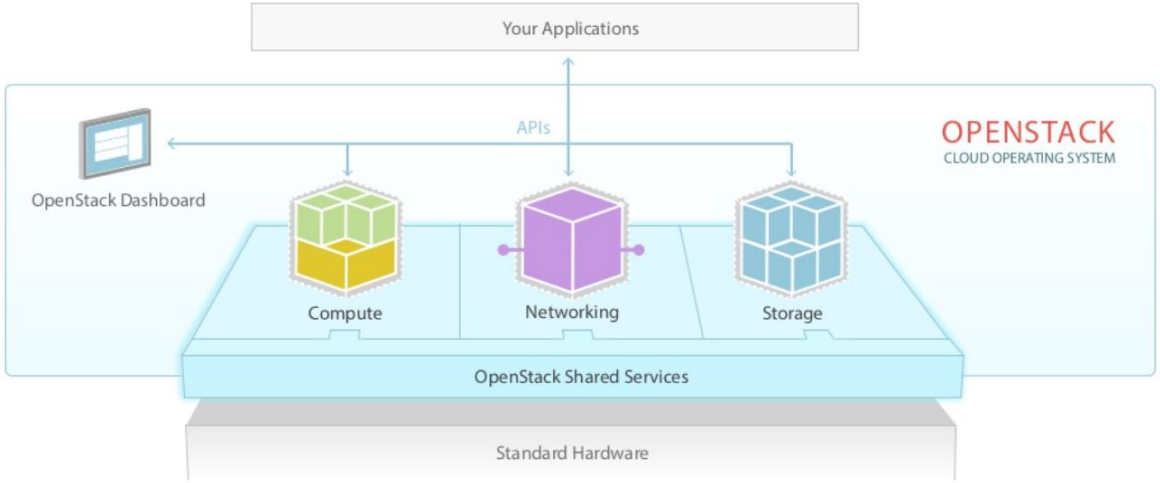
**NC** = *Node Controller.* Jeder Hypervisor (Cloud-Knoten) hat einen Node Controller. Stellt Informationen zur den verfügbaren Ressourcen zur Verfügung.

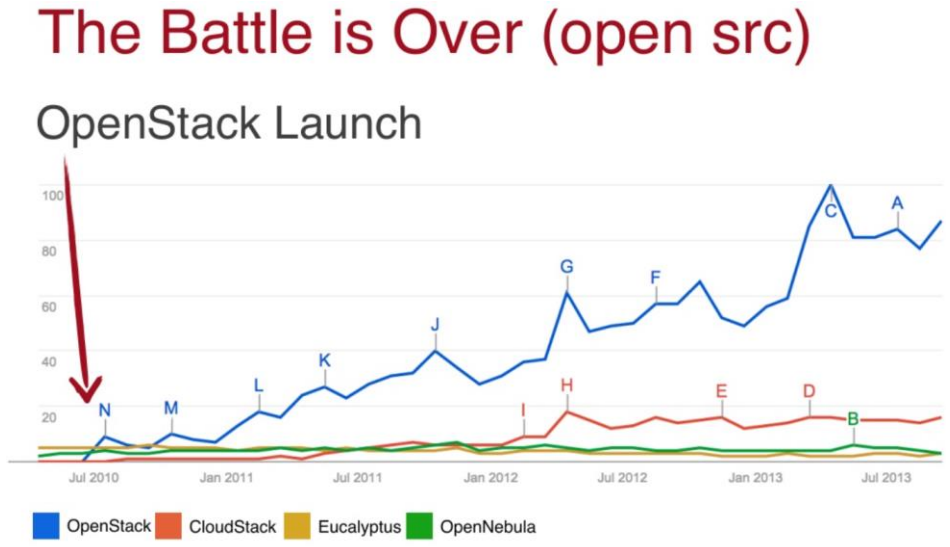
**Cluster** = Räumlich zusammenhängendes Netz an Knoten.

