CSCI 5409 Cloud Computing
Fall, 2023
Week 2 – Lecture 2 (Sep 15, 2023)

Introduction to Cloud Computing – Key Terms

Dr. Lu Yang
Faculty of Computer Science
Dalhousie University
luyang@dal.ca

Housekeeping and Feedback

- Differences of AWS Academy Foundations course and AWS Academy Learner Lab
- Newly registered students please ask me to add you to Teams and AWS courses
- Upload the slide before class

Objectives

- Key terminology in cloud computing
  - we all need to speak the same language
- Understand the differences of cloud roles
- Understand the characteristics of cloud computing

Contents

1

Background, Terms and Concepts

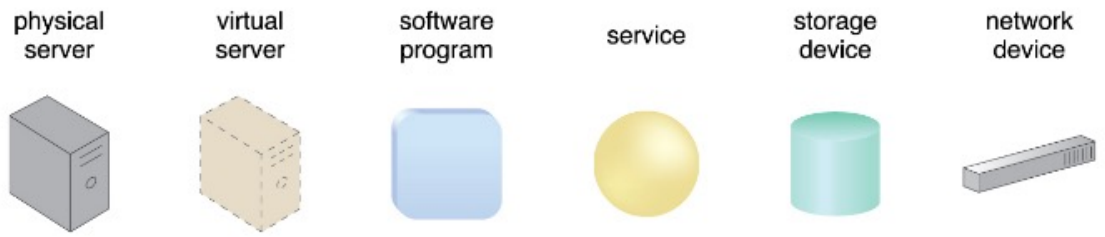
1. Generations of Computer Systems
2. IT Resource
3. On-Premise Resource
4. Cloud Definition
5. Scaling
6. Provisioning

Generations of Computer Systems

Generation	Timeframe	Technologies
0	Before 1940	Mechanical calculating machines
1	1940 – 1956	Vacuum tubes, magnetic drums
2	1956 – 1963	Transistors
3	1964 – 1971	Integrated circuits, keyboard, monitors
4	1971 – 2010	VLSI, CPU, desktop, laptop
5	Present	Network, AI, Virtualization, Cloud

IT Resource

"An IT resource is a physical or virtual IT-related artifact that can be either software-based, such as a virtual server or a custom software program, or hardware based, such as a physical server or network device." [1]



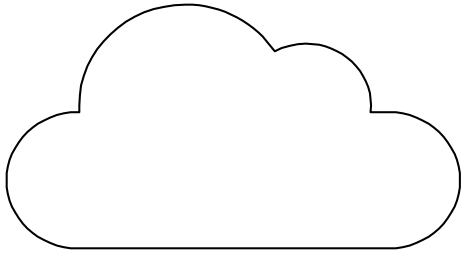
1. Cloud Computing (T. Erl, Z. Mahmood, R.Puttini), page 34

IT Resource (Cont'd)

IT Resource

On-premise

Cloud



AWS, Google, Microsoft Azure, IBM, etc.

On-Premise Resource

- "An IT resource that is hosted in a conventional IT enterprise within an organizational boundary (that does not specifically represent a cloud) is considered located on the premises of the IT enterprise, or on-premise for short."[1]
- An IT resource that is on–premise cannot be cloud-based, and vice-versa
- In simple terms this usually means a server running within an organization's own building or data center connected directly to the internet serving something up
- These are the IT resources an organization wishes to eliminate when transitioning to the cloud

1. Cloud Computing (T. Erl, Z. Mahmood, R.Puttini), page 36

Cloud Definition (1/2)

Definition 1: "It's a term used to describe a global network of servers, each with a unique function. The cloud is not a physical entity, but instead is a vast network of remote servers around the globe which are linked together and meant to operate as a single ecosystem. These servers are designed to:

