

Fundamentals of Cloud Computing

Tejas Parikh (t.parikh@northeastern.edu)

Spring 2018

CSYE 6225
Northeastern University
<https://spring2018.csye6225.com>

I WAS HOPING FOR
A SLIGHTLY MORE DETAILED
EXPLANATION OF HOW
CLOUD COMPUTING WORKS
THAN - "IT'S MAGIC"!



© D.Fletcher for CloudTweaks.com

On-Premise

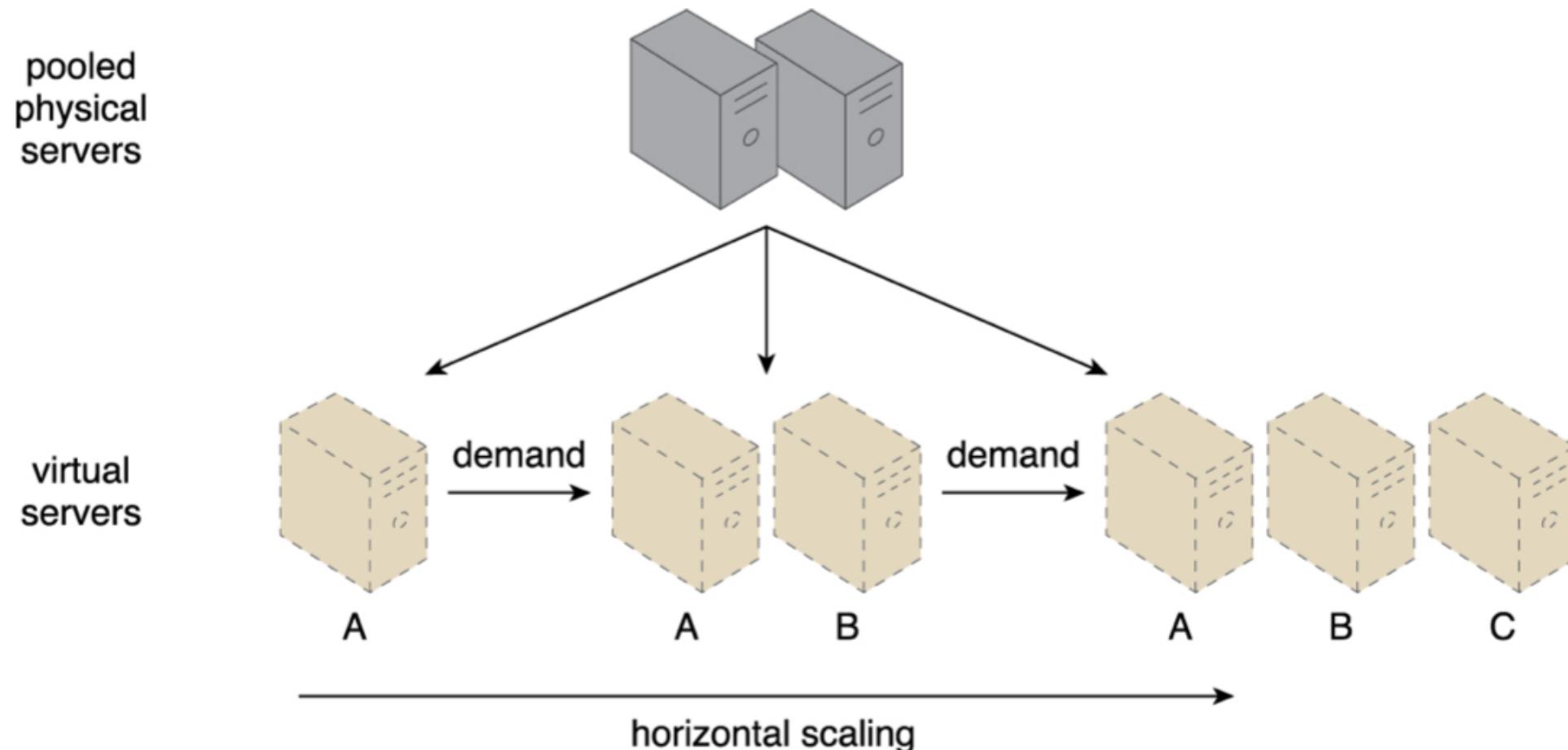
Servers are acquired, operating systems are installed, other hardware may be involved, but all of that lives within your four walls, or the walls of your datacenter.