# IaaS mit OpenStack

# OpenStack: Der de-facto Standard für Open-Source Private IaaS Clouds.

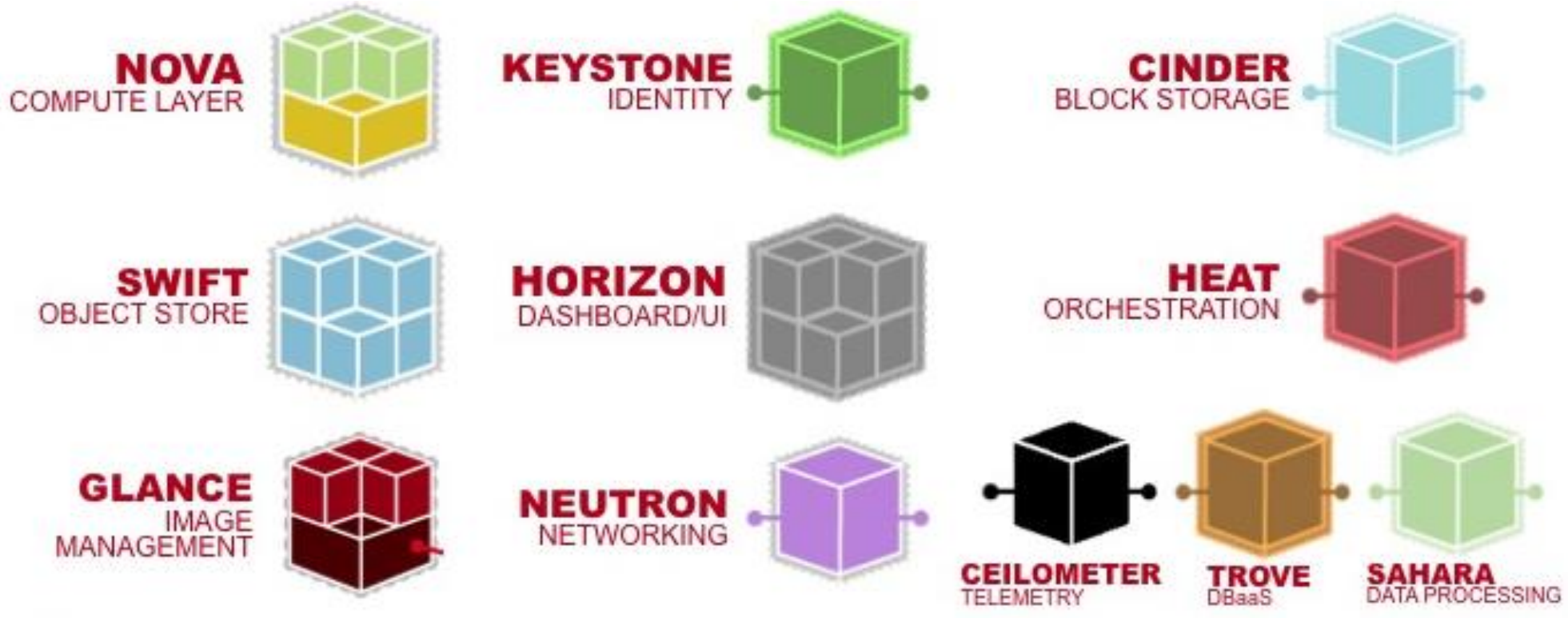
- Open Source Projekt wurde maßgeblich initiiert von RackSpace und der NASA.
- Das erste vollständige Release erfolgte im Oktober 2010.
- Lizenziert unter der Apache Lizenz.
- Eine Vielzahl der klassischen IT-Player (SAP, IBM, vmWare, HP, Oracle, Cisco) sind Teil der OpenStack-Community.
- Sehr aktives Open-Source-Projekt mit > 400 aktiven Committern.
- Ausgelegt eher als Framework denn als fertiges System für IaaS-Clouds.