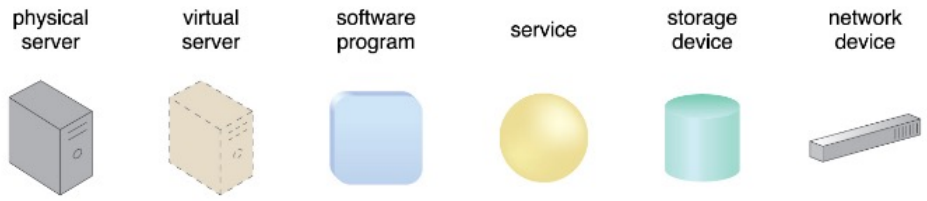
- Store and manage data
- Run applications or deliver content or a service such as streaming videos, web email, office productivity software or social media
- Be accessed online from any Internet-capable device – the information will be available wherever you go and whenever you need it."[1]

1. https://azure.microsoft.com/en-ca/overview/what-is-the-cloud/

Cloud Definition (2/2)

Definition 2: "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. This cloud model promotes availability and is composed of:

- Five essential characteristics (On-demand self-service, Broad network access, Resource pooling, Rapid elasticity, Measured Service);
- Three service models (Cloud Software as a Service (SaaS), Cloud Platform as a Service (PaaS), Cloud Infrastructure as a Service (IaaS));
- Four deployment models (Private cloud, Community cloud, Public cloud, Hybrid cloud); and,
- Key enabling technologies:
  - (1) fast wide-area networks,
  - (2) powerful, inexpensive server computers, and
  - (3) high-performance virtualization for commodity hardware." [1]



1. https://csrc.nist.gov/projects/cloud-computing

Cloud Discussion

- *"If computers of the kind I have advocated become the computers of the future, then computing may someday be organized as a public utility just as the telephone system is a public utility. … The computer utility could become the basis of a new and important industry."* – John McCarthy (Stanford - 1961) (a founder of AI, creator of Lisp)
- A **cloud** refers to a distinct IT environment that is designed for the purpose of remotely provisioning scalable and measured IT resources
- Cloud != Internet, it has a finite boundary

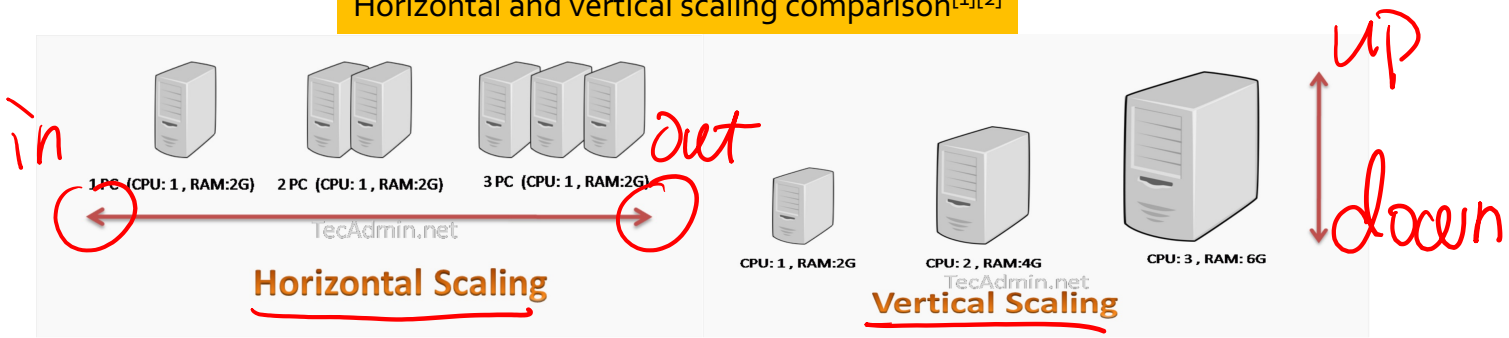
Scaling (1/2)

