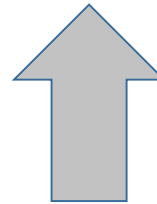


Cloud Computing

Kapitel 5: Infrastructure-as-a-Service

Dr. Josef Adersberger

Ab heute sind wir in der Cloud



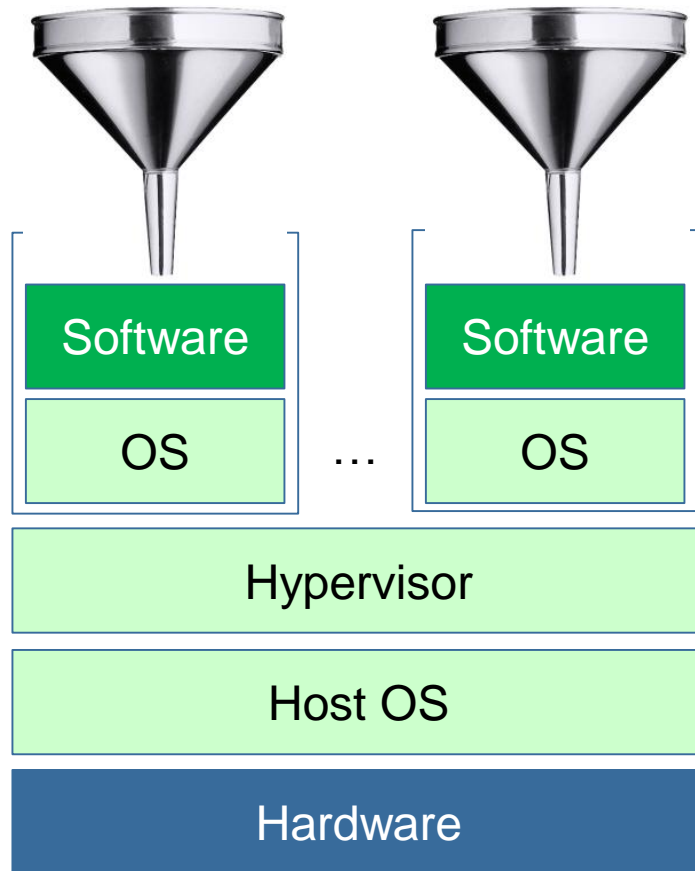
Kommunikationssysteme
im Internet

Virtualisierung

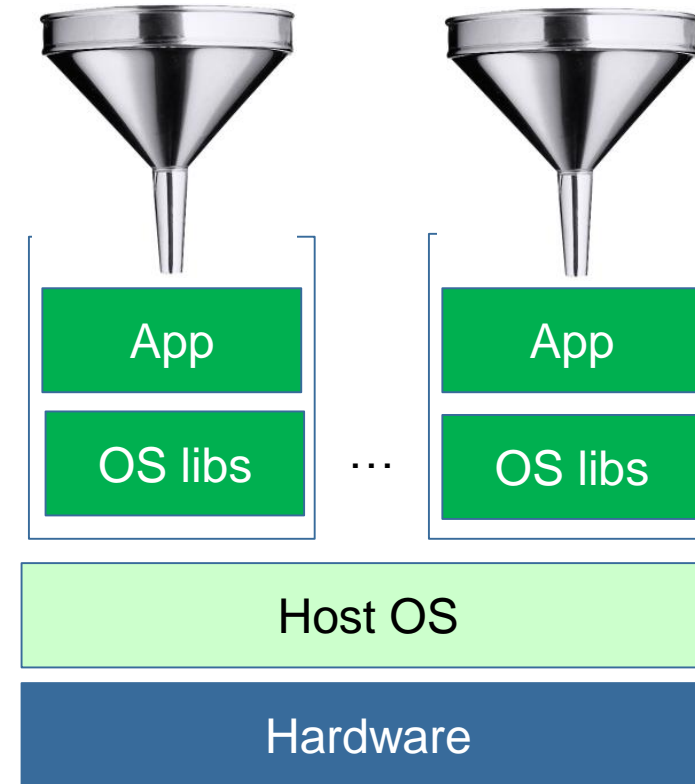
Provisionierung

Reactive Programming

Die letzte Vorlesung: Wie kommt Software auf das Blech?

