Cloud Computing

Nelson L. S. da Fonseca

IEEE ComSoc Summer Scool

Albuquerque, July 17-21, 2017

Acknowledgement

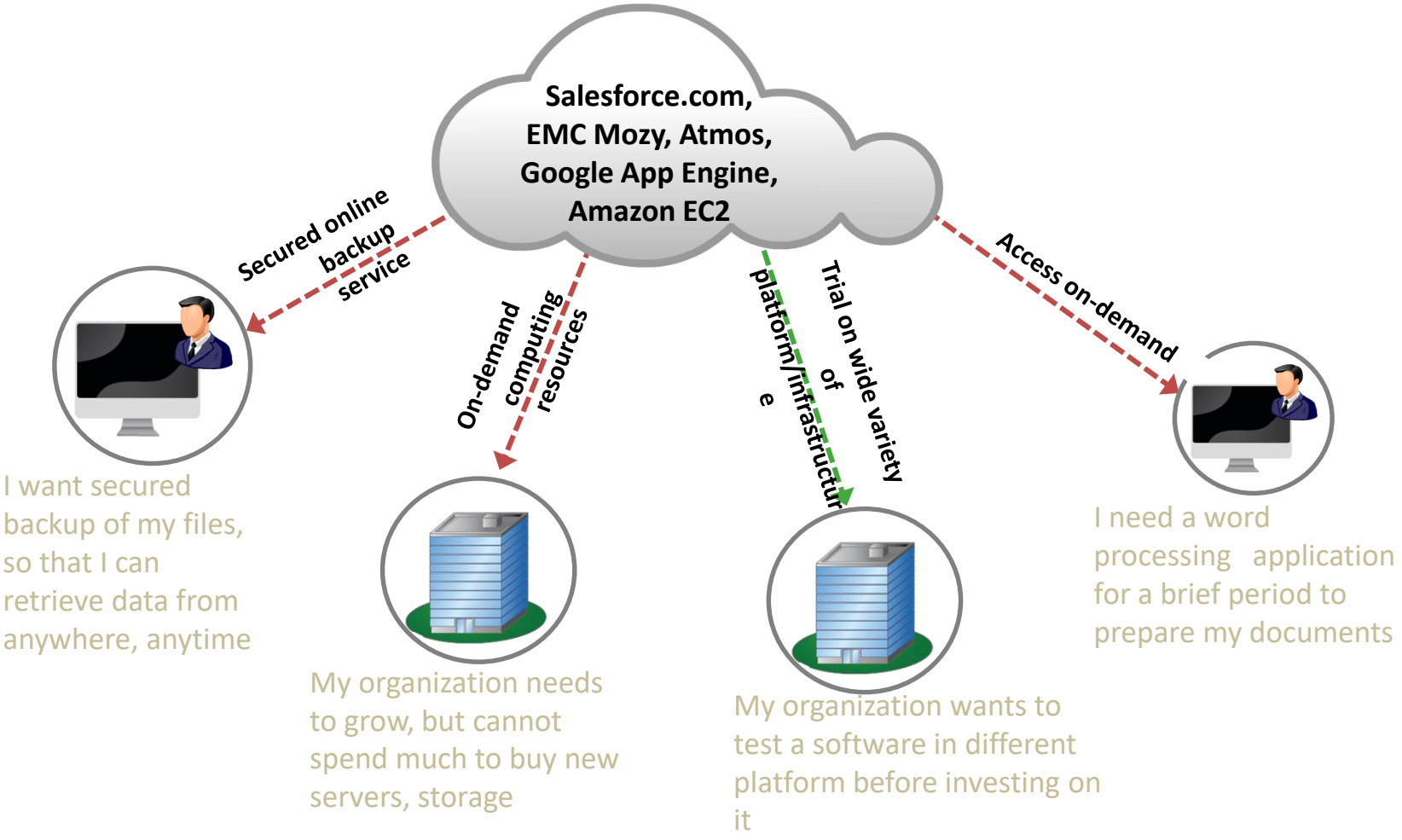
- Some slides in this set of slides were kindly provided by:
  - Dzmitry Kliazovich, University of Luxenbourg
  - Luiz Fernando Bittencourt, University of Campinas
  - EMC Corporation

Cloud Computing

"Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."