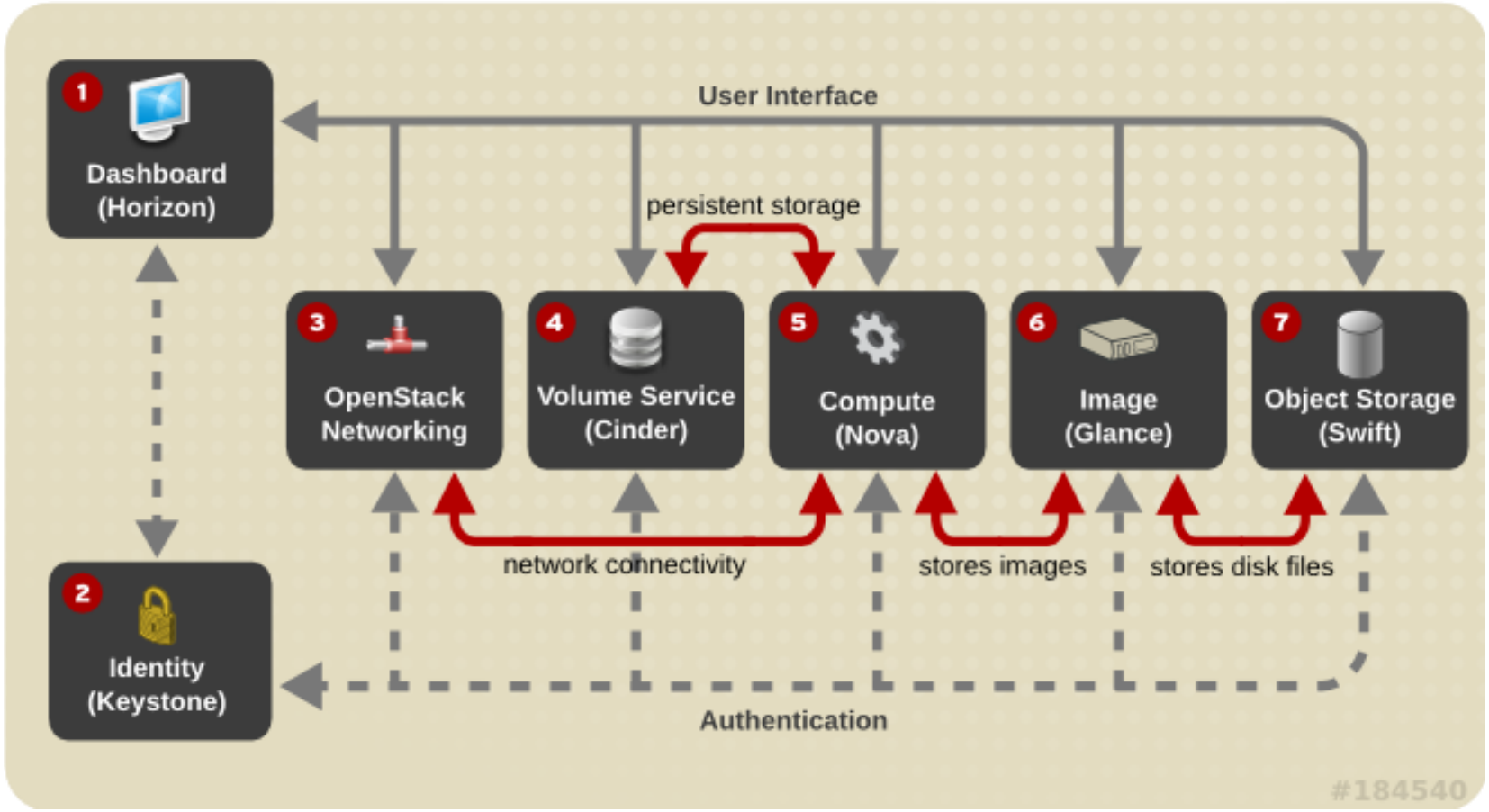

Quellen: http://de.slideshare.net/randybias/state-of-the-stack-v2

# Die OpenStack Komponenten.




Quelle: http://de.slideshare.net/sgordon2/deep-dive-openstack-summit-red-hat-summit-2014