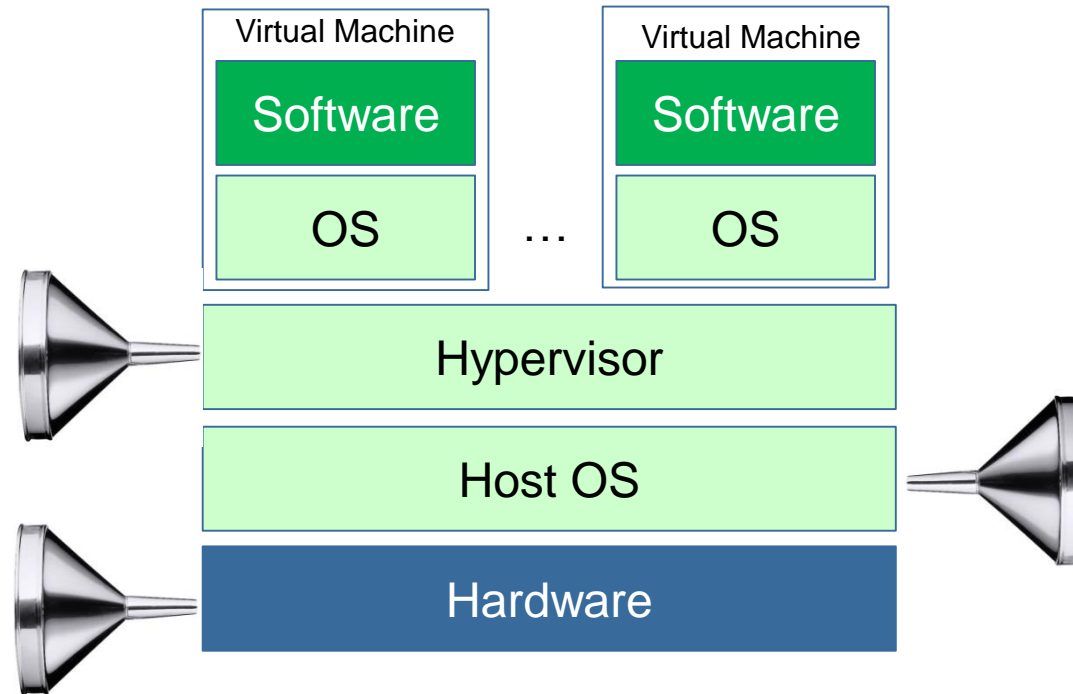


Hardware-Virtualisierung

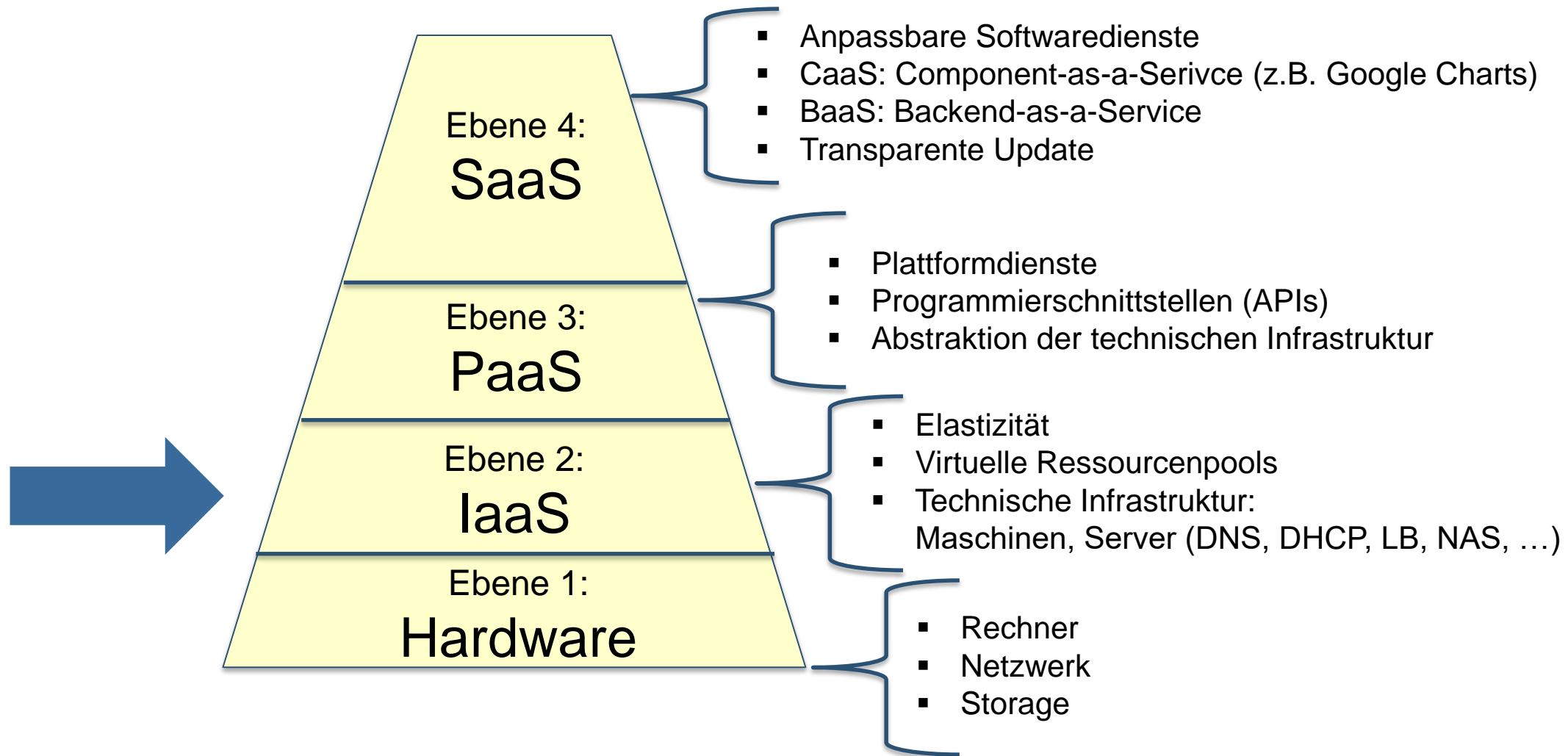


Betriebssystem-Virtualisierung

Heute: Wie kommt Software an das Blech?

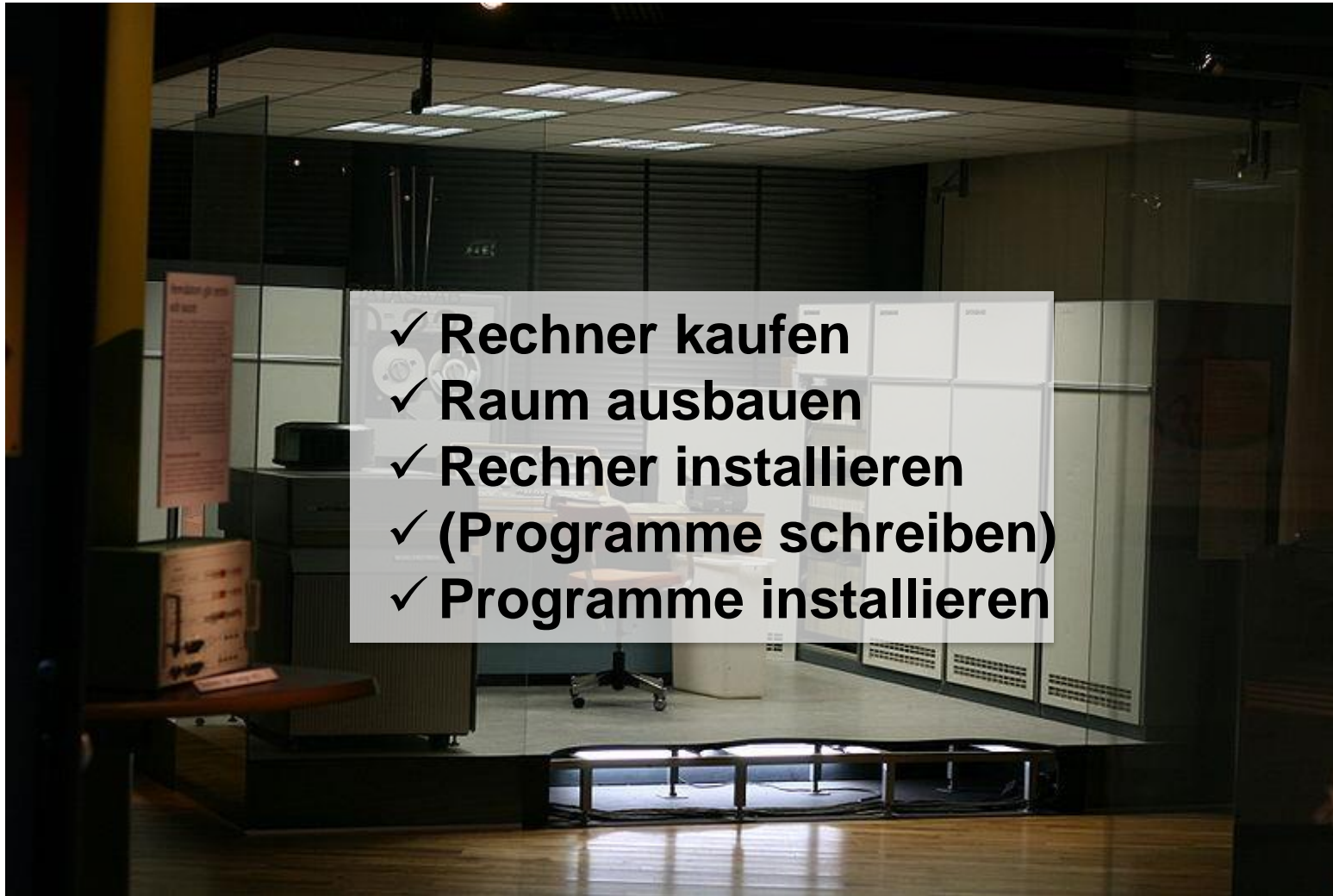


Das Schichtenmodell des Cloud Computing: Vom Blech zur Anwendung.



Einführung: Infrastructure-as-a-Service

Time2System im letzten Jahrhundert: > 1 Jahr.



- ✓ Rechner kaufen
- ✓ Raum ausbauen
- ✓ Rechner installieren
- ✓ (Programme schreiben)
- ✓ Programme installieren

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

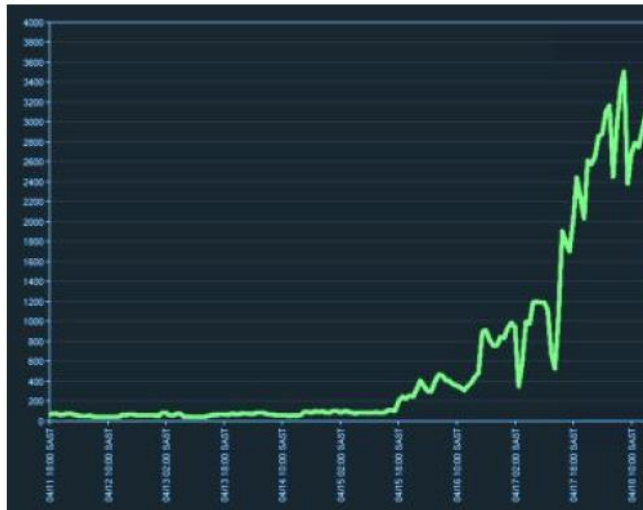
Time2System in der Cloud-Ära: In Echtzeit.

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*).

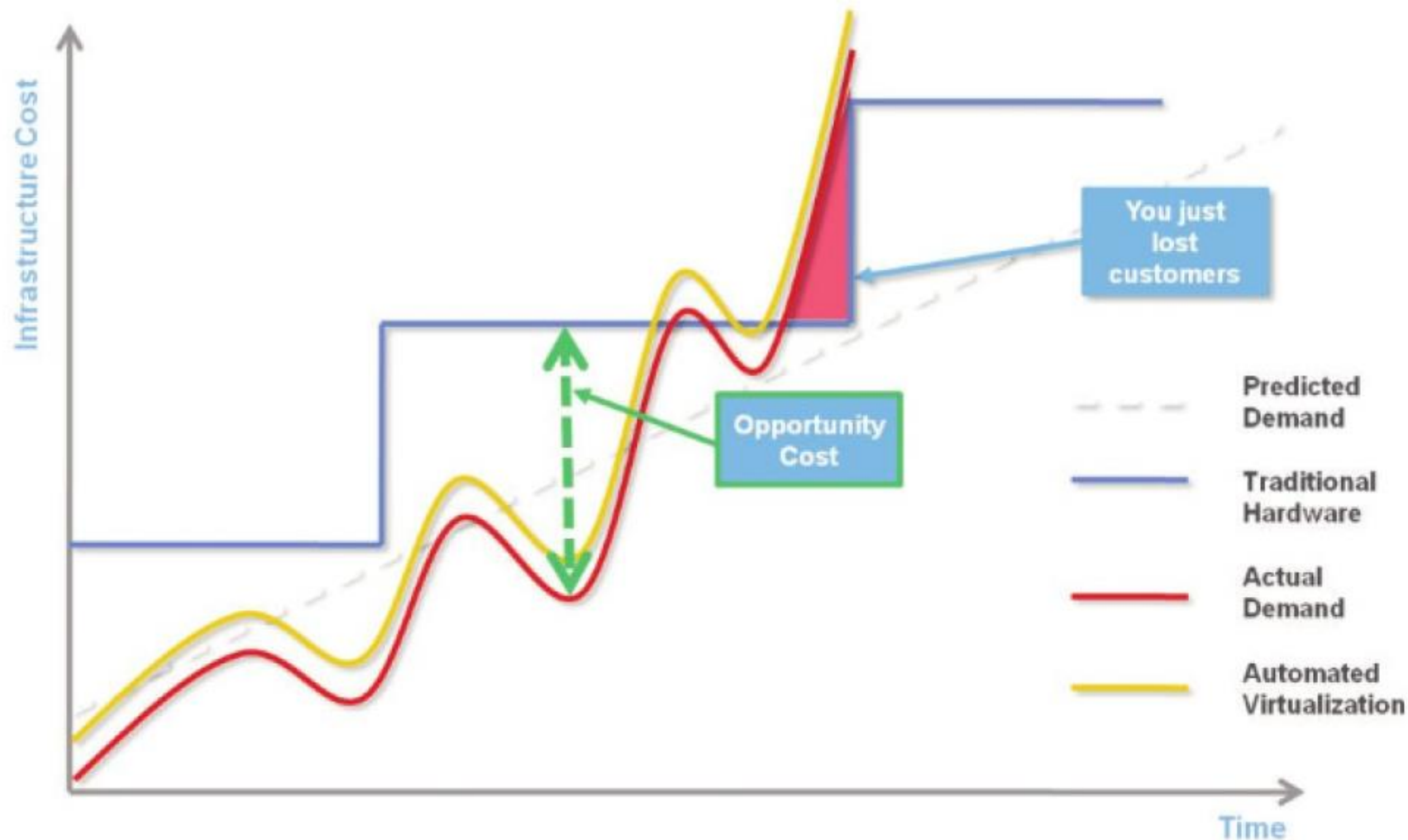
Große Websites, die von einer Server-Farm bedient werden, haben meistens keine Probleme mit dem erhöhten Traffic. Es sind vor allem kleinere Einzel-Server, die einem Slashdot-Effekt zum Opfer fallen. Manchmal wird der Slashdot-Effekt scherzhaft mit einem [Distributed-Denial-of-Service-Angriff](#) verglichen.