NIST 2009

Cloud Services

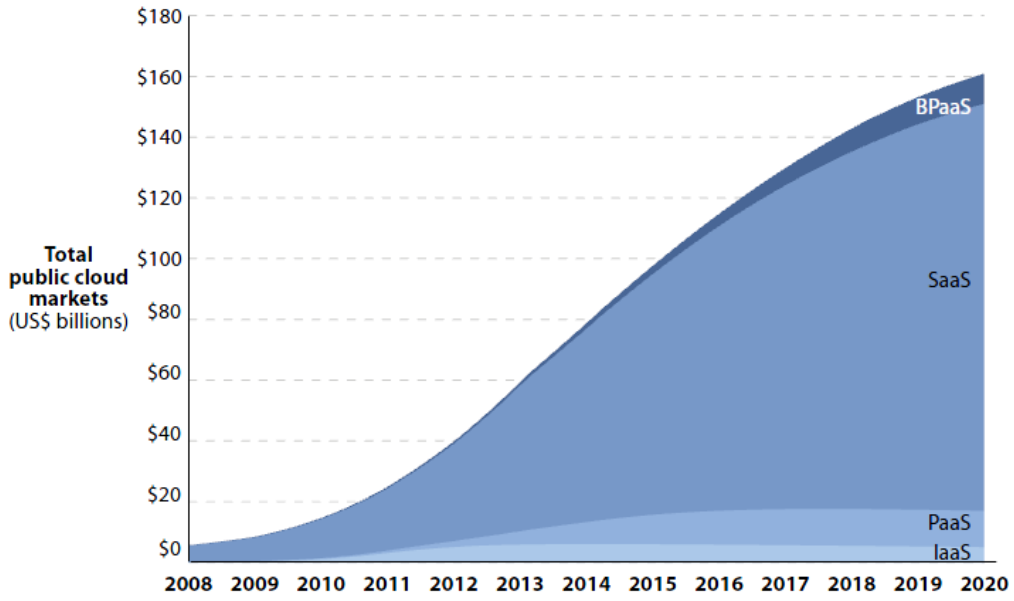


Cloud Players

- Flickr has more than 6 billion photos
- Google serves more than 1.2 billion queries/day on more than 27 billion items
- More than 2 billion videos watched a day on YouTube

Cloud Computing

- Cloud computing market: $241 billion in 2020
- Main focus is on Software-as-a-Service (SaaS)



Source: Larry Dignan, "Cloud computing market", ZDNet, 2011.

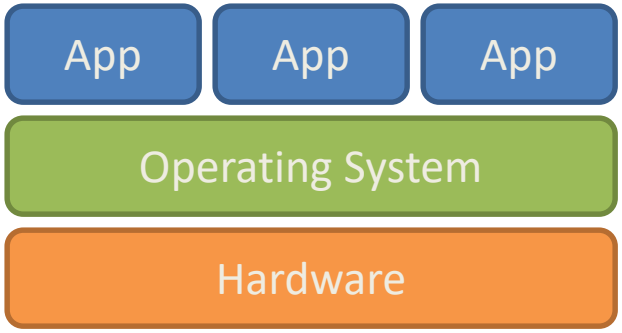
Related Technologies

Technology	Charactetistic
Grid Computing	Form of distributed computing which applies the resources of numerous computers in a network to work on a single complex task at the same time
Utility Computing	Service provisioning model that offers computing resources as a metered service
Virtualization	Provides improved utilization of resources Enables optimization of resources by over subscription
Service Oriented Architecture (SOA)	An architectural approach in which a applications make use of services available in the network Each service provides a specific function, for example, business function (Payroll Tax calculation)

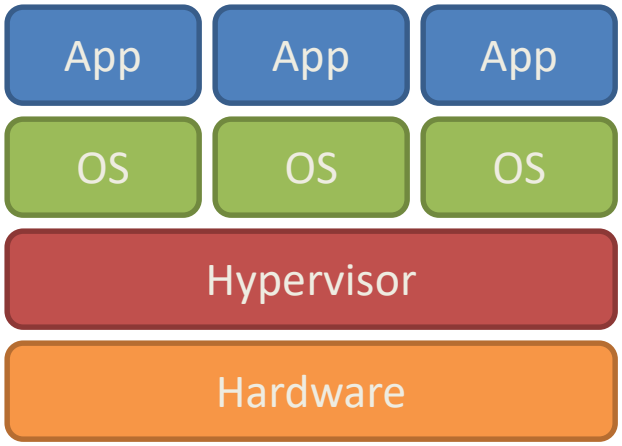
Grids and Clouds

GRID	Clouds
Uniform distribution of resources	Resources allocated on demand
Scientific focus	Commercial focus
Batch programming	Service based
(*Grid Security Infrastructure*)	Not specific security model

Virtualization



Traditional Stack



Virtualized Stack

Envision of Computing as Utility

- "Computing may someday be organized as a public utility, just as the electricity is organized as a public utility"

John McCarthy, speech at MIT in 1961

- "A of now, computer networks are still in their infancy, but as they grow up and become sophisticated, we will probably see the spread of computer utilities"

Leonard Kleinrock "Father" of the , Internet

Cloud Computing Services

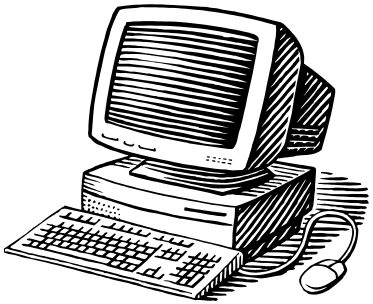
Cloud Computing Services

- Resources provided as services
- Pay-per-use business model
- Elasticity
- Service guarantees
- Ubiquitous access to resources

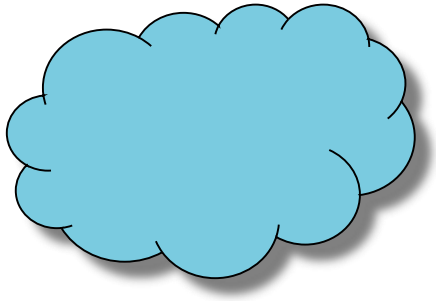
Resources provided as services

- User can request, configure and access cloud resources using cloud-specific APIs
- Enables consumers to get computing resources as and when required, without any human intervention
- Facilitates consumer to leverage "ready to use" services or, enables to choose required services from the service catalog

Resources provided as services



Private computational resources



Resources provided by the Cloud

Pay-per-use business model

- Consumers only pay for resources actually used
- Resource usage is monitored and reported, which provides transparency for chargeback to both Cloud service provider and consumer about the utilized service
- Pricing/billing model is tied up with the required service levels

