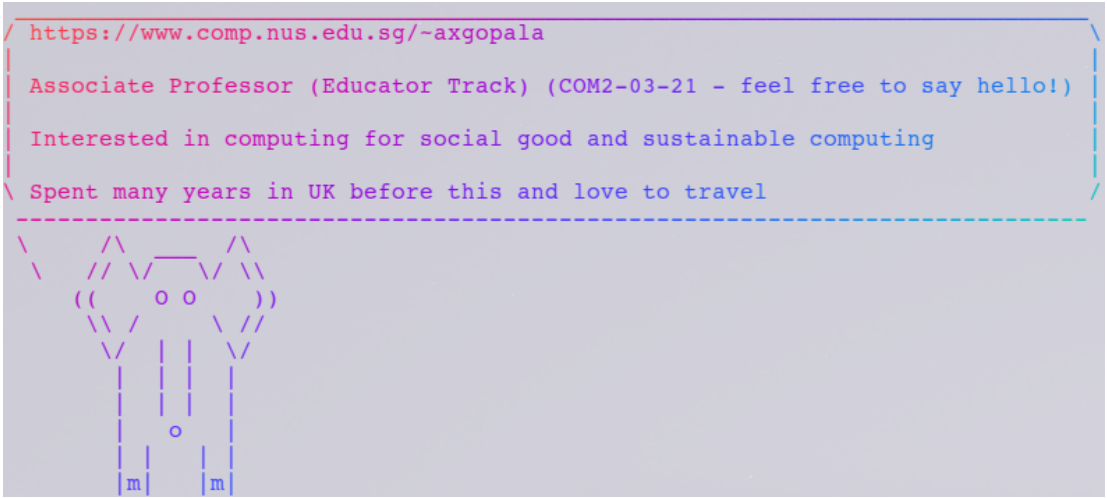
Introduction

Anandha Gopalan
(with thanks to Teo Yong Meng)

axgopala@nus.edu.sg

Who am I?


https://www.comp.nus.edu.sg/~axgopala
Associate Professor (Educator Track) (COM2-03-21 - feel free to say hello!)
Interested in computing for social good and sustainable computing
Spent many years in UK before this and love to travel

Course Information

Canvas will be used as the LMS

- Lecture slides (available at start of the week)
- Quizzes
- Assessments
- Link to webcast of lectures available on Panopto

EdStem is preferred for Q & A

- Join using: `https://edstem.org/us/join/XdRzxa`
- Allows creation of threads anonymously, though not anonymous to the teaching staff
- Allows creation of private threads – visible only to teaching staff

NUS National University of Singapore

Course TAs

Name: Tanuj Sur
PhD student

Name: Dan Yi Jia
MComp student

Name: Ryan Lu Soon Han
MComp student

Name: Yaswanth Tavva
PhD student

Name: Fu Yihao
PhD student

Teaching Style

Interactive, so please do engage ☺

Feel free to ask any questions ☺

No separate tutorial or lab sessions

You learn by doing in this course

Lectures provide the theory, while labs and group project help augment your learning

Topic	Lecture hours
Principles of Cloud Computing	6
Technologies behind Cloud Computing	4
Applications and Programming	8
Cloud Management	2

Detailed schedule available on *Canvas*

Assessments

Assessment	Weightage
Quizzes	15%
Labs	20%
Group Project	35%
Final Test	30%

Final Test

- Held on 14/04/26 (in lecture slot)
- 90 minutes
- **Open book** $\Rightarrow$ no access to Internet

NUS National University of Singapore

AI and Plagiarism I

Use of AI tools **is allowed** $\Rightarrow$ ensure you follow the college guidelines on this and appropriately cite its use

Zero-tolerance for plagiarism

- Students will be reported to the university for disciplinary action for plagiarism/cheating offence
- Resources
  - `https://www.comp.nus.edu.sg/cug/plagiarism/`
  - `https://www.nus.edu.sg/celc/statements-and-e-resources-on-plagiarism/`