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](#).



The screenshot shows the Animoto website. At the top, there's a navigation bar with the Animoto logo, links for 'Sign In or Sign Up', 'Sign In with Facebook', and a 'Create Video' button. Below the navigation bar, there are links for 'Features', 'Sample Videos', and 'Plans & Pricing'. The main heading is 'Create stunning videos!'. Below this, a subheading reads: 'Turn your photos, video clips, and music into stunning video masterpieces to share with everyone. Fast, free, and shockingly easy!'. There are two buttons: 'Sign Up free' and 'Learn more'. On the right side, there's a video player showing a couple smiling, with a large play button icon and the text 'Play Sample Video' overlaid.

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.

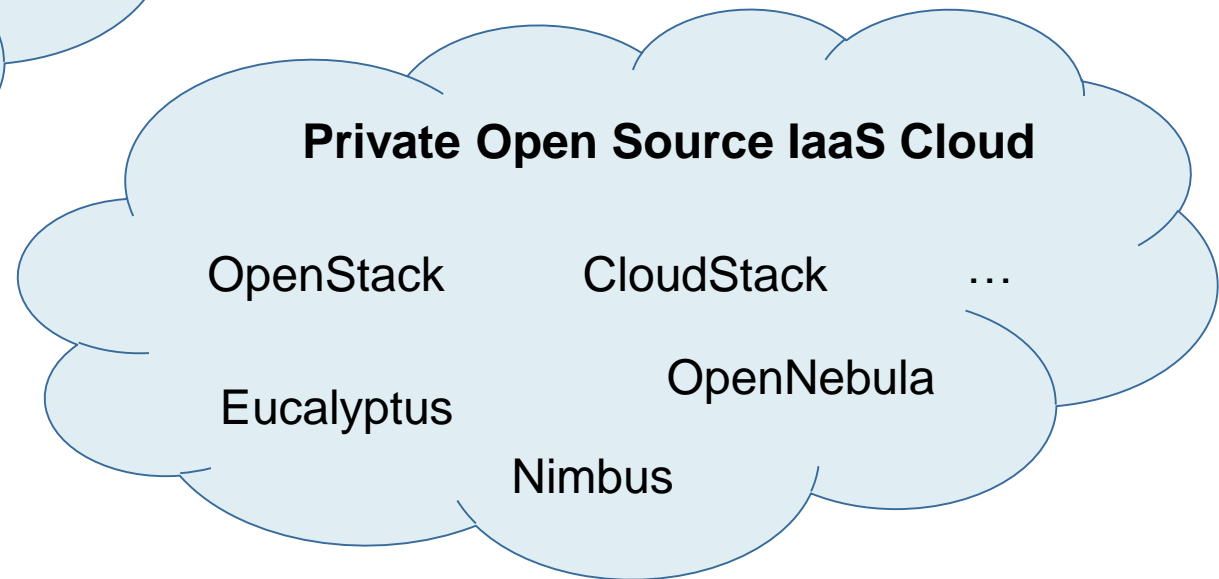
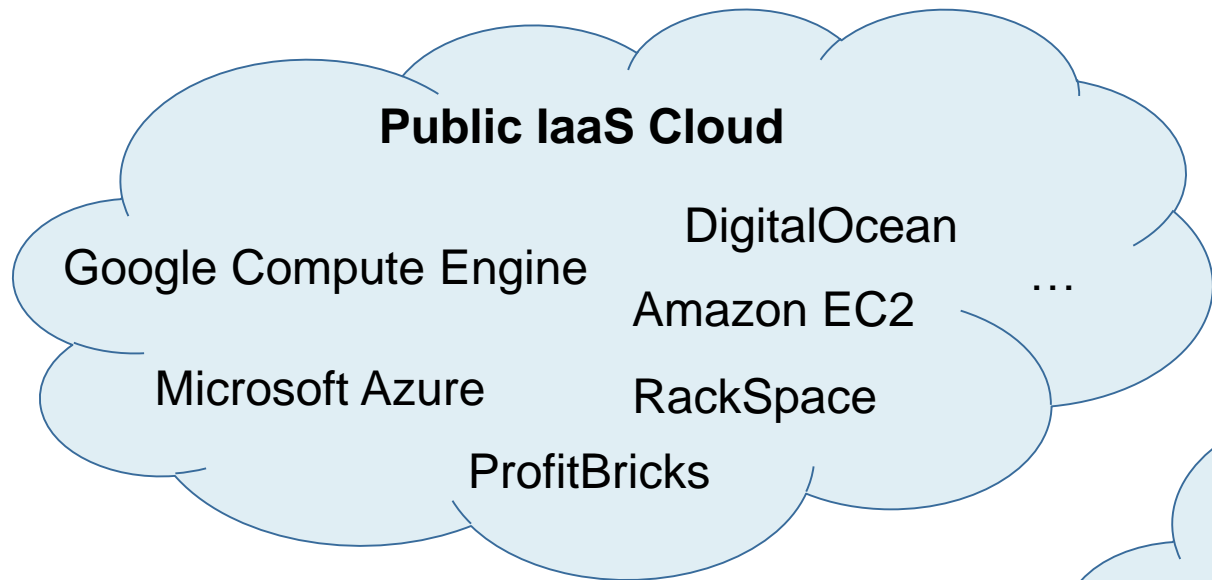
■ Ressourcen-Typen in einer IaaS-Cloud:

- **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.