Scaling

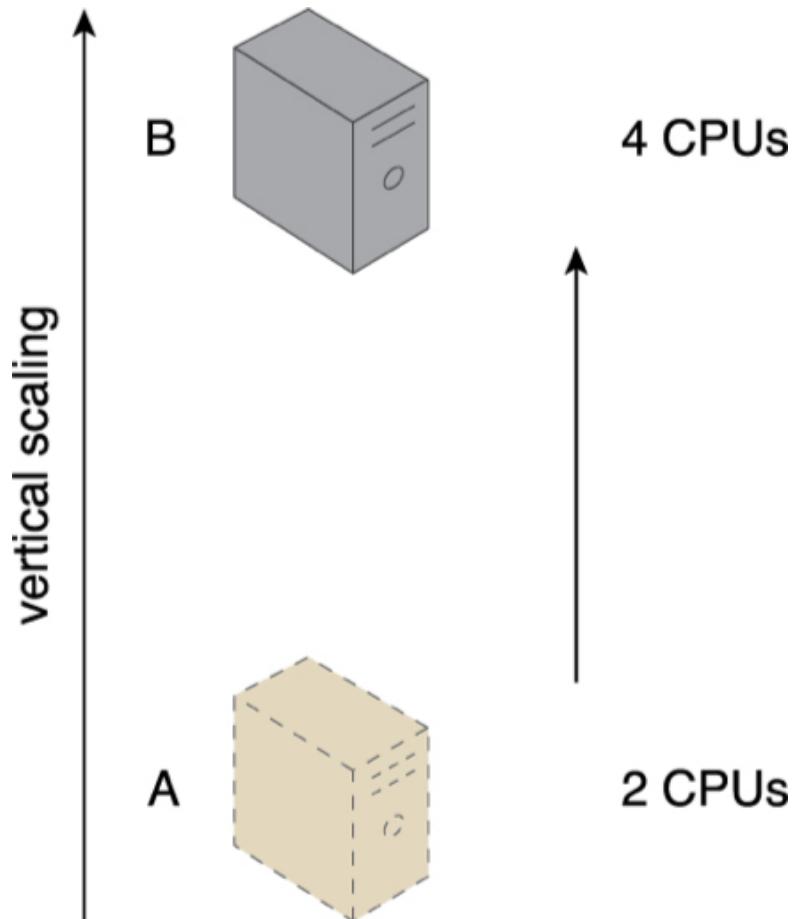
Scaling represents the ability of the resource to handle increased or decreased usage demands. The following are types of scaling:

- Horizontal Scaling – scaling out and scaling in
- Vertical Scaling – scaling up and scaling down

Horizontal Scaling



Vertical Scaling



Business Drivers

- Capacity Planning
- Cost Reduction
- Organizational Agility

Business Driver - Capacity Planning

- Capacity planning is the process of determining and fulfilling future demands of an organization's IT resources, products and services.
- A discrepancy between the capacity of an IT resource and its demand can result in a system becoming either inefficient (over-provisioning) or unable to fulfill user needs (under-provisioning).
- Different capacity planning strategies:
 - Lead Strategy – adding capacity to an IT resource in anticipation of demand
 - Lag Strategy – adding capacity when the IT resource reaches its full capacity
 - Match Strategy – adding IT resources capacity in small increments, as demand increases
- Planning for capacity can be challenging because it requires estimating usage load fluctuations.
- Peak capacity is only rarely used and, consequently, average server utilization levels are often under 20%. Since few users deliberately provision for less than the expected peak, resources are idle at nonpeak times. The more pronounced the variation, the more the waste.