- "Scaling, from an IT resource perspective, represents the ability of the IT resource to handle increased or decreased usage demands"[1]
- **Horizontal Scaling** is the allocation or release of IT resources that are of the same type
  - Horizontal allocation is referred to as **scaling out**
  - Horizontal release is referred to as **scaling in**
- **Vertical Scaling** is the replacement of an existing IT resource by another with higher or lower capacity
  - Replacing with higher capacity is referred to as **scaling up**
  - Replacing with lower capacity is referred to as **scaling down**

1. Cloud Computing (T. Erl, Z. Mahmood, R.Puttini), page 37

Scaling (2/2)

Horizontal and vertical scaling comparison[1][2]



Horizontal Scaling	Vertical Scaling
Less expensive	More expensive
IT resources instantly available	IT resources normally instantly available
Resource replication and automated scaling	Additional setup is normally needed
Additional IT resources needed	No additional IT resources needed
Not limited by hardware capacity	Limited by maximum hardware capacity

1. https://tecadmin.net/tutorial/what-is/horizontal-scaling/
2. https://crmtrilogix.com/Cloud-Blog/Cloud-Design-Goals/Scaling-and-Cloud-Services/181

Provisioning

- "Provisioning is the process of setting up IT infrastructure. It can also refer to the steps required to manage access to data and resources and make them available to users and systems.
- Provisioning is not the same thing as configuration, but they are both steps in the deployment process. Once something has been provisioned, the next step is configuration.
- When the term "provisioning" is used, it can mean many different types of provisioning, such as server provisioning, network provisioning, user provisioning, service provisioning, and more."[1]

1. https://www.redhat.com/en/topics/automation/what-is-provisioning

2

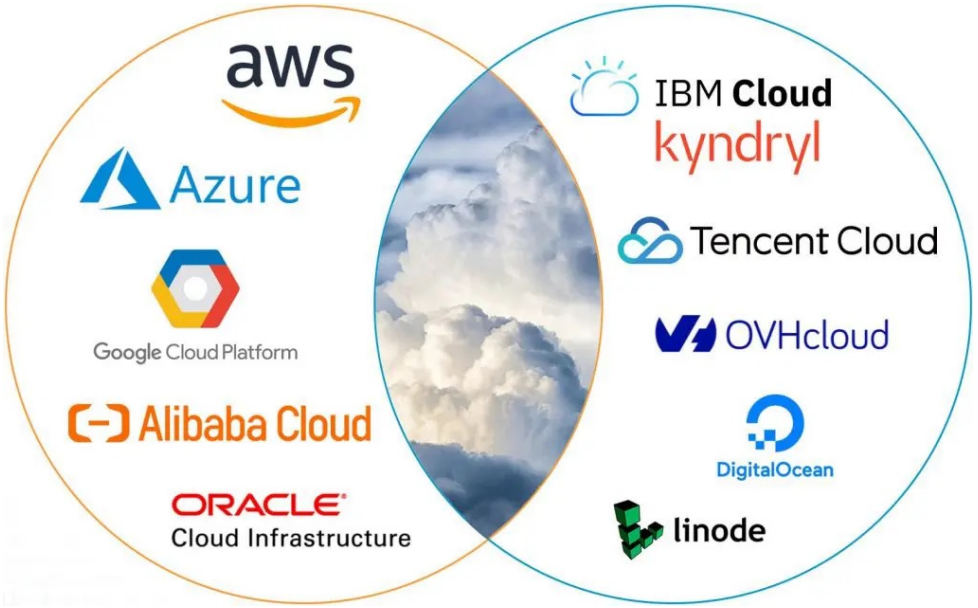
Cloud Roles

1. Cloud Provider
2. Cloud Consumer
3. Cloud Service Consumer
4. Cloud Service Owner
5. Cloud Resource Administrator

Cloud Provider

- The organization that provides cloud-based IT resources is the cloud provider [1]

#	Cloud Service Provider
1	Amazon Web Services (AWS)
2	Microsoft Azure
3	Google Cloud Platform (GCP)
4	Alibaba Cloud
5	Oracle Cloud
6	IBM Cloud (Kyndryl)
7	Tencent Cloud
8	OVHcloud
9	DigitalOcean
10	Linode (Akamai)