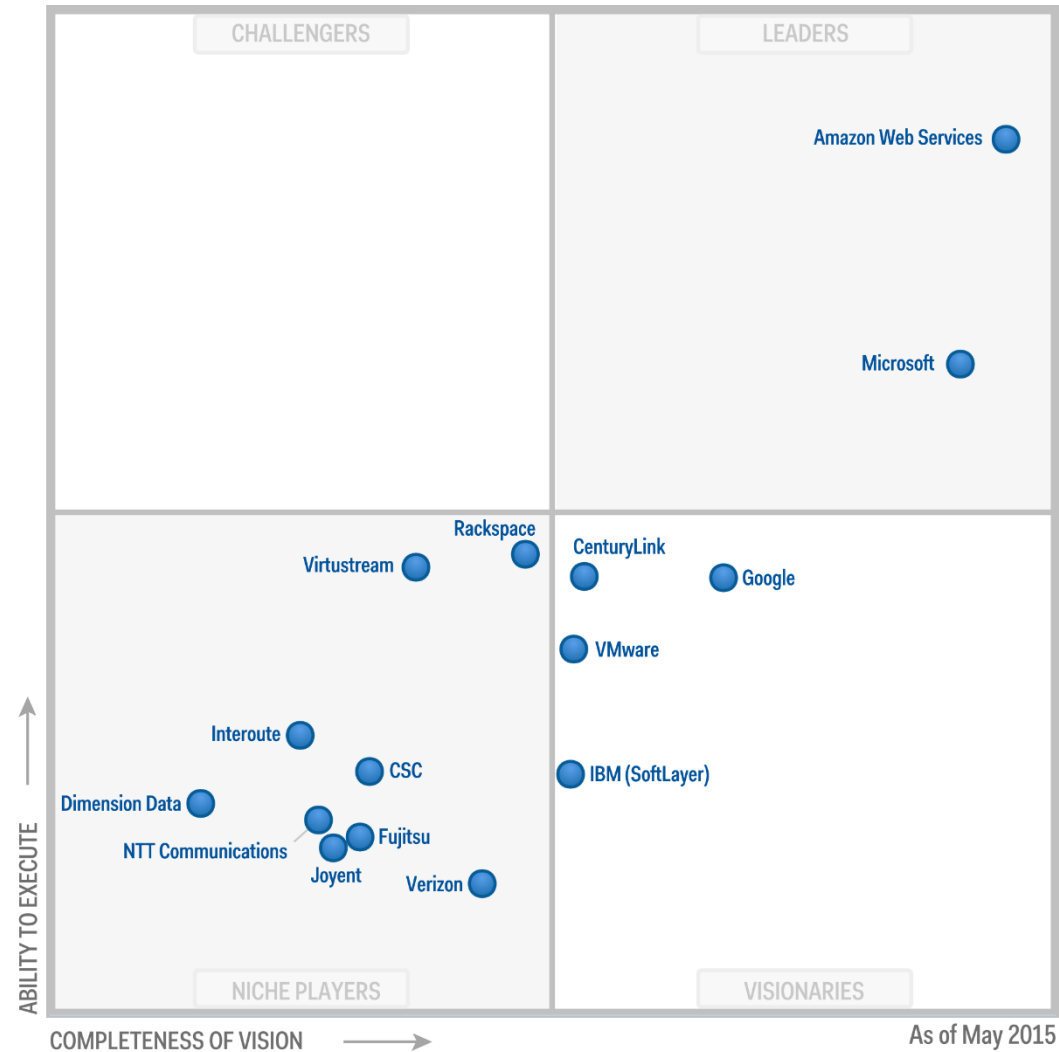
■ Infrastruktur-Dienste einer IaaS-Cloud:

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

**Der Markt expandiert gerade stark und ist wenig konsolidiert.
Das macht die Auswahl schwer.**



Der IaaS Markt 2015 laut Gartner



2015 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: Offene APIs
- Haftung
- Support

Characteristics of ...

Rubber Bands

■ Base Length



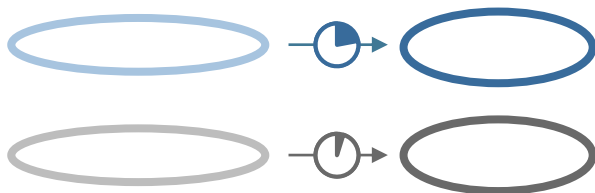
■ Width / Thickness / Force



■ Stretchability



■ Elasticity

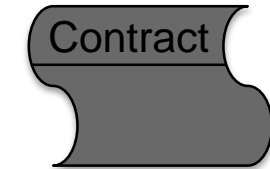
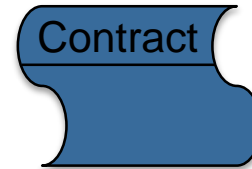


IaaS Clouds

■ Performance (1 resource unit)



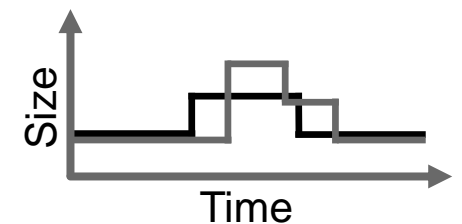
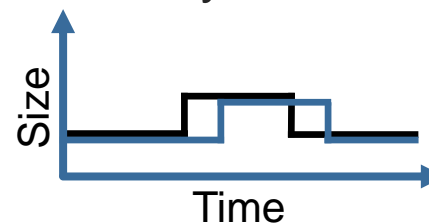
■ Quality Criteria / SLA



■ Scalability

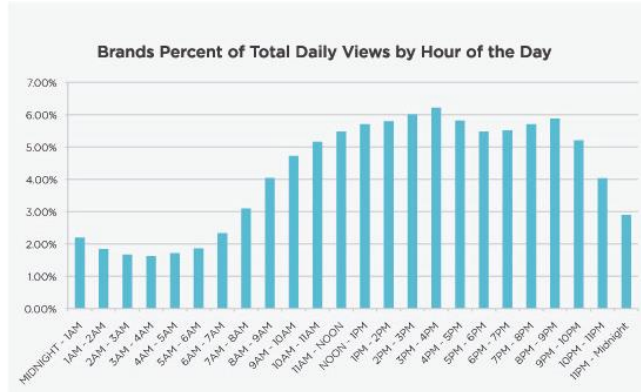


■ Elasticity

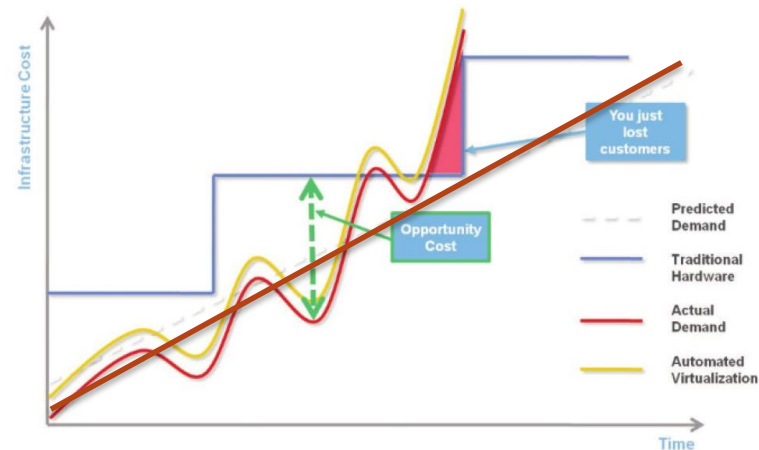


Skalierbarkeit: Effekte

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

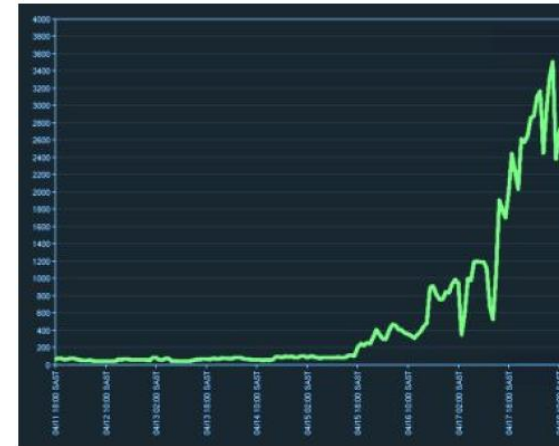


- **Kontinuierliches Wachstum**

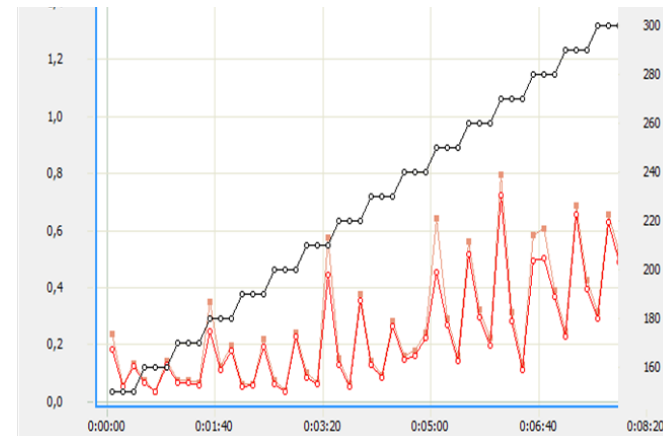


Source: Amazon Web Services

- **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.

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>

Preismodell: Die Kosten entstehen i.d.R. flexibel abhängig vom Ressourcen-Verbrauch.

■ Kosten =

Fixe Kosten

- + *Kosten pro Compute Hour x Anzahl der Compute Hours*
- + *Kosten pro gespeicherten Daten x Anzahl der gespeicherten Daten*
- + *Kosten pro eingehenden Datentransfer x Anzahl des eingehenden Datentransfers*
- + *Kosten pro ausgehenden Datentransfer x Anzahl des ausgehenden Datentransfers*
- + *Kosten pro Storage-Transaktion x Anzahl der Storage-Transaktionen*
- + *weitere flexible Kosten (Load Balancer, Monitoring, ...).*

■ Die IaaS-Kosten der verschiedenen Anbieter sind prinzipiell kalkulier- und vergleichbar.