NUS National University of Singapore

AI and Plagiarism II

https://www.comp.nus.edu.sg/cug/plagiarism/

> **Plagiarism is generally defined as the practice of taking someone else's work or ideas and passing them off as one's own** *(The New Oxford Dictionary of English).*
>
> *All students share the responsibility of protecting the academic standards and reputation of the University. This responsibility can extend beyond each student's own conduct, and can include reporting incidents of suspected academic dishonesty through the appropriate channels. Students who have reasonable grounds to suspect academic dishonesty should raise their concerns directly to the relevant Head of Department, Dean of Faculty, Registrar, Vice Provost or Provost. The University does not condone plagiarism.*

NUS National University of Singapore

AI and Plagiarism III

https://www.nus.edu.sg/celc/statements-and-e-resources-on-plagiarism/

> *Students should adopt this rule* – **You have the obligation to make clear to the assessor which is your own work, and which is the work of others.** *Otherwise, your assessor is entitled to assume that everything being presented for assessment is being presented as entirely your own work. This is a minimum standard.*
>
> *A student may not knowingly intend to plagiarise, but that should not be used as an excuse for plagiarism.* **Students should seek clarification from their instructors or supervisors if they are unsure whether or not they are plagiarising the work of another person.**

Quizzes

Ensure you keep up with the module

Review key concepts

Each quiz has 10 questions in total

Each quiz takes best out of 3 attempts

Four quizzes in total

Labs

To reinforce what you have learnt in a practical manner

Two labs in total

Knowledge gained helps towards the group project

Can use **any** cloud platform $\rightarrow$ available free tier should be more than enough

NUS National University of Singapore

Group Project

Develop a SaaS product in an area of interest ⇒ working prototype is **required**

Teams of 4 or 5

- **Recommendation**: Have team members from different cohorts

Activity	**Deadline**
Group formation	09/02/26
Preliminary report	09/03/26
Final submissions	19/04/26

Can use **any** cloud platform → available free tier should be more than enough

Intra-group – everyone needs to make their equal contribution

Inter-group – you are **not** to share your work with another group, even if you have discussions

NUS National University of Singapore

SoC Term Project Showcase

Great event to showcase your group project to industry guests – prizes also available ☺

Winner of the **Innovative Open-Source Excellence Award** last semester ☺

Likely date: **15/04/26** from **13:00 – 19:00** (approx.)

Last semester's event: `https://uvents.nus.edu.sg/event/27th-steps`

More details will follow after recess week

Resources

Lots of free online resources here:
`https://github.com/cloudcommunity/Free-Books`

Cloud Computing: Concepts, Technology & Architecture, Thomas Erl, et al., Prentice-Hall, 2013, [chapters 3, 4, 5, 11, 15, 16]

The Datacenter as a Computer – Designing Warehouse-Scale Machines, 3^{rd} Edition, Morgan & Claypool Publishers, 2019 (available online) [chapters 1, 2, 3, 4, 6]

NUS National University of Singapore

Previously on CS5224 . . .

Content is too technical or not technical enough

- More support available – SE toolbox from CS3219
- Group project provides enough breadth – product must have a very good business case and an excellent implementation and presentation
- Lots of resources (including videos) for students to explore, as necessary

Specification and expectation for group project was unclear

- Will be improved with feedback from previous students and help from TAs

No peer assessment in the group project

- Peer assessment is a requirement in the group project

Infrastructure as Code requested

- New topic this semester ☺

NUS National University of Singapore

What is Cloud Computing?

Gartner Report (`https://www.gartner.com/en/documents/3947472`)

- "... a style of computing in which scalable and elastic IT-enabled capabilities are delivered as a service using Internet technologies."

Forrester Research (`https://www.forrester.com/blogs/09-10-02-assessing_the_maturity_of_cloud_computing_services/`)

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

NIST 2011 (`https://csrc.nist.gov/pubs/sp/800/145/final`)