1. Cloud Computing (T. Erl, Z. Mahmood, R.Puttini)
2. Table and figure from: https://dgtlinfra.com/top-10-cloud-service-providers-2022/

Cloud Consumer

- "The party that uses cloud-based IT resources is the cloud consumer" [1]
- A cloud consumer is an organization (or a human) that has a formal contract or arrangement with a cloud provider to use IT resources made available by the cloud provider
- Major cloud consumers:[2]
  - Video streaming (youtube, twitch, pornography, etc.) (65% internet traffic)
  - Netflix (15% internet traffic)

Featured Customers

Moderna

As a biotechnology company, one of Moderna's biggest breakthroughs was the development of its mRNA COVID-19 vaccine with a level of speed, effectiveness, and scale made possible by AWS.

Read more about Moderna on AWS »

Capital One

Capital One's all-in cloud migration to AWS has strengthened and expanded its reputation and reach as an industry disrupter—making it a magnet for top-tier engineers and non-technical talent.

Read more about Capital One on AWS »

Netflix

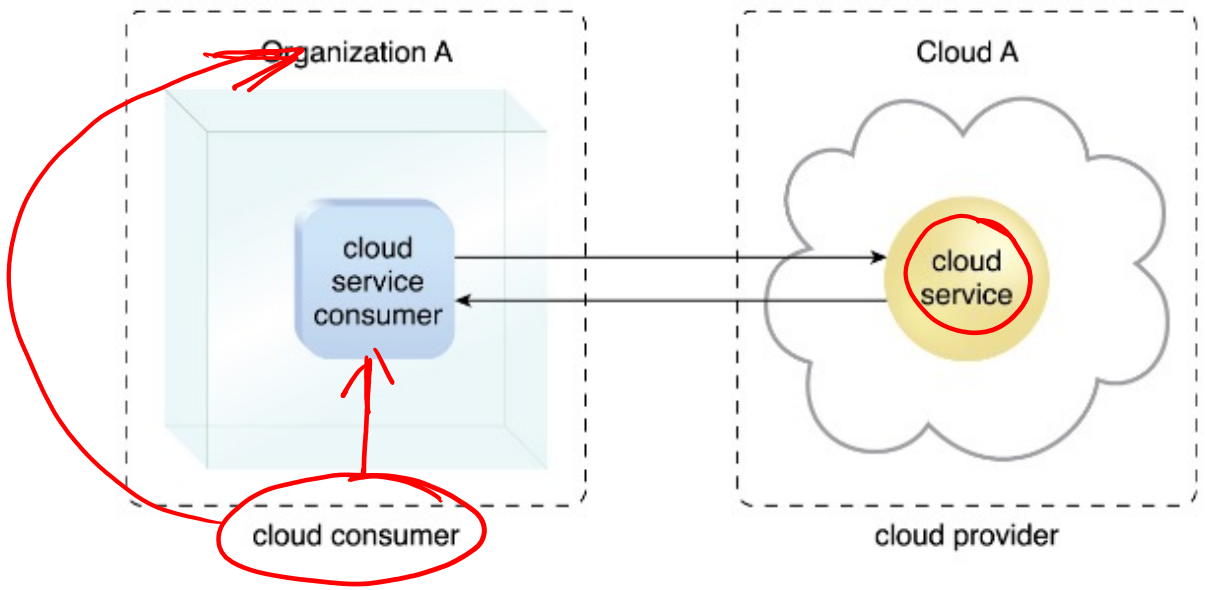
Netflix built a visual effects (VFX) studio on AWS that eliminates technological and geographic barriers and allows top VFX and animation artists to seamlessly collaborate, expanding the internet television network's content production capabilities.