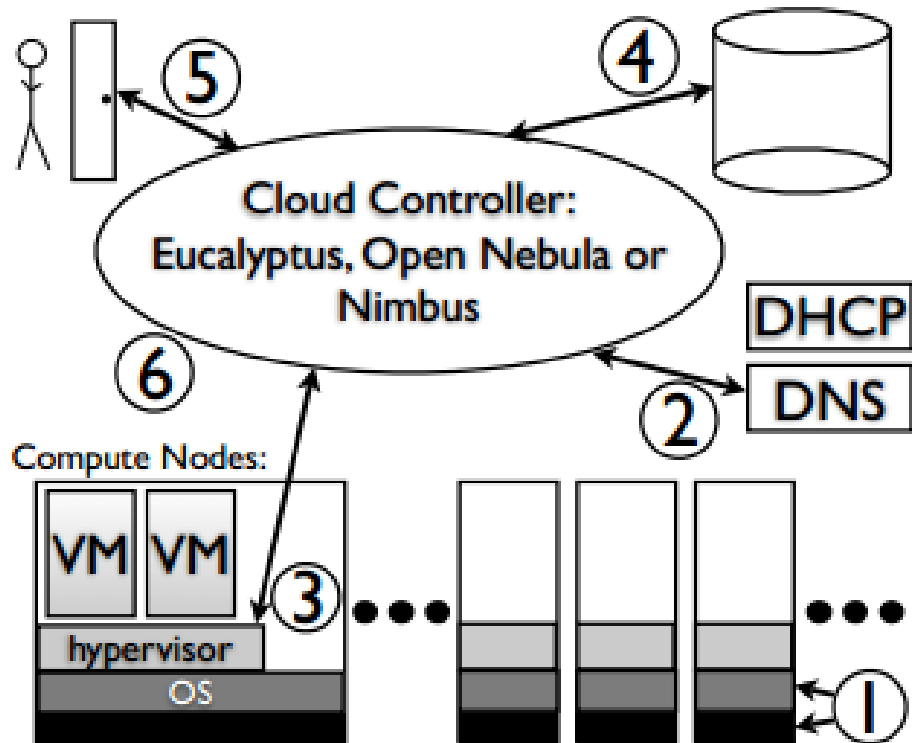
<http://calculator.s3.amazonaws.com/calc5.html>



<http://www.cloudorado.com>

Architektur einer IaaS Cloud

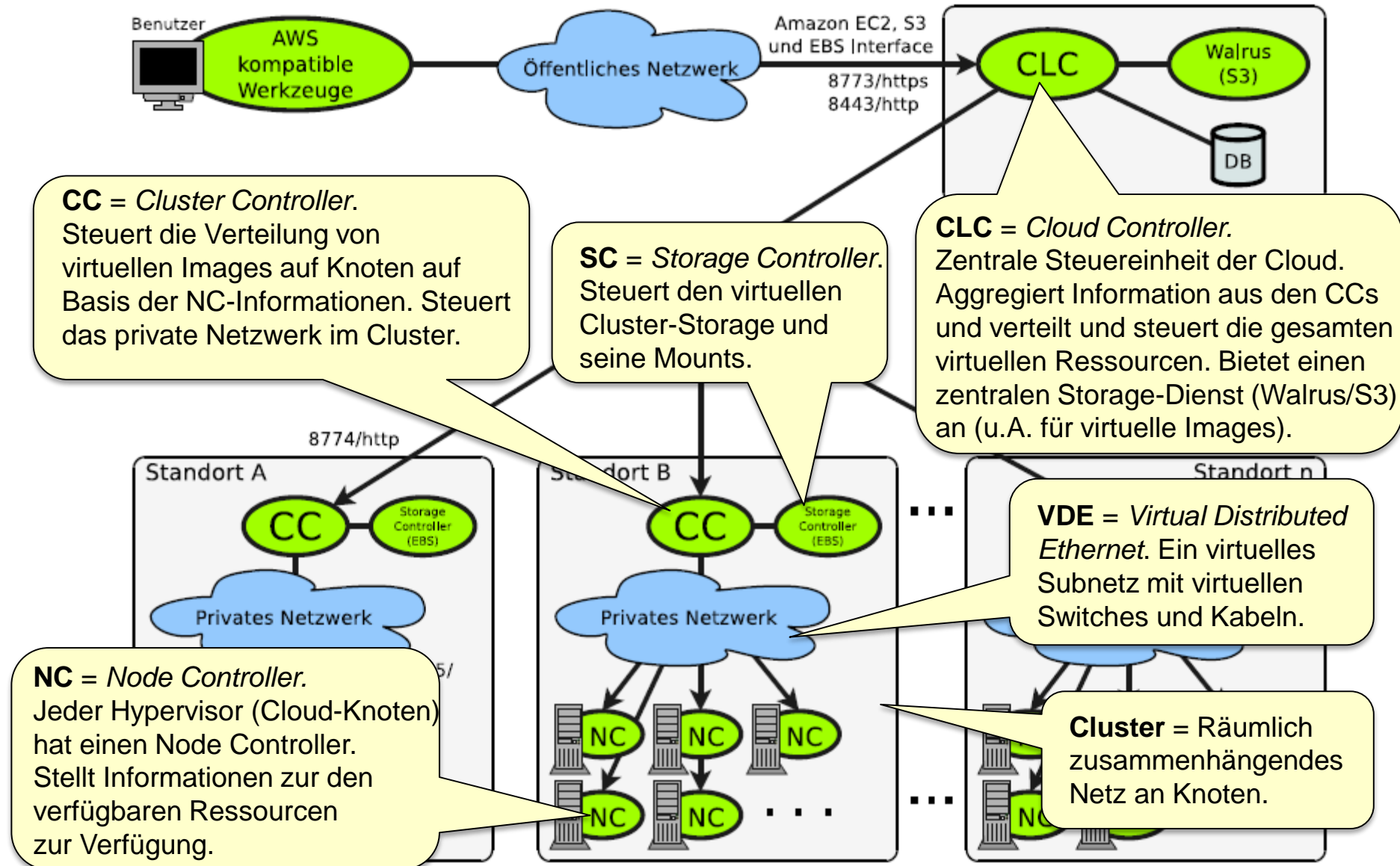
Eine IaaS-Referenzarchitektur



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

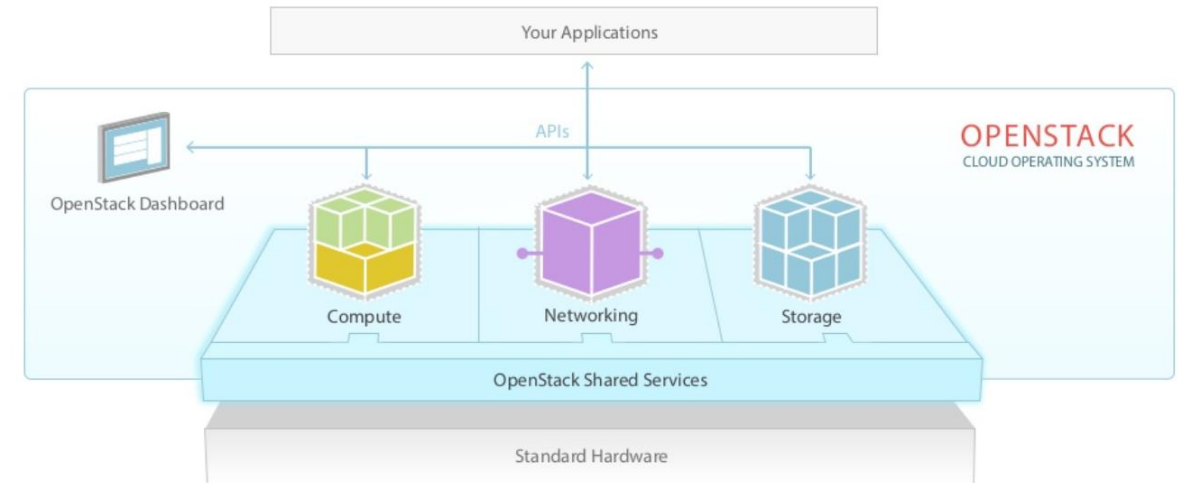
Peter Sempelinski and Douglas Thain,
“A Comparison and Critique of Eucalyptus, OpenNebula and Nimbus”,
IEEE International Conference on Cloud Computing Technology and Science, 2010.

Der interne Aufbau einer IaaS-Cloud am Beispiel Eucalyptus.



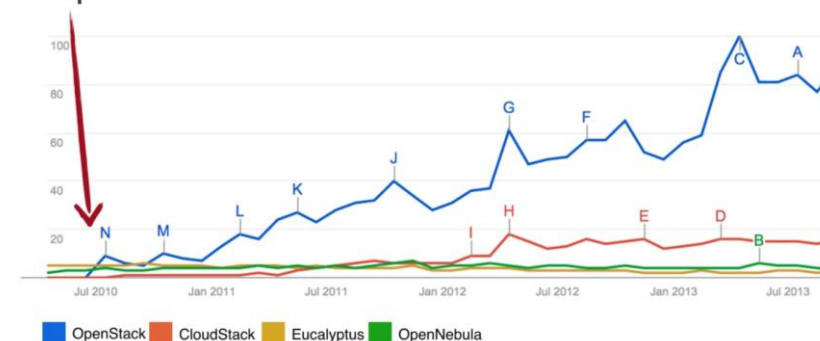
OpenStack: Der de-facto Standard für Open-Source Private IaaS Clouds.

- 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.