Pay-per-use business model

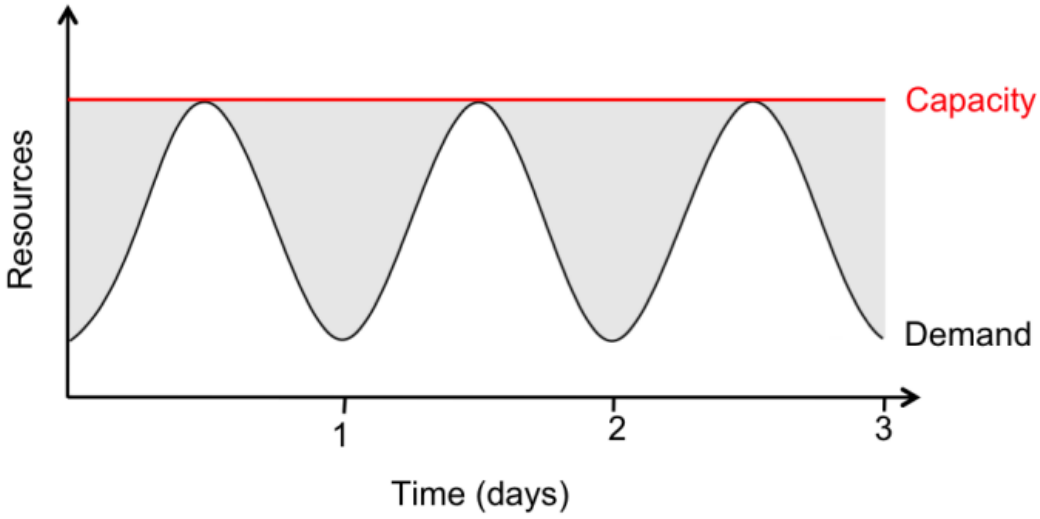
API name	Memory (GiB)	Cores/Compute units	Instance storage (GB)	32bit/64bit	I/O performance	EBS-optimizable	On-demand cost (Linux: per hour on US East 1)
t1.micro	0.594	Variable/Variable	0 (EBS only)	32/64	Variable	No	$0.020
t2.micro	1	1/Variable	0 (EBS only)	32/64	Low	No	$0.013
t2.small	2	1/Variable	0 (EBS only)	32/64	Low	No	$0.026
t2.medium	4	2/Variable	0 (EBS only)	32/64	Low	No	$0.052
m1.small	1.7	1/1	160	32/64	Moderate	No	$0.060
m1.medium	3.75	1/2	410	32/64	Moderate	No	$0.120
m1.large	7.5	2/4	850	64	Moderate	500 Mbit/s	$0.240
m1.xlarge	15	4/8	1600	64	High	1000 Mbit/s	$0.480
m2.xlarge	17.1	2/6.5	420	64	Moderate	No	$0.410
m2.2xlarge	34.2	4/13	850	64	High	500 Mbit/s	$0.820
m2.4xlarge	68.4	8/26	1690	64	High	1000 Mbit/s	$1.640
m3.xlarge	15	4/13	0 (EBS only)	64	Moderate	500 Mbit/s	$0.500
m3.2xlarge	30	8/26	0 (EBS only)	64	High	1000 Mbit/s	$1.000
c1.medium	1.7	2/5	350	32/64	Moderate	No	$0.145
c1.xlarge	7	8/20	1690	64	High	1000 Mbit/s	$0.580
c3.8xlarge	60	32 (Xeon E5-2680 v2)		64			
c4.8xlarge	60	36 (Xeon E5-2666 v3)		64			

Elasticity

- Consumers can acquire or release resources on demand
- Ability to scale IT resources rapidly, as required, to fulfill the changing needs without interruption of service
- Resources can be both scaled up and scaled down dynamically
- To the consumer, the Cloud appears to be infinite
- Consumers can start with minimal computing power and can expand their environment to any size

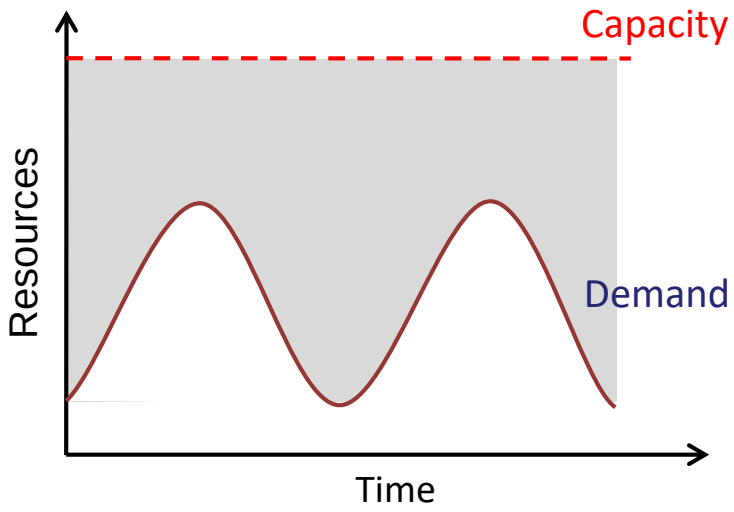
Elasticity

- Pay by use instead of provisioning for peak
  - Recall: DC costs >$150M and takes 24+ months to design and build



Elasticity

- Risk of over-provisioning: underutilization
  - Huge sunk cost in infrastructure