Business Driver – Cost Reduction

- A direct alignment between IT costs and business performance can be difficult to maintain.
- Two costs need to be accounted for
 - The cost of acquiring new infrastructure
 - Cost of its ongoing ownership
- Convert capital expenses to operating expenses (CapEx to OpEx).
- Operational overhead represents a considerable share of IT budgets, often exceeding up-front investment costs.
- Absence of up-front CapEx allows capital to be redirected to core business investment.

Business Driver – Organizational Agility

- Businesses need the ability to adapt and evolve to successfully face change caused by internal and external factors.
- Organizational agility is the measure of an organization's responsiveness to change.
- Up-front investments and infrastructure ownership costs that are required to enable new or expanded business solutions may themselves be prohibitive.

What is Cloud Computing?



Definition of Cloud Computing

According to **Gartner** report Cloud Computing is:

“... a style of computing in which scalable and elastic IT-enabled capabilities are delivered as a service to external customers using Internet technologies.”

Definition of Cloud Computing

Forrester Research defines cloud computing as:

“... a standardized IT capability (services, software, or infrastructure) delivered via Internet technologies in a pay-per-use, self-service way.”

Definition of Cloud Computing

National Institute of Standards and Technology (NIST) defines Cloud Computing as:

“.. a model for enabling ubiquitous, 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. This cloud model is composed of five essential characteristics, three service models, and four deployment models.”

Benefits of Cloud Computing

- Reduce upfront cost of hardware, software and ownership
- Reduce or eliminate need for data center, cooling, backup datacenter & operations related cost
- On-demand access to pay-as-you-go computing resources
- Access to virtually unlimited computing resources
- Access to different kind of computing resources
- Flexibility in scaling resources almost instantaneously
- Increased Availability & Reliability



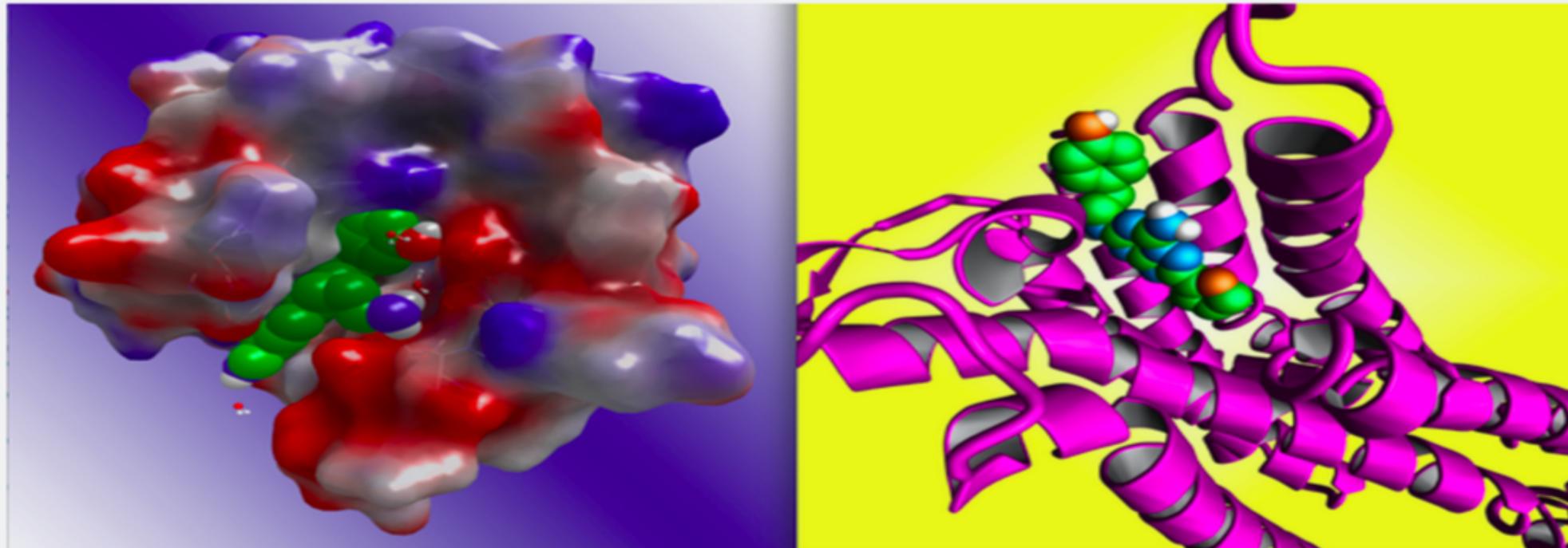
There is no cloud
it's just someone else's computer