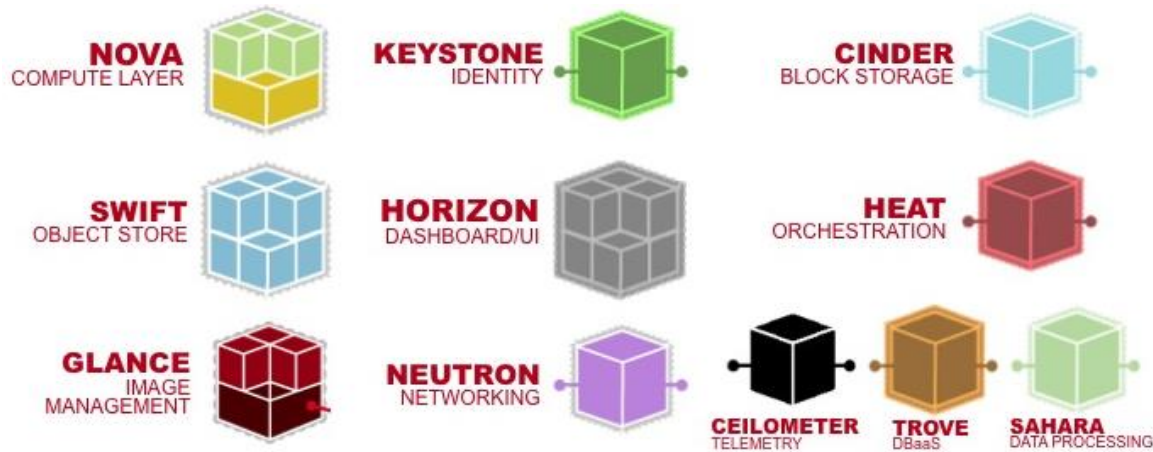
The Battle is Over (open src)

OpenStack Launch



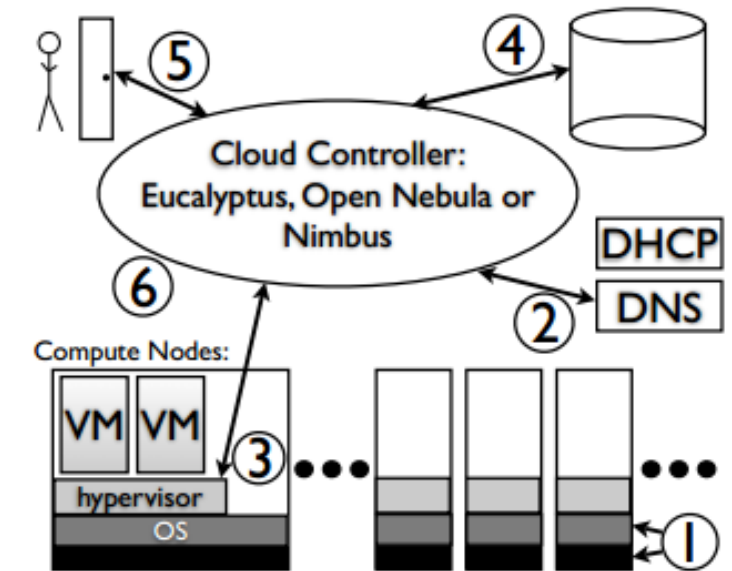
OpenStack Quiz!?

OpenStack Komponenten:



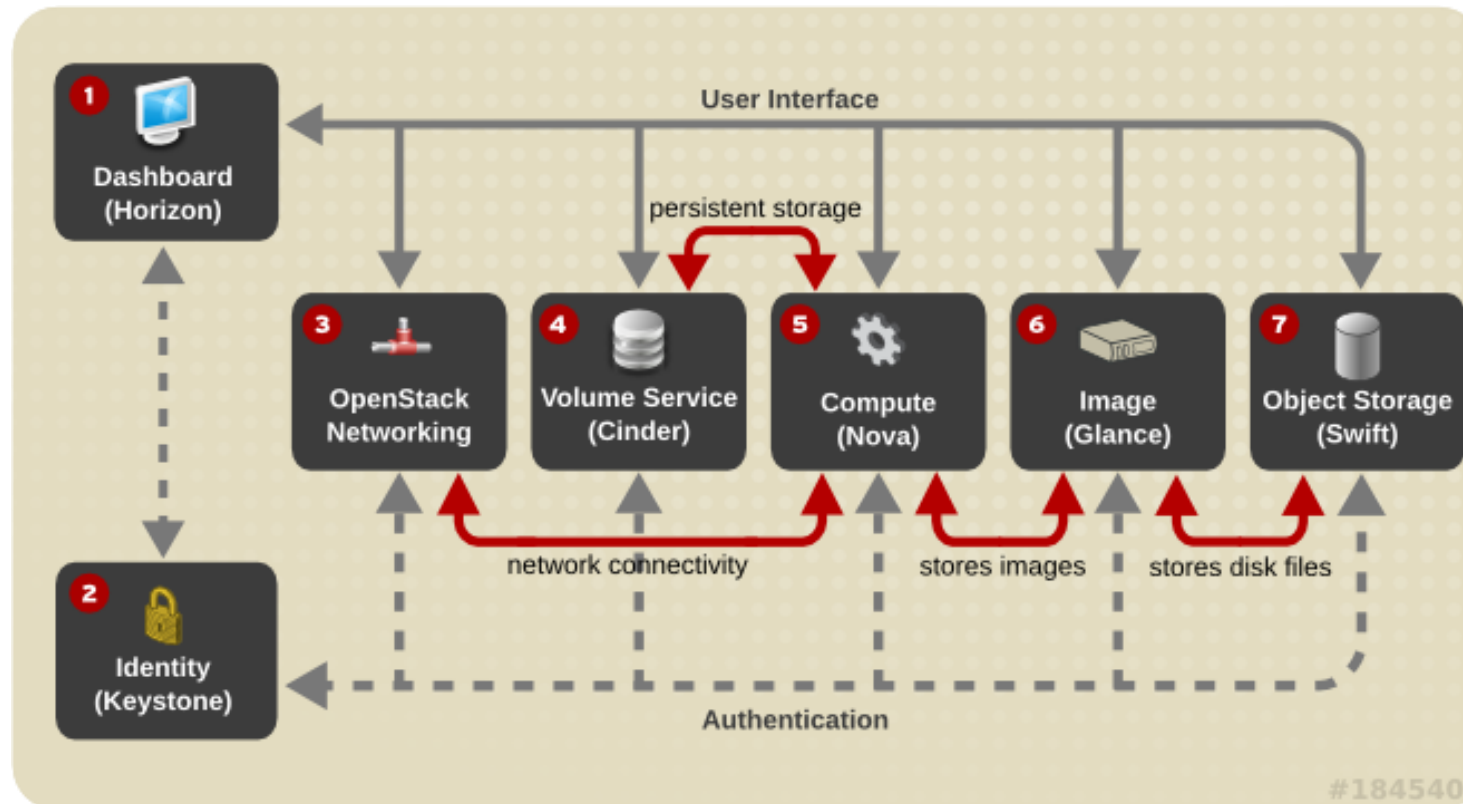
Quelle: <http://de.slideshare.net/sgordon2/deep-dive-openstack-summit-red-hat-summit-2014>

Komponenten der IaaS Referenz-Architektur:

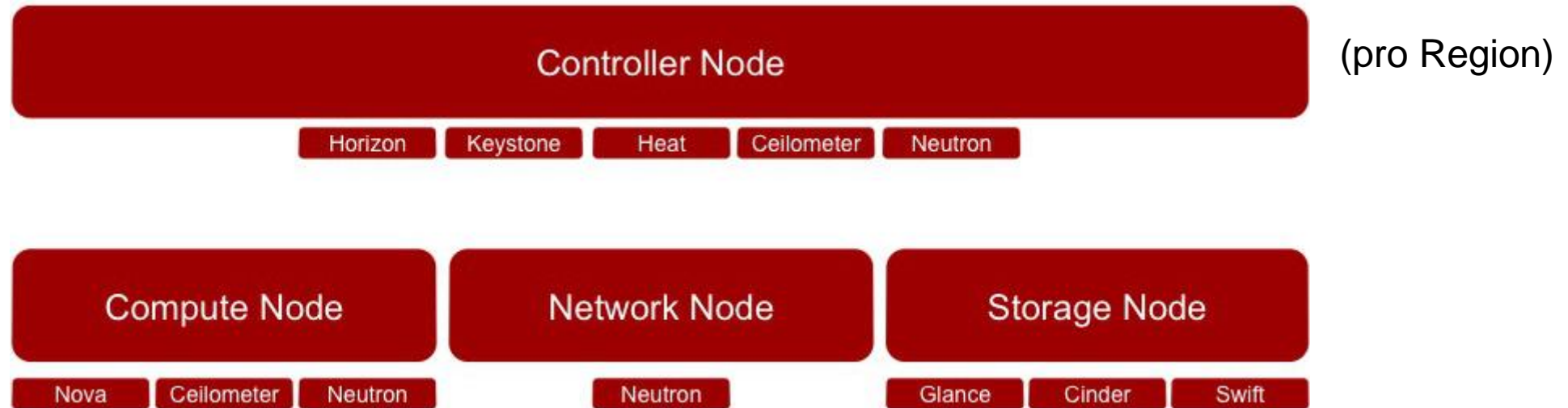


1. Hardware und Betriebssystem
2. Virtuelles Netzwerk und Netzwerkdienste
3. Virtualisierung
4. Datenspeicher und Image-Verwaltung
5. Managementschnittstelle
6. Cloud Controller

OpenStack: Das Zusammenspiel der Kern-Komponenten.