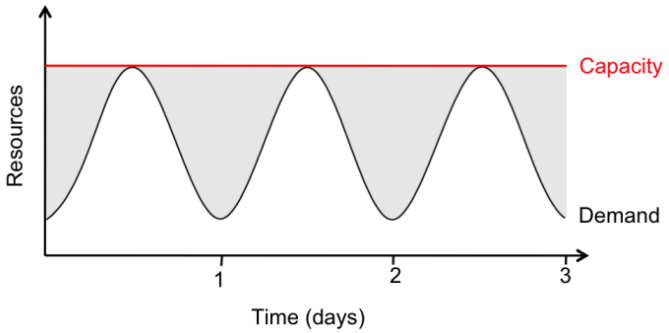


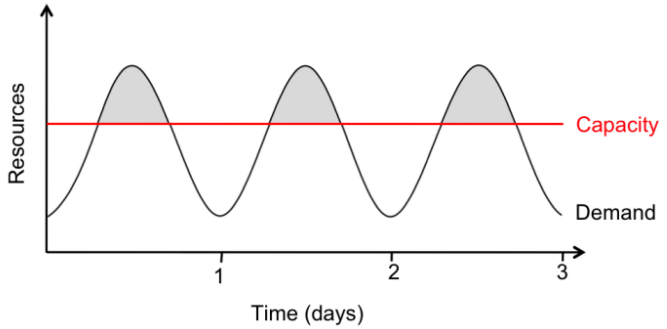
Static data center

Unused resources

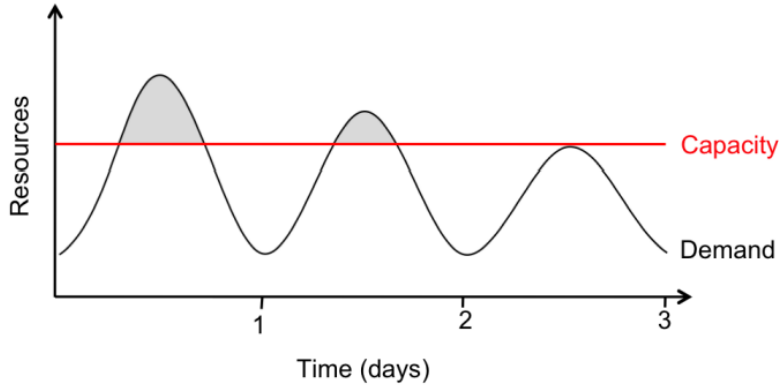
Elasticity

- Heavy penalty for under-provisioning





Lost revenue



Lost users

- Founded in 2006, launched app in 2008
- Allows users to create video with sound out of photos
- Elasticity: in 3 days, 7500,00 users signed for the app –demand went from 50 to 3,500 machines (Amazon EC2)

http://animoto.com/blog/company/amazon-com-ceo-jeff-bezos-on-animoto

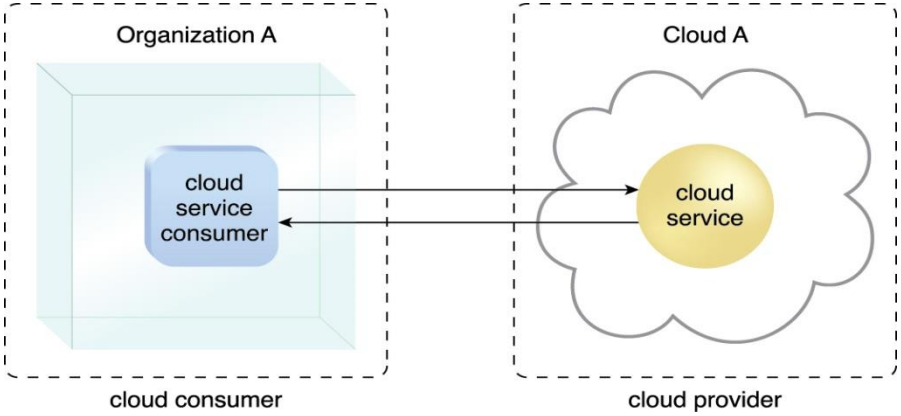
The Washington Post

- In, 2008, the White House schedule released to the public 17,481 pages of non-searchable, low-quality files in PDF;
- Content of great interest to journalists, The Washigton post employed optical character recognition (OCR) programto make content searchable;
- Used 1,479 hours of VM time (200 EC2 instance) to make content available on the web in 26 hours after release of low quality pdf files;

http://aws.amazon.com/solutions/case-studies/washington-post/

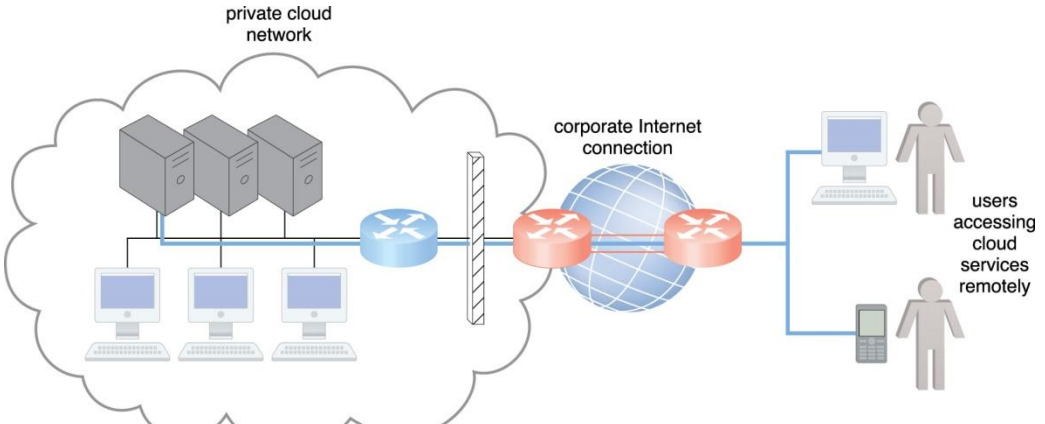
Service guarantees

- Cloud providers can offer guarantees on the use of certain services (e.g., 99.9% availability) in the form of service-level agreements (SLAs)