MINISTRY OF INNOVATION —

\$4,829-per-hour supercomputer built on Amazon cloud to fuel cancer research

A 50,000-core supercomputer deployed on Amazon shows the cloud's potential

JON BRODKIN - 4/19/2012, 9:00 AM



<https://arstechnica.com/business/2012/04/4829-per-hour-supercomputer-built-on-amazon-cloud-to-fuel-cancer-research/>

Obstacles for Cloud Computing

- Reduced Control over operations.
- Cloud Provider (Vendor) lock-in. Applications & Solutions may become vendor specific over time and make it difficult to migrate to different cloud provider or to on-prem.
- Reduce control over service configuration. Cloud services consumers can no longer fine tune parameters for services such as databases, etc.
- Data Confidentiality & Auditability. Regional compliance and legal issues can make it challenging to move application to cloud if the cloud provider does not meet the requirements. For e.g. HIPPA regulations.
- Cloud consumers usually have no recourse when cloud platform suffers outage.
- Performance Unpredictability
- Software Licensing

Salesforce Outage: Can Customers Trust The Cloud?

Salesforce experienced an outage and service disruption to the NA14 instance, sending customers to Twitter to complain and organizations to evaluate the best way to work with cloud software providers.

AWS outage knocks Amazon, Netflix, Tinder and IMDb in MEGA data collapse

Cloudopocalypse stalks Sunday sofa surfers



BUSINESS
INSIDER

ENTERPRISE

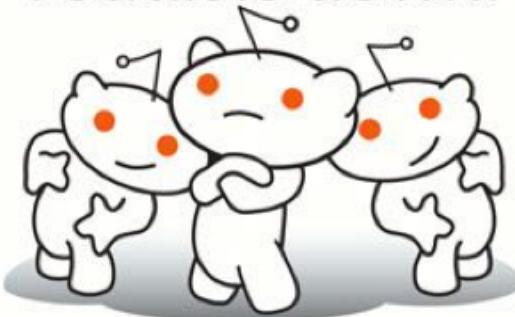
Google apologizes for cloud outage that one person describes as a 'comedy of errors'

PayPal, 2009

In 2009, PayPal experienced a worldwide system outage for approximately five hours. The company handled \$2,000 in online commerce every second at the time, suggesting the event interfered with \$36 million worth of personal and business transactions.

Amazon is currently experiencing a degradation. They are [working on it](#).

reddit is down.



Follow

Pinterest is currently unavailable due to server outages. Our goal is to be back up by 10:30PM PST. Thanks for your patience!

Reply Retweet Favorite

10:07 PM - 29 Jun 12 via Mobile Web · Embed this Tweet

Azure status



We're having issues.

But we're working on it...



The sky is falling! Amazon's cloud seems to be down (raining?) so we're experiencing some issues too. Be back soon!

5 hours ago via web

Retweeted by [RealAmandaStone](#) and others



Amazon's massive AWS outage was caused by human error

One incorrect command and the whole internet suffers.

BY JASON DEL REY | @DELREY | MAR 2, 2017, 2:20PM EST



<https://www.recode.net/2017/3/2/14792636/amazon-aws-internet-outage-cause-human-error-incorrect-command>

Cloud Characteristics

For an IT environment to be considered cloud, it must meet following 5 essential characteristics:

1. On-demand self service
2. Broad Network Access
3. Resource Pooling (multitenancy)
4. Elasticity
5. Measured Service



Cloud Characteristics - On-demand self service

- A cloud consumer can unilaterally access cloud-based IT resources giving the cloud consumer the freedom to self-provision these IT resources.
- Once configured, usage of the self-provisioned IT resources can be automated, requiring no further human involvement by the cloud consumer or cloud provider. This results in an on-demand usage environment.