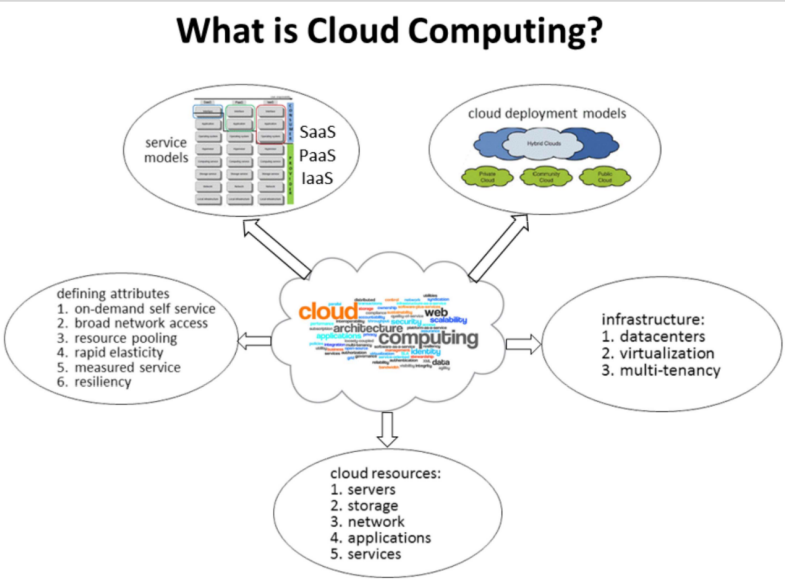
- "Cloud computing is a model for enabling *ubiquitous*, *convenient*, on-demand *network access* to a shared pool of configurable computing resources (e.g., *networks*, *servers*, *storage*, *applications*, *and services*) that can be rapidly provisioned and released with minimal management effort **or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

on-demand service + elastic resource

Program is an Internet (cloud) **service**

and

platform are datacenters


What is Cloud Computing?
service models
SaaS
PaaS
IaaS
cloud deployment models
Hybrid Clouds
Private Cloud
Community Cloud
Public Cloud
defining attributes
1. on-demand self service
2. broad network access
3. resource pooling
4. rapid elasticity
5. measured service
6. resiliency
cloud
computing
architecture
web
scalability
security
infrastructure:
1. datacenters
2. virtualization
3. multi-tenancy
cloud resources:
1. servers
2. storage
3. network
4. applications
5. services

History I

1961 – "computing may someday be organized as a *public utility*", John McCarthy (1927 – 2011)

1996 – "cloud computing" was coined at Compaq Computer (MIT Technology Review)

1999 – **Salesforce.com** pioneered the concept of delivering enterprise applications via a simple website

Jul 2002 – **Amazon Web Services (AWS)** provided a suite of cloud-based services including computation, storage, and even human intelligence through *Amazon Mechanical Turk*

May 2006 – **Amazon Simple Storage Service (S3)**, a "pay-per-use" storage

History II

Aug 2006 – **Amazon Elastic Compute Cloud (EC2)**, a commercial web service (IaaS) that allows small companies and individuals to rent computers to run their own applications

Apr 2008 – **Google App Engine**, Google's PaaS, Bigtable and GFS for storage, MapReduce, etc.

Nov 2009 – **Microsoft Windows Azure**, an operating environment "designed to manage extremely large pools of computational resources". Customers run Windows-based applications over the Internet using Microsoft's data centers, while Azure organizes resources and handles spikes in demand

2011 onwards – many many many cloud providers!

Why Cloud Computing?

Technology (cloud-enabled platforms and services) for current/future innovations and disruptions

Next wave of cloud disruption delivers advanced capability around AI, IoT, Blockchain, among others!

Reduces business cost

- Improves match between elastic resource demand and elastic computing resource

Improves **availability** and **elasticity**

Reduces cost

But . . .

What caused the AWS outage - and why did it make the internet fall apart?

21 October 2025

Share Save

Zoe Kleinman

Technology editor

https://www.bbc.com/news/articles/cev1en9077ro

Cloudflare apologises for outage which took down X and ChatGPT

18 November 2025

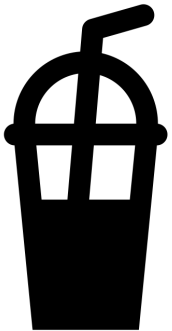
Share Save

Liv McMahon

Technology reporter

https://www.bbc.com/news/articles/c629pny4gl7o

Take a Break



NUS National University of Singapore

To help deliver **computing as-a-service**

- Parallel and Distributed Computing
- Programming Models
- Utility Computing
- Virtualization
- Web Technologies
- Storage and Network Technologies
- Internet

Key Terms

1. Elastic resource
2. Availability
3. Capacity planning (resource provisioning)
4. Scaling (horizontal, vertical)
5. Cloud-based IT resources
6. Cloud service
7. Trust boundary

NUS National University of Singapore

Key Business Drivers!