# Das Zusammenspiel der Kern-Komponenten in OpenStack.

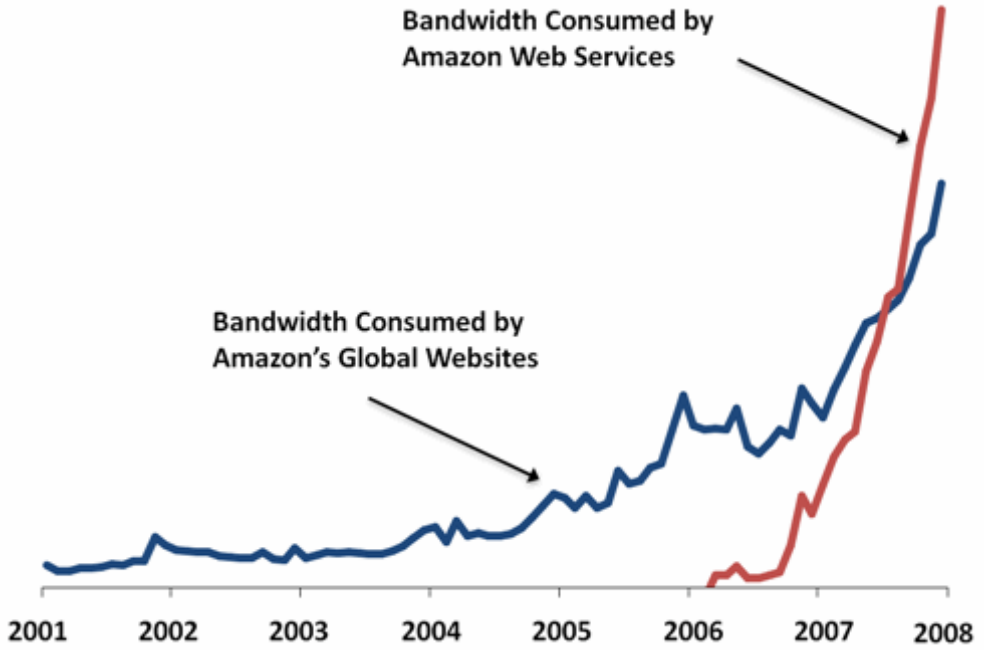



Quelle: https://access.redhat.com/documentation/en-US/Red_Hat_Enterprise_Linux_OpenStack_Platform/2/html/Getting_Started_Guide/ch01.html