Ubiquitous access to resources

- Consumers can access resources using standard web protocols (REST or SOAP) from a variety of clients (web browsers, PDAs, cell phones)
- Eliminates the need for accessing a particular client platform to access the services
- Enables accessing the services from anywhere across the globe



Cloud Computing Benefit

Data ownership

Security & Privacy

Legal competence

Backup services

Location independence

Elastic performance

Elastic pricing

Reduced costs

Quality of Service

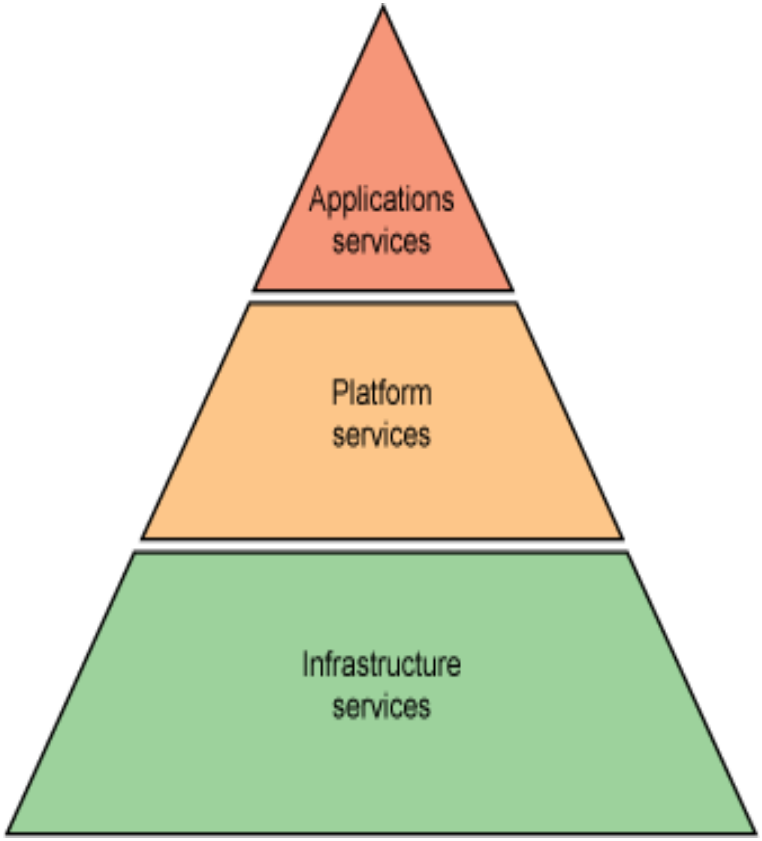
Flexibility and speed to market

Advantages Cloud Computing

Reduced IT Cost	Prevents the up-front capital expenditure
Business agility support	ability to add new resources quickly
Flexible scaling	Scales easily and instantly and on demand
High availability	Ensures application availability at varying levels
Less energy consumption	reduced power consumption

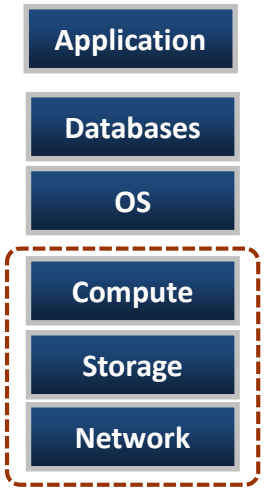
As a Service

- SaaS – *Software as a Service*
- PaaS – *Platform as a Service*
- – *Infrastructure as a Service*
  Infrastructure as a Service



Infrastructure-as-a-Service

- Provides capability to the consumer to hire infrastructure components such as servers, storage, and network
- Enables consumers to deploy and run software, including OS and applications. Clients have control of virtual resources
- Virtualization
- Ex.: AWS, VirtualBox, VMWare, OpenStack



amazon web services™

GOGRID Control in the Cloud™

Platform-as-a-Service

- Capability provided to the consumer to deploy consumer-created or acquired applications on the Cloud provider's infrastructure
- Consumer has control over
  - Deployed applications
  - Possible application hosting environment configurations
- Consumer is billed for platform software components
- Clients use language and proprietary tolls
- Ex.: Google App Engine (GAE), Azure, Facebook platform