Cloud Characteristics - Broad Network Access

- Ability for a cloud service to be widely accessible.
- Establishing ubiquitous access for a cloud service can require support for a range of devices, transport protocols, interfaces, and security technologies.

Cloud Characteristics - Resource Pooling (multitenancy)

- The characteristic of a software program that enables an instance of the program to serve different consumers (tenants) whereby each is isolated from the other, is referred to as multitenancy.
- A cloud provider pools its IT resources to serve multiple consumers by using multitenancy models that frequently rely on the use of virtualization technologies.
- Through the use of multitenancy technology, IT resources can be dynamically assigned and reassigned, according to cloud service consumer demands.

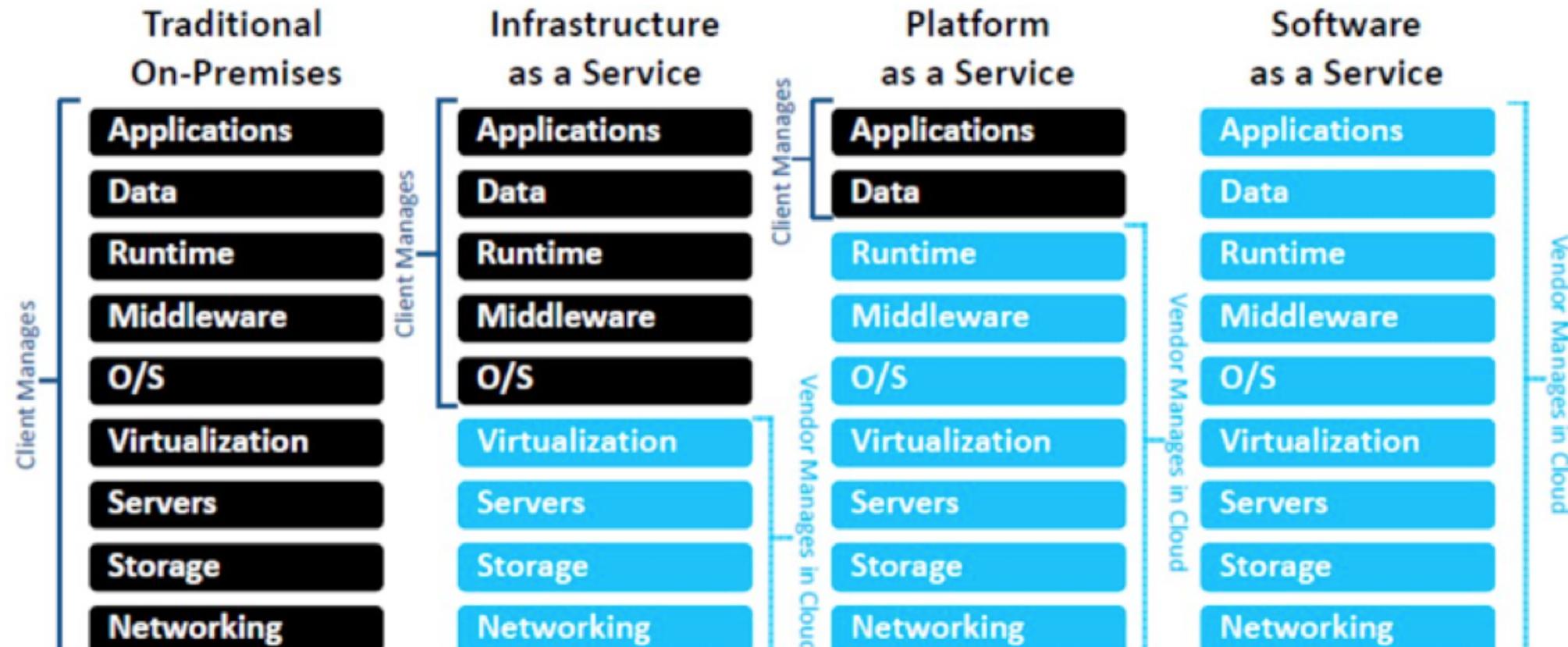
Cloud Characteristics - Elasticity

- Elasticity is the automated ability of a cloud to transparently scale IT resources, as required in response to runtime conditions or as pre-determined by the cloud consumer or cloud provider.
- Elasticity is often considered a core justification for the adoption of the cloud computing, primarily due to the fact that it is closely associated with reduced investment and proportional cost benefit.