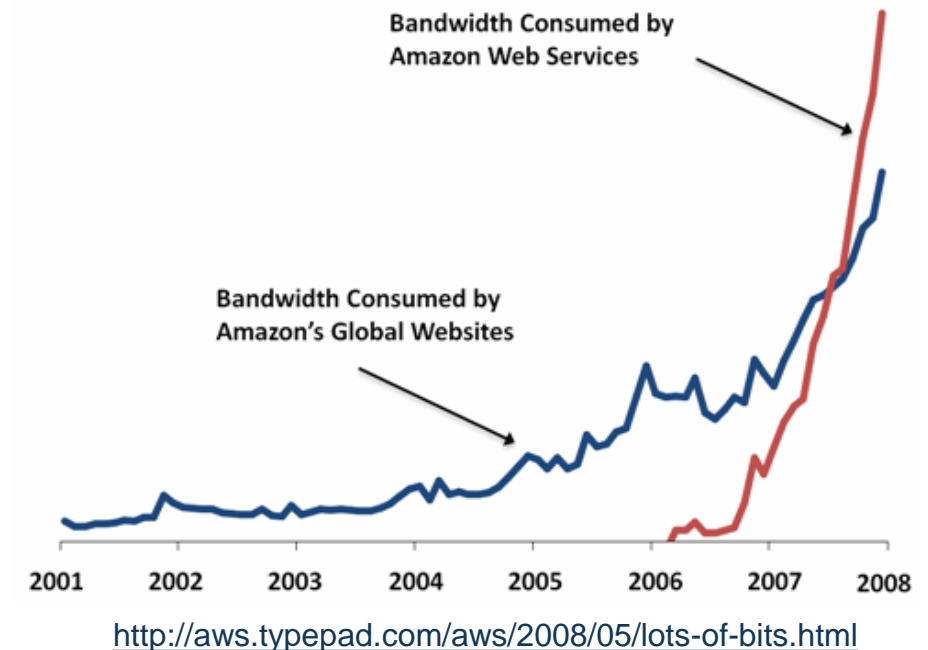
Die Verteilungssicht von OpenStack.








































IaaS mit der 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
 - 2014:
 - Marktführer mit ca. 27% Marktanteil, ca. 1 Mio. Kunden und ca. 4 Mrd. \$ Umsatz gefolgt von Microsoft Azure mit 10%.
 - 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.

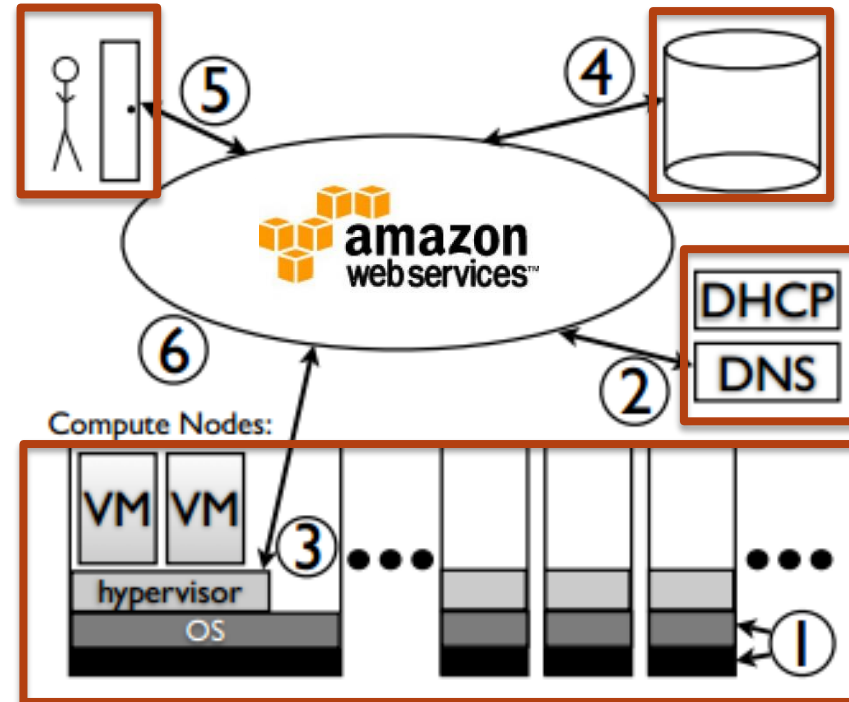


Neben der Amazon EC2 IaaS Cloud bietet Amazon noch viele weitere IaaS-Komponenten, PaaS- und SaaS-Dienste.

All AWS Services	>	 AppStream	 EC2	 Redshift
Compute		 CloudFormation	 ElastiCache	 Route 53
Storage & Content Delivery		 CloudFront	 Elastic Beanstalk	 S3
Database		 CloudSearch	 Elastic Transcoder	 SES
Networking		 CloudTrail	 EMR	 SNS
Administration & Security		 CloudWatch	 Glacier	 SQS
Analytics		 CodeDeploy	 IAM	 Storage Gateway
Application Services		 Cognito	 Kinesis	 SWF
Deployment & Management		 Config <small>PREVIEW</small>	 Lambda <small>PREVIEW</small>	 Trusted Advisor
Mobile Services		 Data Pipeline	 Mobile Analytics	 VPC
Enterprise Applications		 Direct Connect	 OpsWorks	 WorkSpaces
		 Directory Service	 RDS	 Zocalo
		 DynamoDB		

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.

Virginia Datacenter



US East (N. Virginia)

US West (Oregon)

US West (N. California)

EU (Ireland)

EU (Frankfurt)

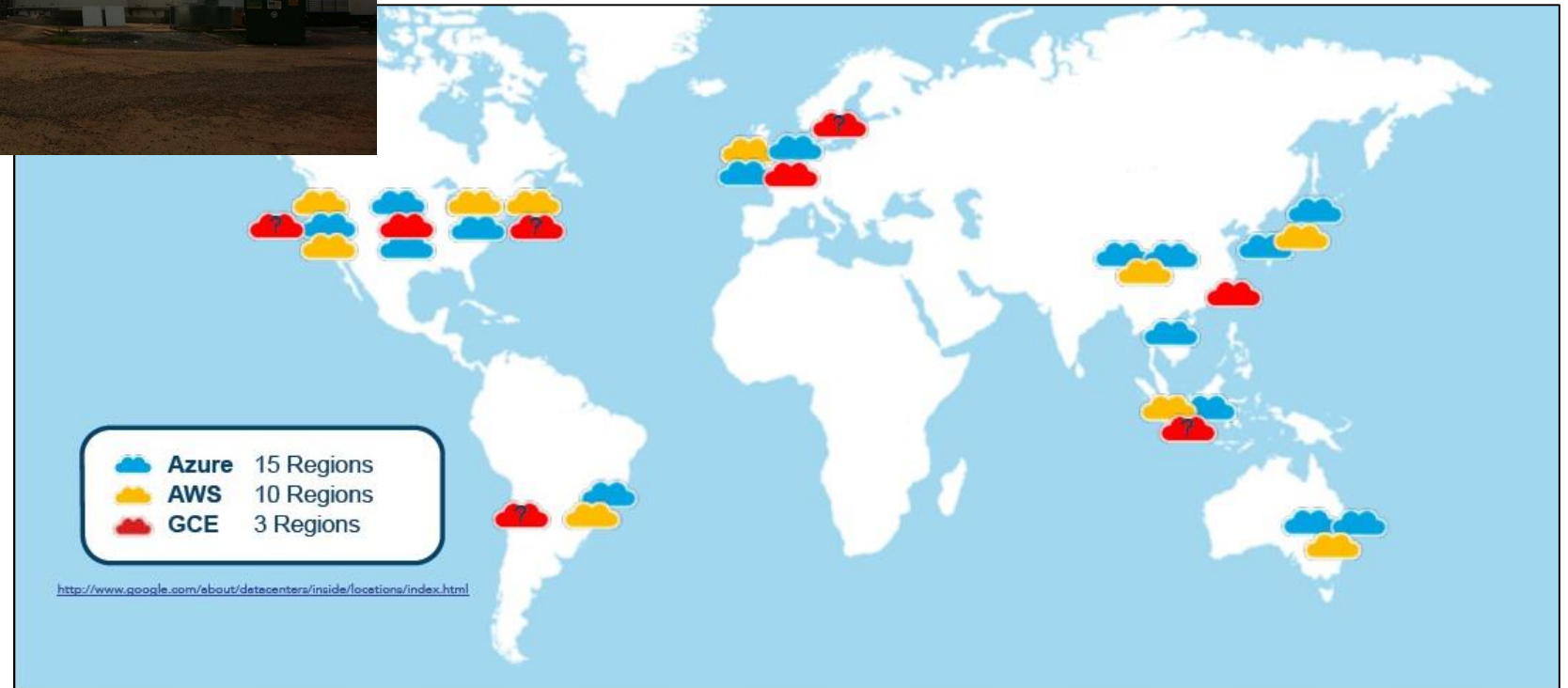
Asia Pacific (Singapore)

Asia Pacific (Tokyo)

Asia Pacific (Sydney)

South America (São Paulo)

+ GovCloud



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. Authentifizierung 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.