Capacity planning

Cost reduction

Organisation agility

3. Capacity Planning

Process of determining and fulfilling future demands of an organisation's IT resources, products and services

Challenges: usage/demand fluctuations, peak usage, cost of resource provisioning, ...

Strategies

- Lead strategy: add capacity in anticipation of demand
- Lag strategy: add capacity when resources reach its full capacity
- Match strategy: add capacity in small increments as demand increases

Should ideally avoid under-provisioning and over-provisioning

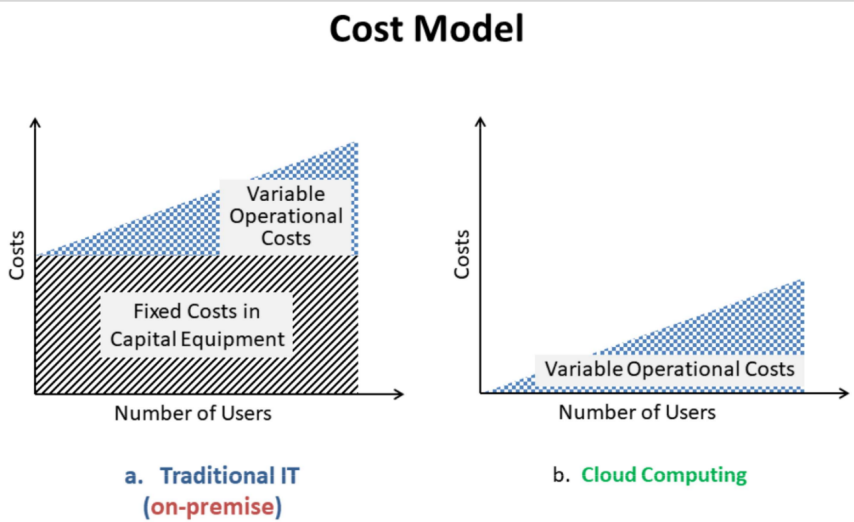
Cost Reduction

Difficult to directly align IT costs and business performance

On-premise system – IT department is a **big cost**

- Up-front investment costs
- Operational costs: technical staff (manpower), utility bills (power and cooling), security and access control to protect infrastructure, admin/account cost (software licences, etc.)

Cloud computing typically offers cost efficiency at scale


Cost Model
Costs
Variable Operational Costs
Fixed Costs in Capital Equipment
Number of Users
a. Traditional IT (on-premise)
Costs
Variable Operational Costs
Number of Users
b. Cloud Computing

Hwang, P199

Organisational Agility

A measure of an organisation's responsiveness to change

On-premise: Up-front investment and infrastructure ownership costs may be prohibitive

Cloud

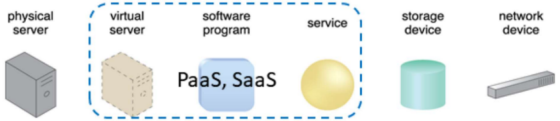
- Elastic IT resources to respond to business cycles beyond what was previously predicted or planned
- Software fixes/updates: Update datacenters vs millions of clients
- Datacenter allows faster introduction of new hardware innovations

Key Terms

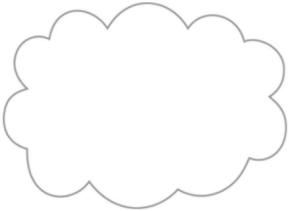
1. Elastic resource
2. Availability
3. Capacity planning (resource provisioning)
4. Cloud-based IT resources
5. Scaling (horizontal, vertical)
6. Cloud service
7. Trust boundary