Cloud Characteristics – Measured Usage

- The measured usage characteristic represents the ability of cloud platform to keep track of the usage of its IT resources, primarily by cloud consumers.
- Based on what is measured, the cloud provider can charge a cloud consumer only for the IT resources actually used and/or for the timeframe during which access to the IT resources was granted.

Cloud Delivery Models



Customization; higher costs; slower time to value

Standardization; lower costs; faster time to value

Cloud Delivery Models



Infrastructure-as-a-Service (IaaS)

- IaaS provides self-contained IT environment consisting of hardware, networking, connectivity, Operating Systems and other “raw” IT resources.
- IaaS provides cloud consumers with a high level of control and responsibility over its configuration and utilization.
- IaaS providers may offer customized physical hardware for servers, load balancers, Gateways, firewalls, Intrusion Detection System (IDS) and Intrusion Prevention Systems.
- With IaaS, cloud consumer is responsible for securing the “raw” IT resources.
- Popular IaaS providers are Amazon Web Services, Google Cloud Platform, Microsoft Azure, IBM SoftLayer, Rackspace

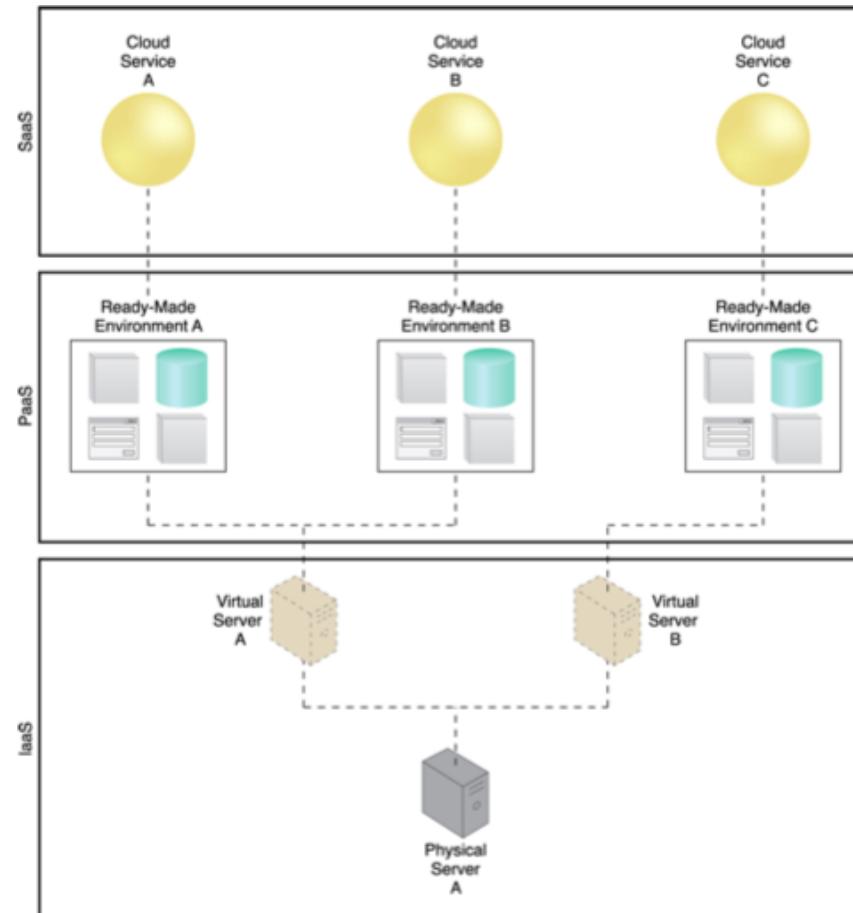
Platform as a Service (PaaS)