# IaaS mit Amazon EC2

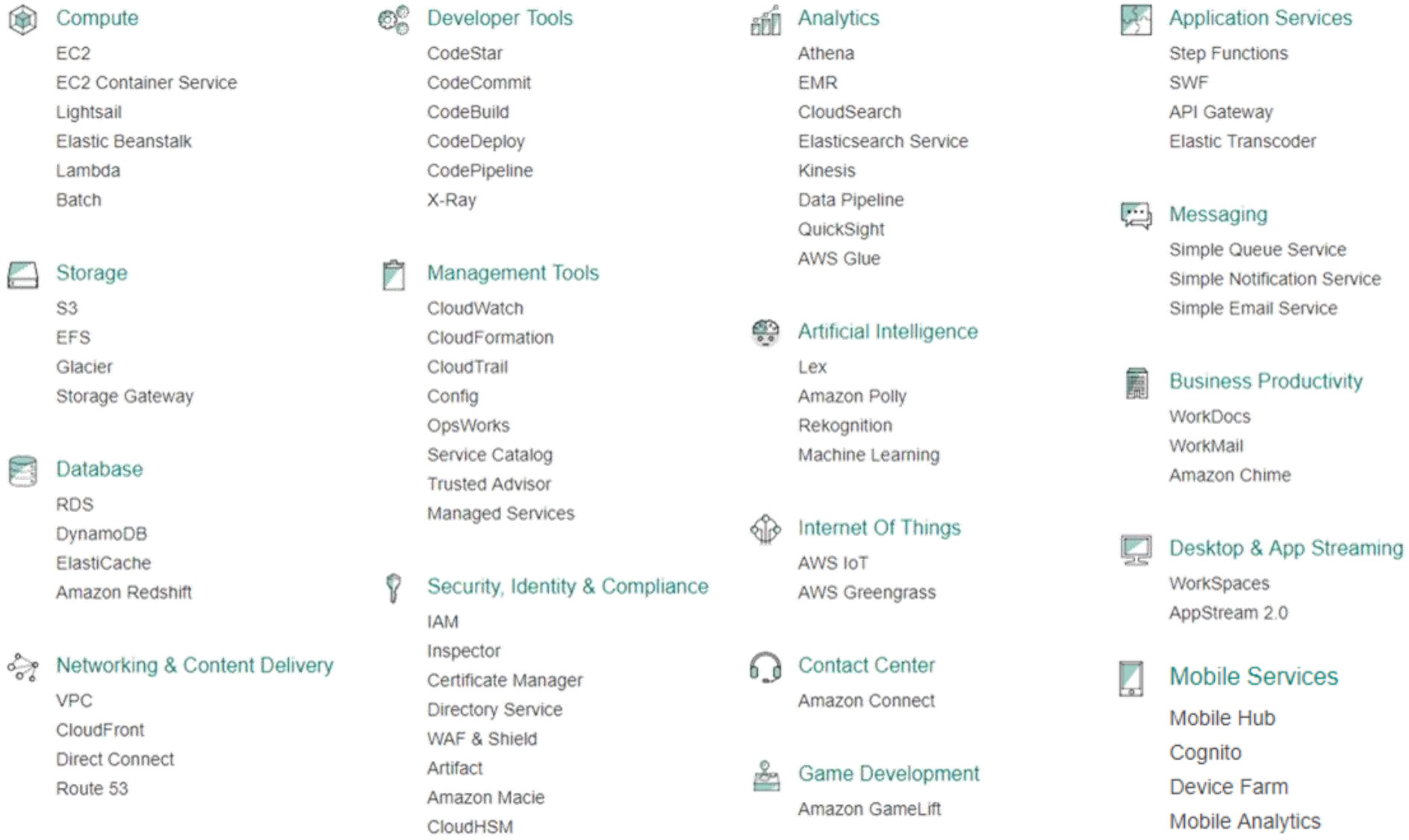
# Die Amazon EC2 IaaS Cloud.

- Amazon bietet im Rahmen der AWS (Amazon Web Service) auch eine IaaS-Cloud an.
- Historie:
  - Start innerhalb von Amazon im Jahr 2001
  - Öffentliche Beta ab 25. August 2006
  - Ab Mitte 2007 mehr Bandbreite durch Dritte in der Cloud konsumiert, als durch die Amazon Webseiten
  - Produktionsreife ab 23. Oktober 2008
  - 2005 bis 2012 ca. 12 Mrd. $ Investment in die Infrastruktur
  - 2015: 1,5 bis 2 Mio. Server in 10 globalen Rechenzentren.
- On-Demand-, Reserved- und Spot-Instanzen in verschiedenen Größen: (http://aws.amazon.com/de/ec2/instance-types) sowie diverse Storage- und Netzwerkdienste.