Read more about Netflix on AWS »

1. Cloud Computing (T. Erl, Z. Mahmood, R.Puttini), page 36
2. https://variety.com/2019/digital/news/netflix-loses-title-top-downstream-bandwidth-application-1203330313/
3. https://aws.amazon.com/solutions/case-studies/?customer-references-cards.sort-by=item.additionalFields.sortDate&customer-references-cards.sort-order=desc&awsf.customer-references-location=*all&awsf.customer-references-segment=*all&awsf.customer-references-industry=*all&awsf.customer-references-use-case=*all&awsf.customer-references-tech-category=*all&awsf.customer-references-product=*all

Cloud Service Consumer

- A cloud service consumer is:
  - "Software programs or applications that programmatically interface with a cloud service's technical contract or API" [1] or
  - "A temporary runtime role assumed by a software program when it accesses a cloud service." [2]
- The cloud consumer uses a cloud service consumer to access a cloud service[2]

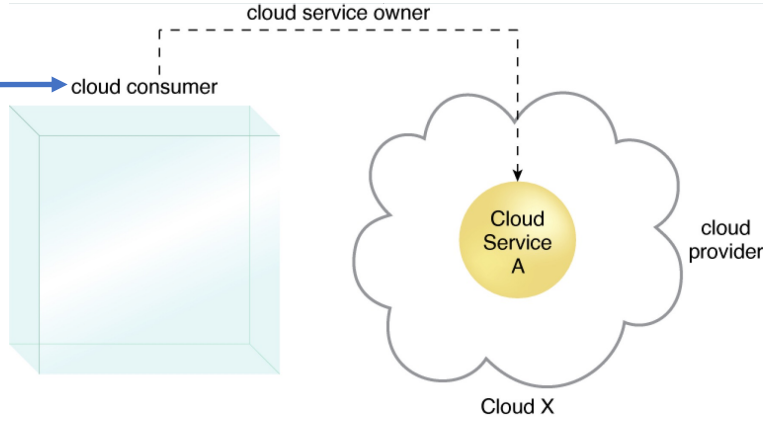


1. Cloud Computing (T. Erl, Z. Mahmood, R.Puttini), page 53
2. https://patterns.arcitura.com/cloud-computing-patterns/basics/roles-and-boundaries/cloud_consumer

Cloud Service Owner

- "The person or organization that legally owns a cloud service is called a cloud service owner" [1]
- A cloud service owner can be a cloud consumer OR a cloud provider that owns the cloud within which the cloud service resides:
  - AWS Lambda is a cloud service owned by Amazon AWS
  - The IMDB API hosted on AWS is a cloud service owned by IMDB
    - AWS is the cloud provider
    - IMDB is the cloud consumer
    - IMDB is the cloud service owner of IMDB API (IMDB API is a cloud service)
    - A cloud service consumer of IMDB calls the IMDB API

A cloud consumer can be a cloud service owner when it deploys its own service in a cloud



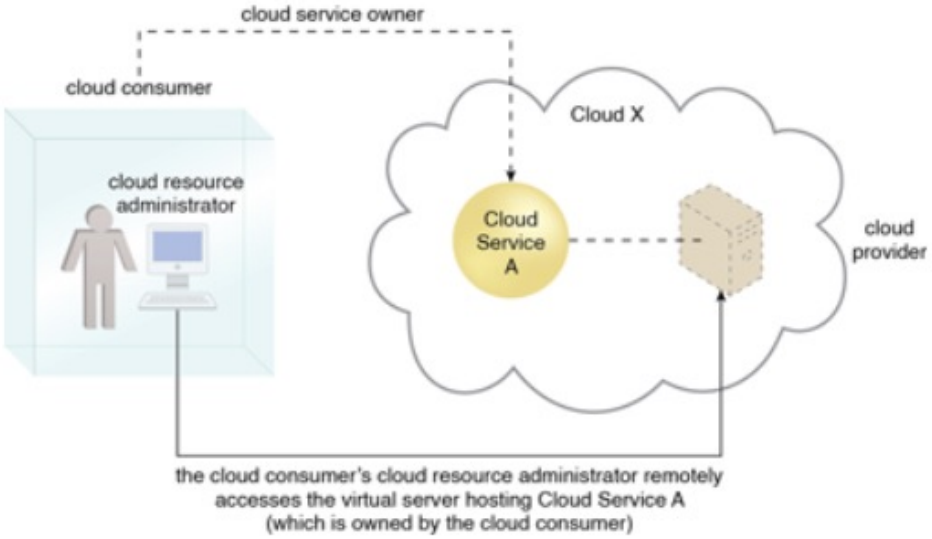
1. Cloud Computing (T. Erl, Z. Mahmood, R.Puttini), page 53
https://patterns.arcitura.com/soa-patterns/basics/soaproject/cloudserviceowner