- PaaS provides “ready-to-use” environment to deploy applications.
- Cloud consumers do not have control over or manage the underlying cloud infrastructure such as network, servers, operating systems, storage, etc.
- Popular PaaS providers are Heroku, IBM BlueMix and Google App Engine

Software as a Service (SaaS)

- A software program positioned as a shared cloud service and made available as a “product” or generic utility represents the typical profile of a SaaS offering.
- Consumer of PaaS does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user specific application configuration settings.
- Popular SaaS offering examples are Gmail, GitHub, Google Drive, etc.

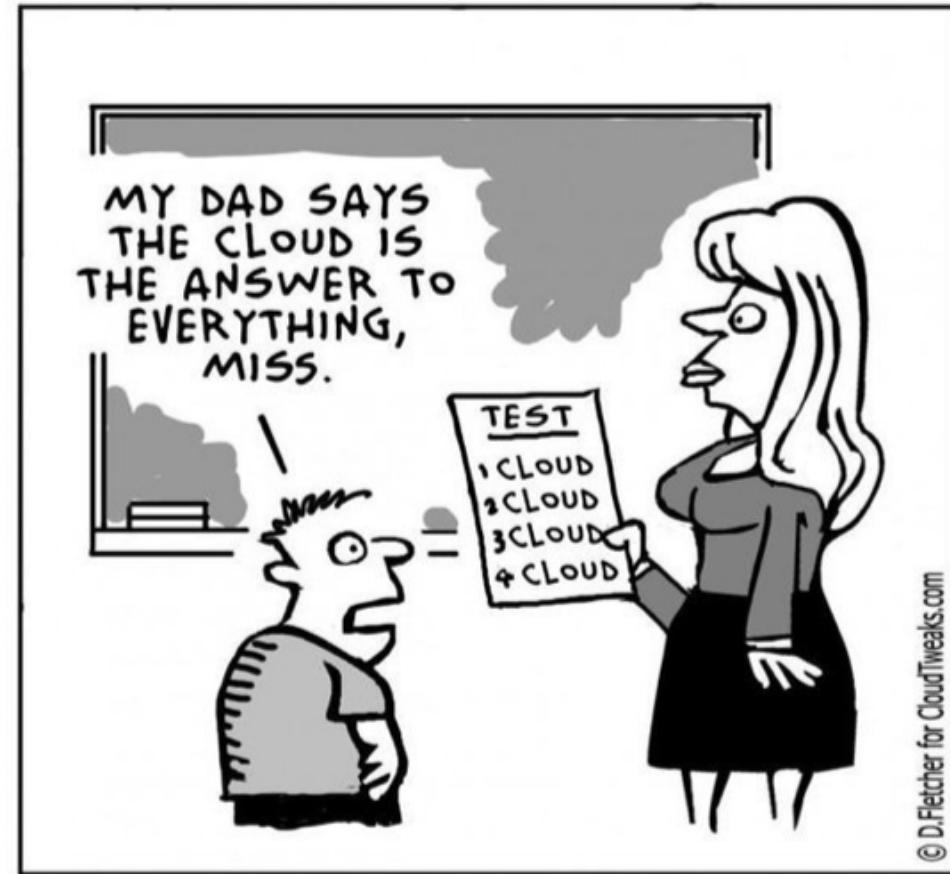
IaaS + PaaS + SaaS



Cloud Deployment Models

There are four common cloud deployment models:

1. Public cloud
2. Community cloud
3. Private cloud
4. Hybrid cloud



© D.Fletcher for CloudTweaks.com

Public Cloud

- Cloud infrastructure is provisioned for open use by the general public.
- Cloud provider is responsible for the creation of on-going maintenance of public cloud and its IT resources