http://aws.typepad.com/aws/2008/05/lots-of-bits.html

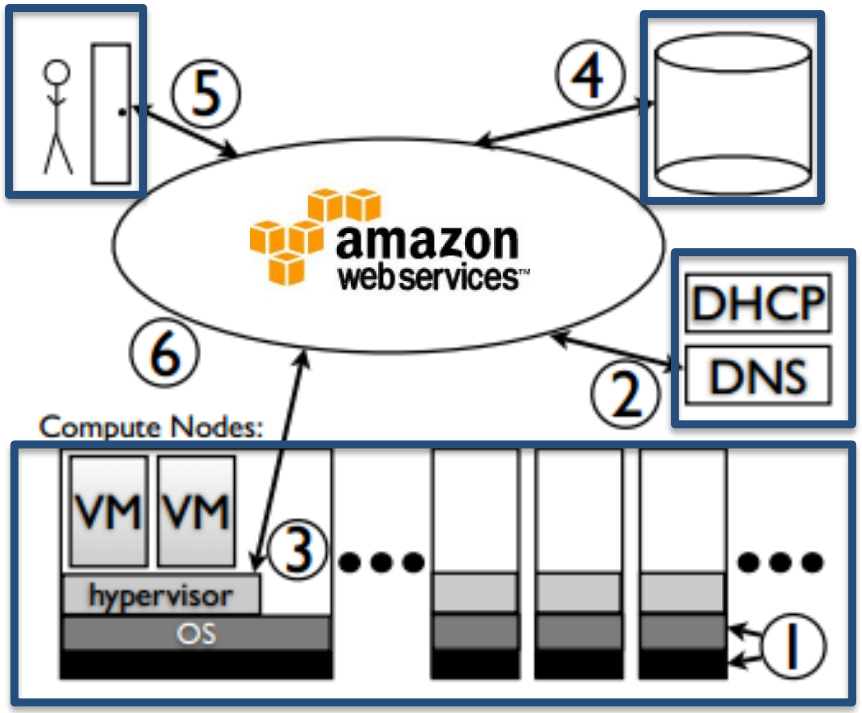
# Neben der Amazon EC2 IaaS Cloud bietet Amazon noch viele weitere IaaS-Komponenten, PaaS- und SaaS-Dienste.



# Architektur der Amazon EC2.

- AWS Management
- Console
- Webservice-API

- EBS (Elastic Block Store)
- S3 (Simple Storage Service)




- DNS / DHCP
- Elastic IPs
- VPC (Virtual Private Cloud)
- Elastic Load Balancer
- CloudFront CDN

- EC2-Knoten mit Xen- und HVM-Virtualisierung
- Monitoring über CloudWatch
- AutoScaling auf Basis von CloudWatch-Metriken