Cloud Resource Administrator (1/2)

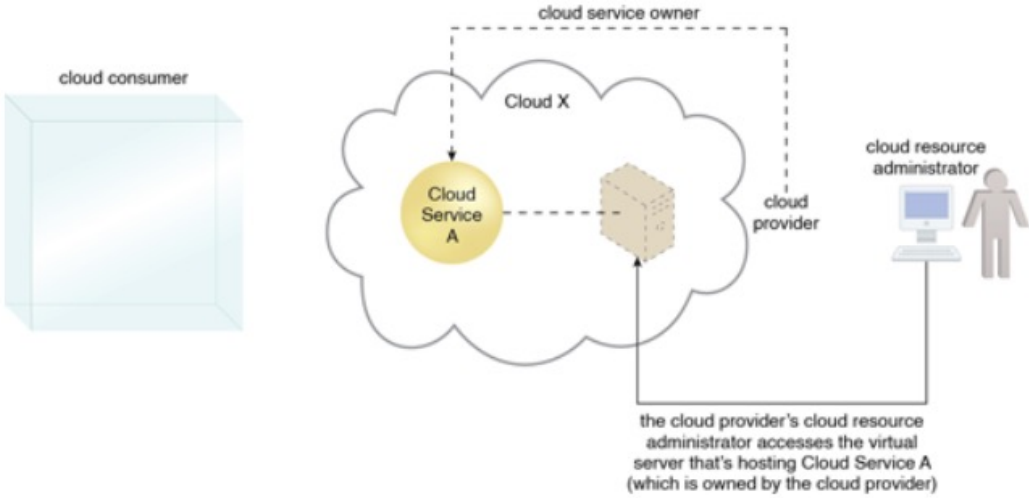
- A cloud resource administrator is the person or organization responsible for administering a cloud-based IT resource (including cloud services)
- The cloud resource administrator can be (or belong to)
  - the cloud consumer or
  - cloud provider of the cloud within which the cloud service resides.
  - Alternatively, it can be (or belong to) a third-party organization contracted to administer the cloud-based IT resource.
- This is the person who can provision IT resources

https://patterns.arcitura.com/cloud-computing-patterns/basics/roles-and-boundaries/cloud_resource_administrator

Cloud Resource Administrator (2/2)



Scenario 1: A cloud resource administrator can be with a cloud consumer organization and administer remotely accessible IT resources that belong to the cloud consumer.



Scenario 2: A cloud resource administrator can be with a cloud provider organization for which it can administer the cloud provider's internally and externally available IT resources.

https://patterns.arcitura.com/cloud-computing-patterns/basics/roles-and-boundaries/cloud_resource_administrator

3

Cloud Characteristics

1. On-demand usage
2. Ubiquitous usage
3. Multitenancy
4. Elasticity Measured Usage
5. Resiliency
6. Summary

On-Demand Usage

disposable

- Freedom to self-provision the IT resources
- Once configured, usage of the self-provisioned IT resources can be automated
- Resources do not incur cost when they are "off"
- Resources incur organization money when they are "on"