Community Cloud

- A community cloud is similar to a public cloud except that its access is limited to a specific community of cloud consumers.
- Example of Community cloud is AWS GovCloud.
- AWS GovCloud (US) is an isolated AWS region designed to host sensitive data and regulated workloads in the cloud, helping customers support their U.S. government compliance requirements, including the International Traffic in Arms Regulations (ITAR) and Federal Risk and Authorization Management Program (FedRAMP). AWS GovCloud (US) is operated by employees who are vetted "U.S. Persons" and root account holders of AWS accounts must confirm they are U.S. Persons before being granted access credentials to the region.

Private Cloud

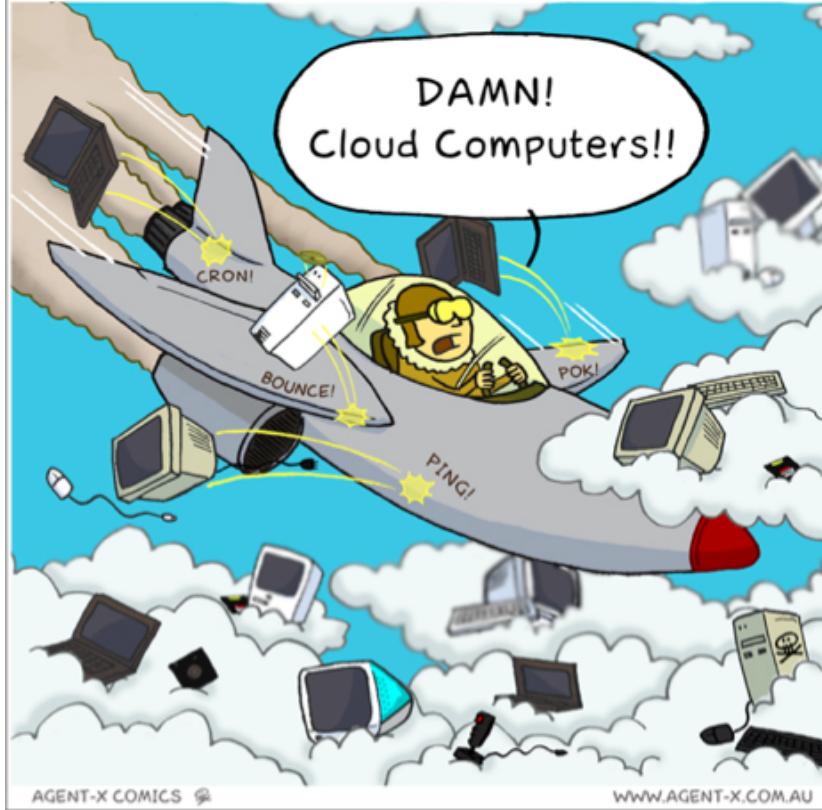
- The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units).
- Private clouds enable an organization to use cloud computing technology as a means of centralizing access to IT resources by different parts, locations, or departments of the organization.
- Private cloud mitigate some of the risks of public cloud as organization has much higher level of control.
- Even though private cloud resources might be hosted and controlled by the organization, it should not be confused with on-prem.

Hybrid Cloud

- A hybrid cloud is a cloud environment comprised of two or more different cloud deployment models.
- For example, a cloud consumer may choose to deploy cloud services processing sensitive data to a private cloud and other, less sensitive cloud services to a public cloud.

New Aspects in Cloud Computing from Hardware Provisioning & Pricing Point of View

- The appearance of infinite computing resources available on demand, quickly enough to follow load surges, thereby eliminating the need for cloud computing users to plan far ahead for provisioning.
- No commitments. Cloud users do not pay upfront fees allowing users to start small and increase resources only when there is an increase in their needs.
- The ability to pay for use of computing resources on a short-term basis as needed.



Additional Resources

<https://spring2018.csye6225.com>