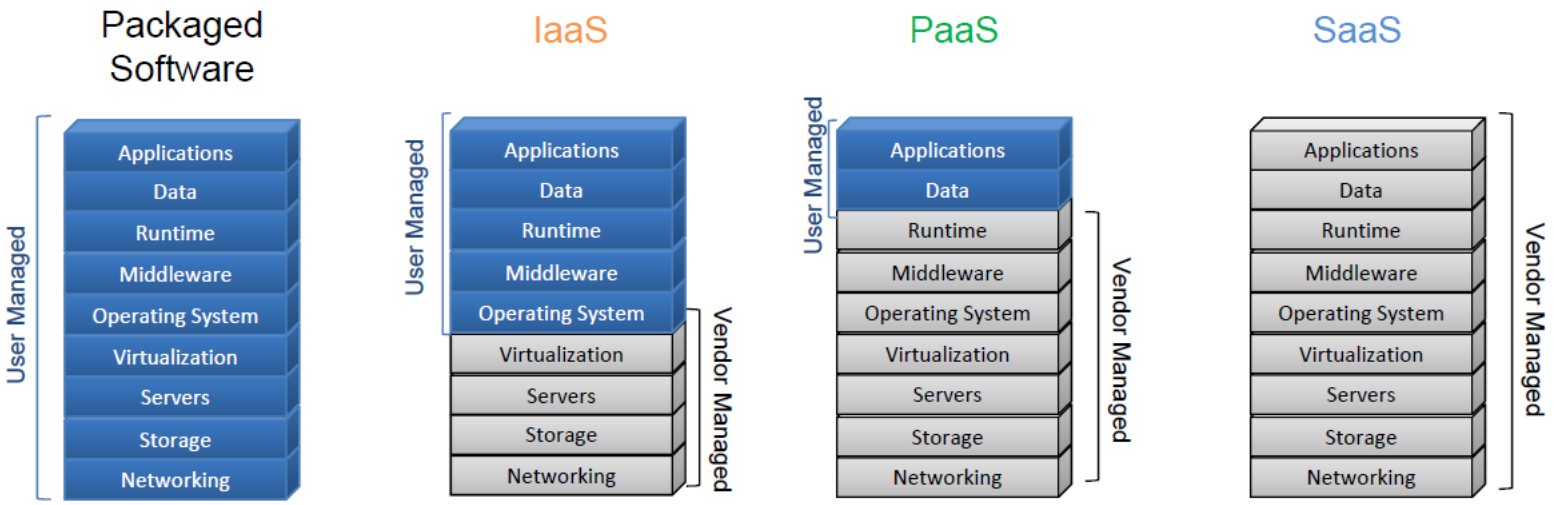
Windows Azure

Google App Engine

Software-as-a-Service

- Capability provided to the consumer to use provider's applications running in a Cloud infrastructure
- Complete stack including application is provided as a service
- Application is accessible from various client devices, for example, via a thin client interface such as a Web browser
- Billing is based on the application usage No control over the cloud
- Ex.: Dropbox, Google Calendar, Yahoo Mail, Gmail, Salesforce.com, Sugar CRM, ...