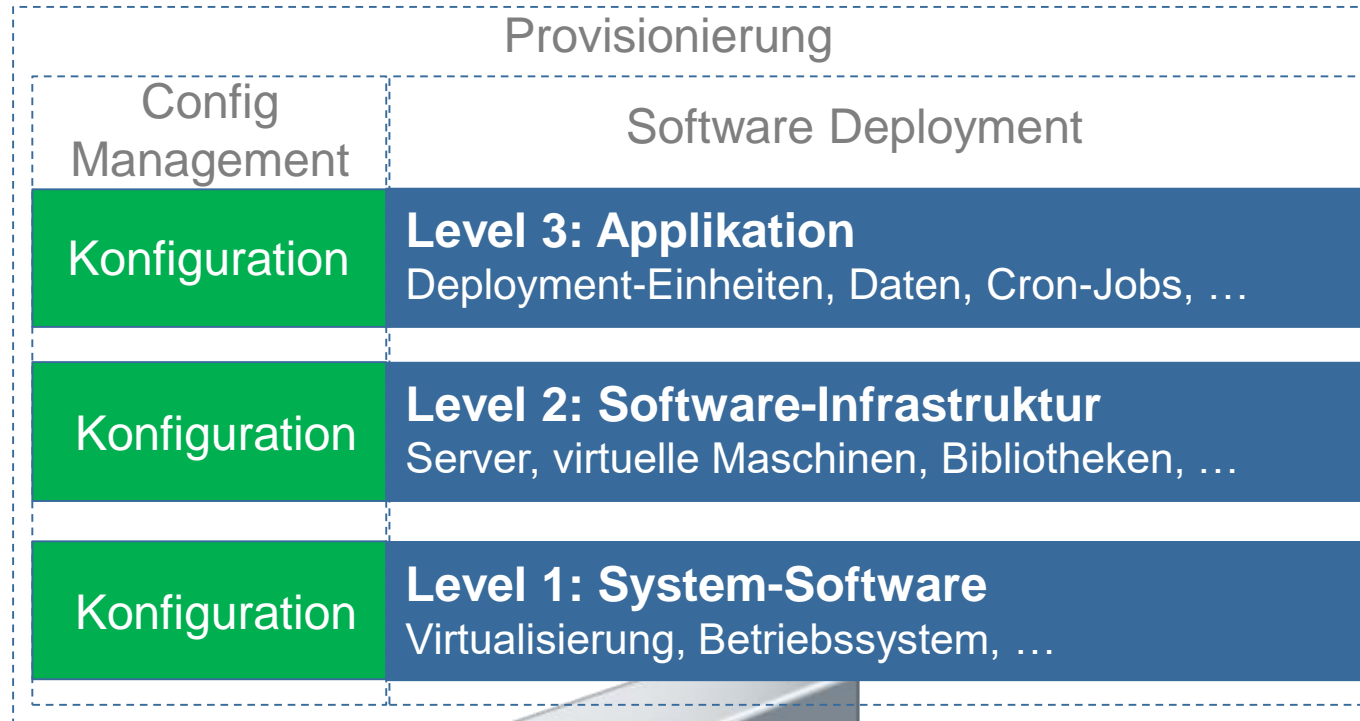
Über die AWS Management Console können alle Dienste der Amazon-Cloud gesteuert werden.

The screenshot displays the AWS Management Console interface for the EC2 service. The top navigation bar shows the user 'Josef Adersberger' and the region 'Frankfurt'. The left-hand navigation menu lists various EC2-related services, including EC2 Dashboard, INSTANCES, IMAGES, ELASTIC BLOCK STORE, NETWORK & SECURITY, and AUTO SCALING. The main content area is divided into several sections:

- Resources:** A section titled 'Resources' showing the number of EC2 resources in the EU (Frankfurt) region. It lists: 0 Running Instances, 0 Elastic IPs, 0 Volumes, 0 Snapshots, 0 Key Pairs, 0 Load Balancers, 0 Placement Groups, and 1 Security Group.
- Create Instance:** A section titled 'Create Instance' with a 'Launch Instance' button. It includes a note: 'Note: Your instances will launch in the EU (Frankfurt) region'.
- Service Health:** A section titled 'Service Health' showing the status of the EC2 service in the EU (Frankfurt) region. It indicates that the service is operating normally in the eu-central-1a and eu-central-1b availability zones.
- Scheduled Events:** A section titled 'Scheduled Events' showing that there are no events scheduled for the EU (Frankfurt) region.
- Account Attributes:** A section titled 'Account Attributes' showing account information, including 'Supported Platforms' (VPC) and 'Default VPC' (vpc-2ef31c47).

Provisionierung einer IaaS Cloud

Provisionierung erfolgt auf drei verschiedenen Ebenen und in vier Stufen.



Hardware

- Rechner
- Speicher
- Netzwerk-Equipment
- ...

Laufende Software!



Application Provisioning



Server Provisioning

Bereitstellung der Laufzeitumgebung für die Applikation.



Bootstrapping

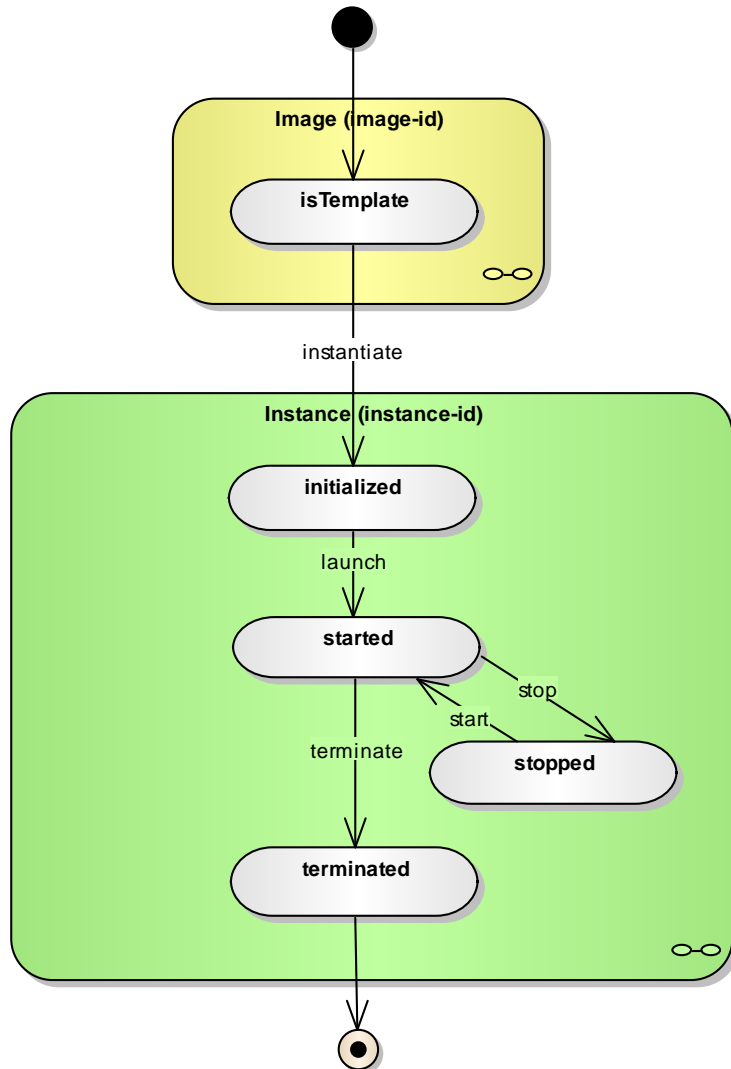
Bereitstellung der Betriebsumgebung für die Software-Infrastruktur.



Bare Metal Provisioning

Initialisierung einer physikalischen Hardware für den Betrieb.

Bootstrapping: Images und Instances



- Ein Image ist ein virtuelles Image für die Virtualisierungstechnik der Cloud.
- Ablage der Images i.d.R. auf zentralem Storage-Dienst.
- Verteilung von Images über Dateitransfer-Protokolle (z.B. BitTorrent)
- *terminated* bedeutet, dass die Instanz verworfen wird inkl. aller lokaler Daten und Daten auf dem Cluster-Storage.
- *stopped* bedeutet, dass die Instanz verworfen wird inkl. aller lokaler Daten, nicht jedoch die Daten auf dem Cluster-Storage gelöscht werden.
- Ablage von gestoppten Instanzen i.d.R. auf Cluster-Storage.