Rephased from: Cloud Computing (T. Erl, Z. Mahmood, R.Puttini), page 59

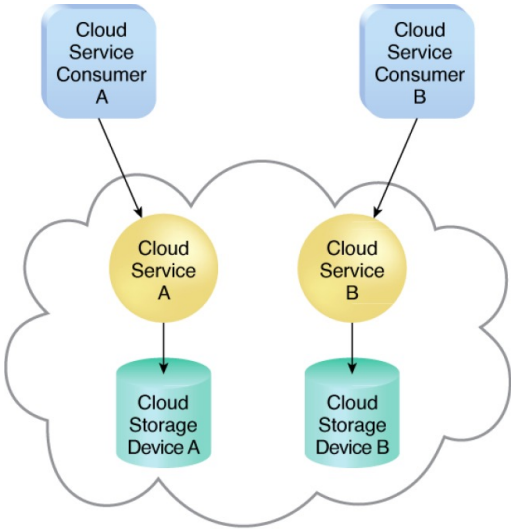
Ubiquitous Access

- The ability for a cloud service to be widely accessible across a network
- Requires:
  - Support for a range of devices (smartphones, tablets, computers, fridges, etc.)
  - Transport protocols (TCP, HTTP, etc.) 4 7
  - Security technologies (Transport Layer Security/Secure Socket Layer, etc.)

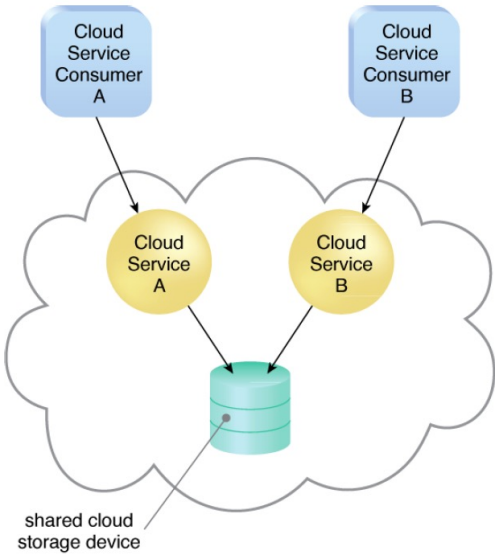
Rephased from: Cloud Computing (T. Erl, Z. Mahmood, R.Puttini), page 59

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."[1]
- "Different physical and virtual IT resources are dynamically assigned and reassigned according to cloud consumer demand, typically followed by execution through statistical multiplexing."[1]
- Without virtualization this would be impossible, we'd need one physical server every customer!



In a single-tenant environment, each cloud consumer has a separate IT resource instance.



In a multitenant environment, a single instance of an IT resource, such as a storage device, serves multiple consumers.

1. Cloud Computing (T. Erl, Z. Mahmood, R.Puttini), page 59-60

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 could consumer or cloud provider."[1]
- This is the core reason we are in the cloud. Cloud providers are assuming the costs of physical resources, cloud consumers are saved the cost of physical resources they are not using.
- Cloud consumers benefit from economies of scale
- Cloud providers make money through multitenancy, there is always something in demand somewhere! Hmm maybe this explains the recent move into game technology, to make use of evening hours?