# Die globale Verteilung der Amazon EC2.

**Region und Anzahl der Availability Zones**

**USA Ost**
Nord-Virginia (6), Ohio (3)

**USA West**
Nordkalifornien (3), Oregon (3)

**Asien-Pazifik**
Mumbai (2), Seoul (2), Singapur (2), Sydney (3), Tokio (3)

**Kanada**
Zentral (2)

**China**
Peking (2)

**Europa**
Frankfurt (3), Irland (3), London (2)

**Südamerika**
São Paulo (3)

**AWS GovCloud (US-West)** (2)

**Neue Region (in Kürze verfügbar)**

**Bahrain**

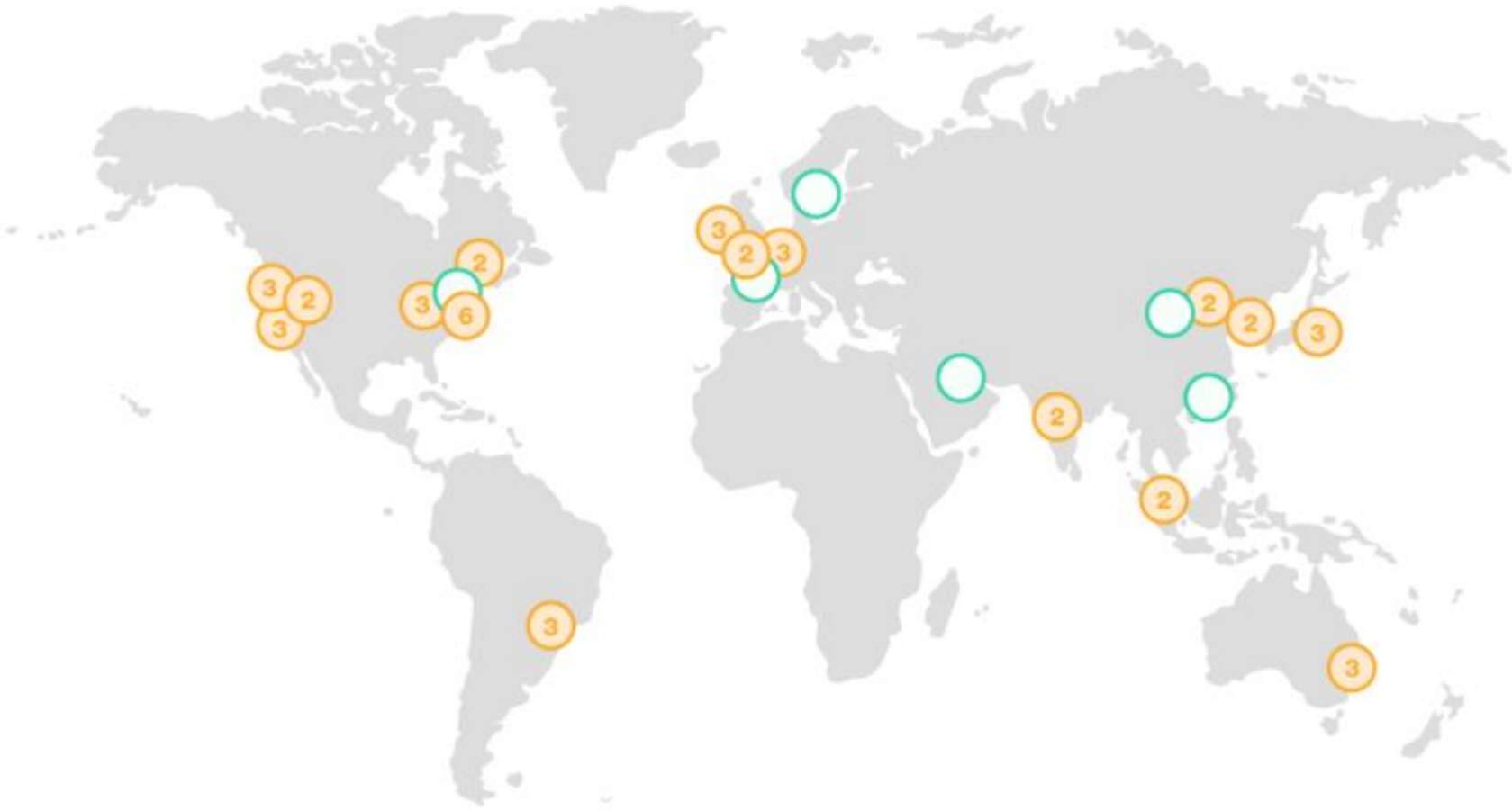
**China**

**Frankreich**

**Hongkong**

**Schweden**

**AWS GovCloud (US-East)**



# Sicherheitsaspekte der Amazon EC2.

- Zertifiziert nach ISO 27001 (Empfehlung BSI). Im deutschen und irischen Datencenter den EU-Datenschutzrichtlinien unterworfen. Amazon ist ebenso global dem US Patriot Act unterworfen.
- Daten und Instanzen können global auf alle Rechenzentren verteilt werden. Jedes dieser Rechenzentren besteht aus mehreren Verfügbarkeitszonen, die ein in sich geschlossenes Rechen-Cluster darstellen.
- Jede EC2-Instanz muss einer Security Group zugeordnet sein. Eine Security Group ist die Konfiguration der Inbound-Firewall für Instanzen.
- Der Zugriff auf die EC2-Administrationsfunktionen kann über das zentrale IAM-System gesteuert werden. Es können Benutzer angelegt, autorisiert und authentifiziert werden. Für den Zugriff per API können Zugriffsschlüssel und Zertifikate vergeben und widerrufen werden. Eine Multi-Faktor-Authentifizierung wird unterstützt.
- Zugriff auf Linux-Instanzen per SSH. Authentifzierung an der Instanz über SSH-Zertifikat (Keypair) und Benutzername („root“/“ec2-user“/„ubuntu“).
- Zugriff auf Windows-Instanzen per Remote Desktop. Das Admin-Passwort für die Maschine kann per Weboberfläche / API abgefragt werden.

# Über die AWS Management Console können alle Dienste der Amazon-Cloud gesteuert werden.