Cloud Service Layers in the Service Levels



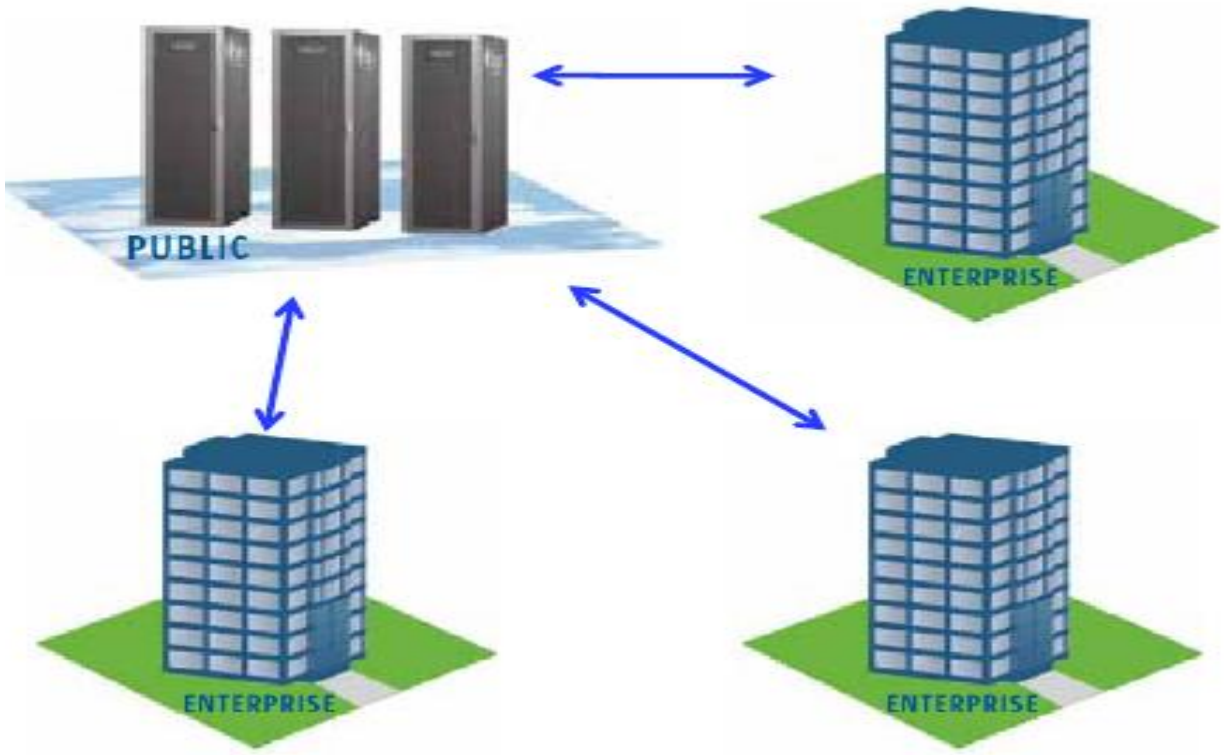
Type of Clouds

Type of Clouds

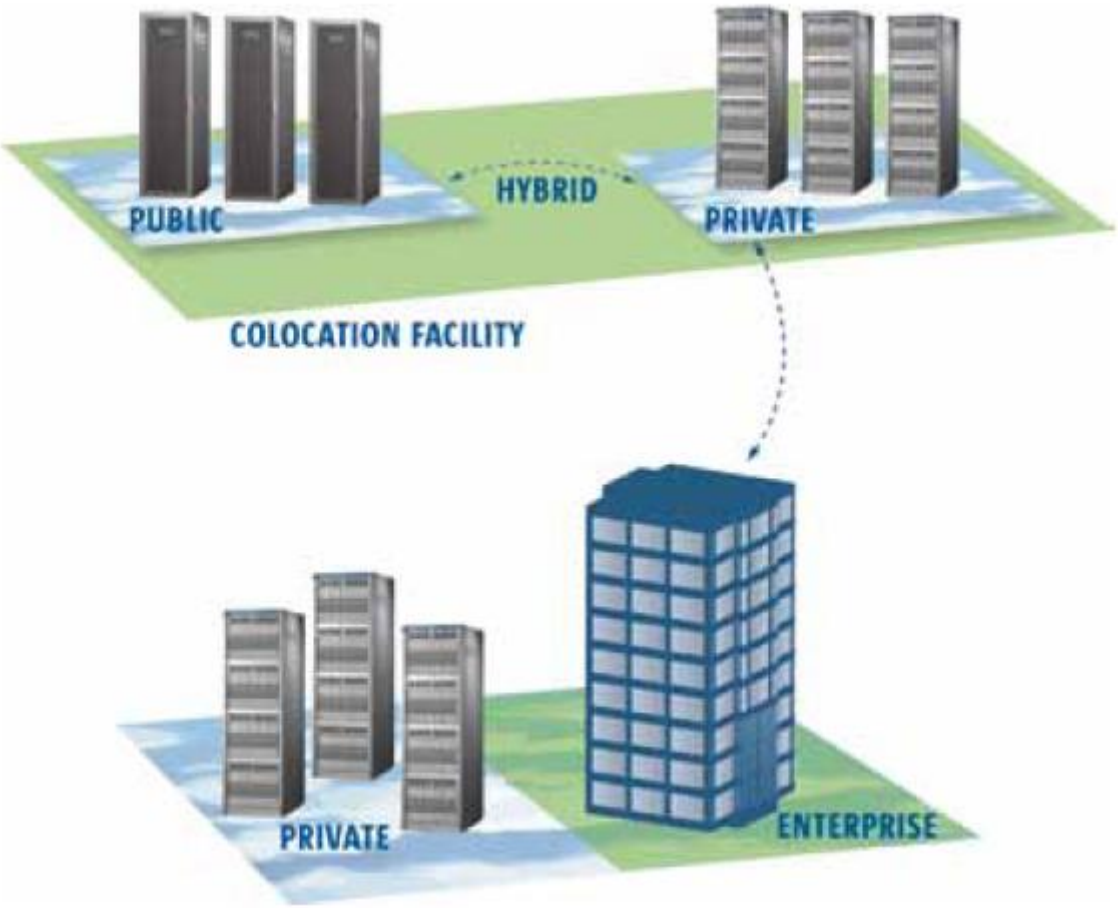
- Private clouds
- Public clouds
- Hybrid clouds
- Community clouds