CloudWatch

1. Cloud Computing (T. Erl, Z. Mahmood, R.Puttini), page 61

Measured Usage

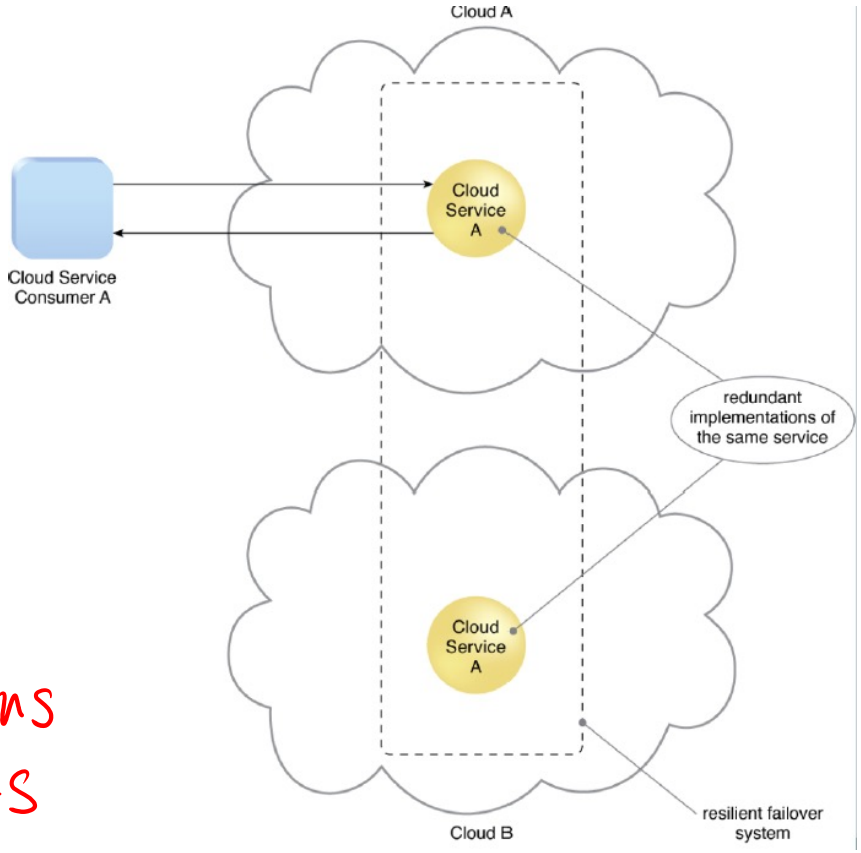
- "The ability of a cloud platform to keep track of the usage of its IT resources, primarily by cloud consumers."[1]
- Enables cloud providers to charge cloud consumers for exactly what they use
- Measured in different ways, depending on the resources:
  - Bandwidth consumed
  - Storage space used
  - Computing cycles
  - Memory
  - GPU and memory
  - CPU cores
  - Requests
  - Etc…the options are endless and custom to the service provided

1. Cloud Computing (T. Erl, Z. Mahmood, R.Puttini), page 61

Resiliency

- "Resilient computing is a form of failover that distributes redundant implementations of IT resources across physical locations."[1]
- Cloud consumers can increase both the reliability and availability of their applications by leveraging the resiliency of cloud-based IT resources
- Can refer to:
  - Redundant resources within the same cloud, but in different physical locations (e.g. ASW regions)
  - Multiple clouds (e.g. AWS and Azure)

AZs
Regions
Clouds

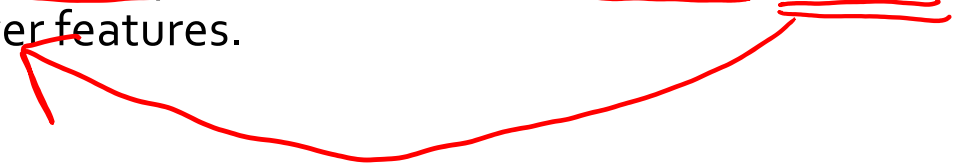


A resilient system in which Cloud B has a redundant implementation of Service A to provide a failover capacity in case Service A on Cloud A becomes unavailable.

1. Cloud Computing (T. Erl, Z. Mahmood, R.Puttini), page 61

Cloud Characteristics Summary

- On-demand usage is the ability of a cloud consumer to self-provision and use necessary cloud-based services without requiring cloud provider interaction. This characteristic is related to measured usage, which represents the ability of a cloud to measure the usage of its IT resources.
- Ubiquitous access allows cloud-based services to be accessed by diverse cloud service consumers, while multitenancy is the ability of a single instance of an IT resource to transparently serve multiple cloud consumers simultaneously.
- The elasticity characteristic represents the ability of a cloud to transparently and automatically scale IT resources out or in. Resiliency pertains to a cloud's inherent failover features.




Kahoot!