Private Clouds

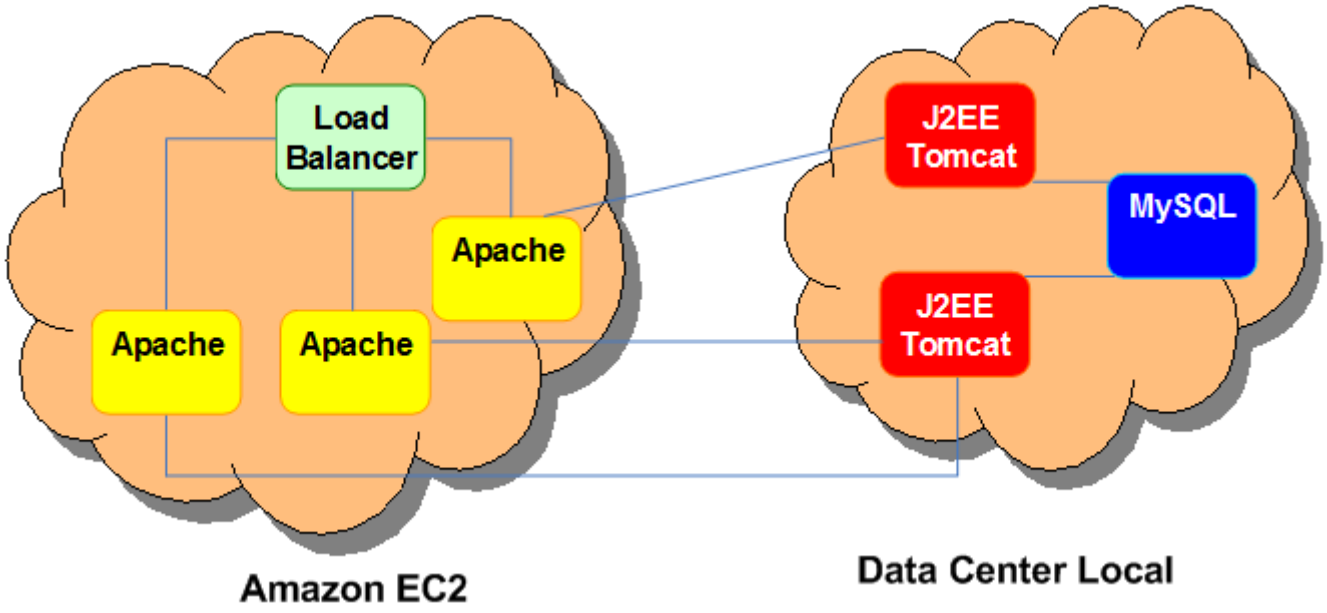
Public Clouds



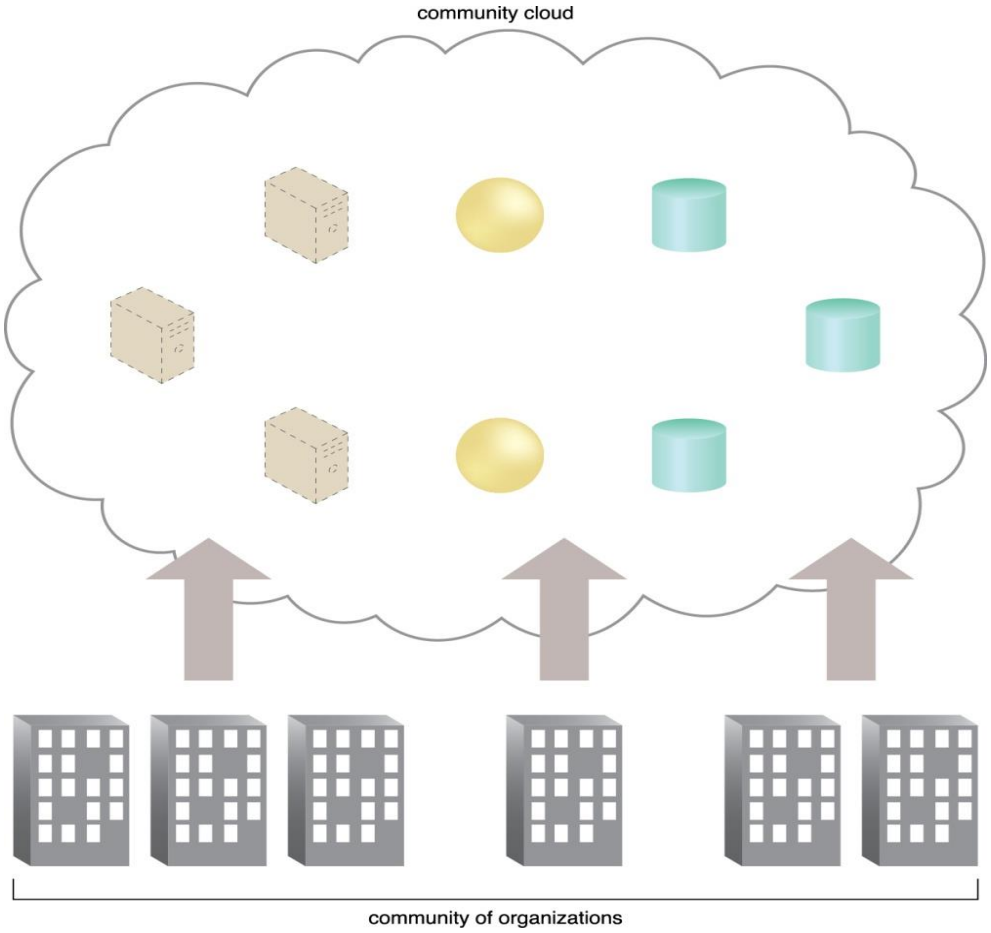
Hybrid Clouds



Hybrid Clouds



Community Clouds



Cloud Computing References

IEEE Cloud Computing Initiative

- http://cloudcomputing.ieee.org/
  - Education
  - Publications
  - Conferences
  - Standards
  - Career

cloudcomputing.ieee.org

Pesquisar

IEEE CLOUD COMPUTING

IEEE

Home | About Us | Conferences | Education & Careers | Publications | Standards | Intercloud | Innovations | Volunteers | Resources

Connect:

March/April 2015 Edition of Cloud Computing Magazine now available for download

How does the emerging cloud of sensors build on the Internet of Things? What are some of the challenges with cloud brokering in enterprise deployment? This issue of IEEE Cloud Computing answers these questions and more.

Read more

In the News

5th IEEE International Conference on Consumer Electronics-Berlin (ICCE-Berlin) call for papers final deadline extension - June 15

IEEE Transactions on Cloud Computing call for EIC nominations

Call for Papers: See the latest calls for papers

The Seven Habits Of Highly Effective Big Data Users (Forbes)

The Institute - Special Report: Big Data - September 2014 (The Institute)

Newly minted MacArthur 'genius' could make cloud computing more secure (VentureBeat)

IEEE Signal Processing Magazine
September 2014

IEEE Network
July 2014

IT Professional
November/December 2013

IEEE Pervasive Computing
October/December 2013

Computer
June 2013

IT Professional
March/April 2013

IEEE Wireless Communications
August 2012

The Institute
June 2012

IEEE Internet Computing
May/June 2012

CLOUD SERVICES, NETWORKING, AND MANAGEMENT

Edited by
Nelson L. S. da Fonseca
Raouf Boutaba

IEEE
IEEE PRESS

IEEE Press Series on Networks and Services Management
Thomas Plevyak and Veli Sahin, *Series Editors*

WILEY