NUS National University of Singapore

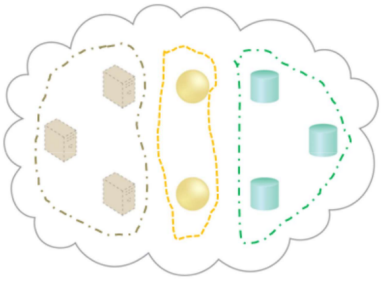
More Concepts and Terminology



Examples of common Cloud-based IT resources and their corresponding symbols



Symbol denotes the Boundary of a Cloud Environment



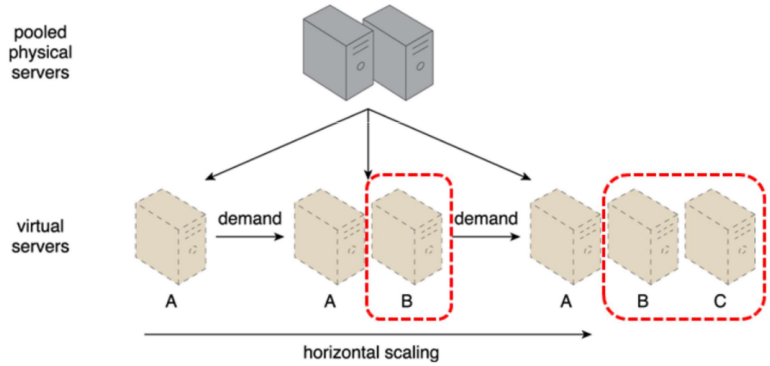
Example: A cloud hosting 8 IT resources: 3 virtual servers, 2 **cloud services**, and 3 storage devices

Ability of the IT resource to handle increased or decreased usage demands

Horizontal Scaling (out or in) – add same resource type

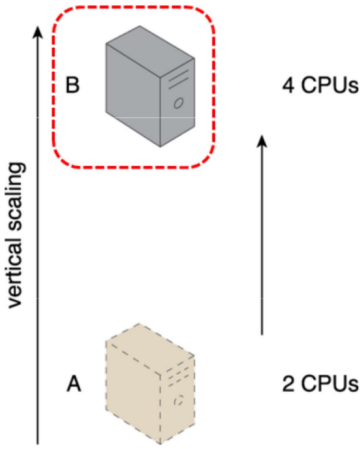
Vertical Scaling (up or down) – replace resource with higher or lower capacity or add resources to a single node

Horizontal Scaling



An IT resource (Virtual Server A) is **scaled out** by adding more of the same resource (Virtual Servers B and C)

Vertical Scaling



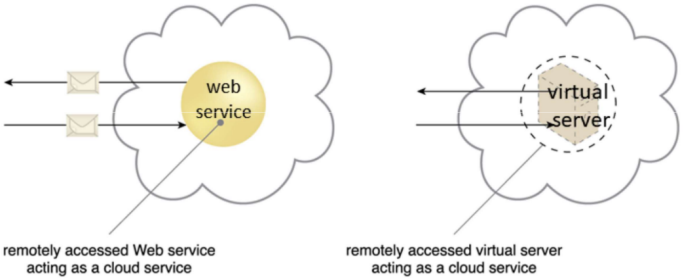
An IT resource (Virtual Server A) with 2 CPUs is **scaled up** by replacing it with a more powerful resource (Server B with 4 CPUs)

Scaling Comparison

	Horizontal	Vertical
Cost	Less expensive using commodity hardware	More expensive using specialized hardware
Availability	Resources instantly available	Resources normally instantly available
Ease of setup	Resource replication and automated scaling	May need additional setup
Hardware capacity	Not limited by hardware capacity	Limited by maximum hardware capacity

6. Cloud Service

Any IT resource made remotely accessible via a cloud



Multitude of service usage models – next lecture

- IaaS (Infrastructure-as-a-Service)
- PaaS (Platform-as-a-Service)
- SaaS (Software-as-a-Service)

NUS National University of Singapore

Technical Challenges

Software Development – different cloud platforms and services across cloud providers

Tools are continuously evolving

Moving large data is still expensive

Security

Internet dependence

Quality of service

Energy costs

...

NUS National University of Singapore

Non-Technical Challenges

Increased security vulnerabilities

Reduced operational governance control

Privacy/legal issues: data localization, multi-regional compliances

Vendor lock-in

Service level agreement

...

Increased Security Vulnerabilities

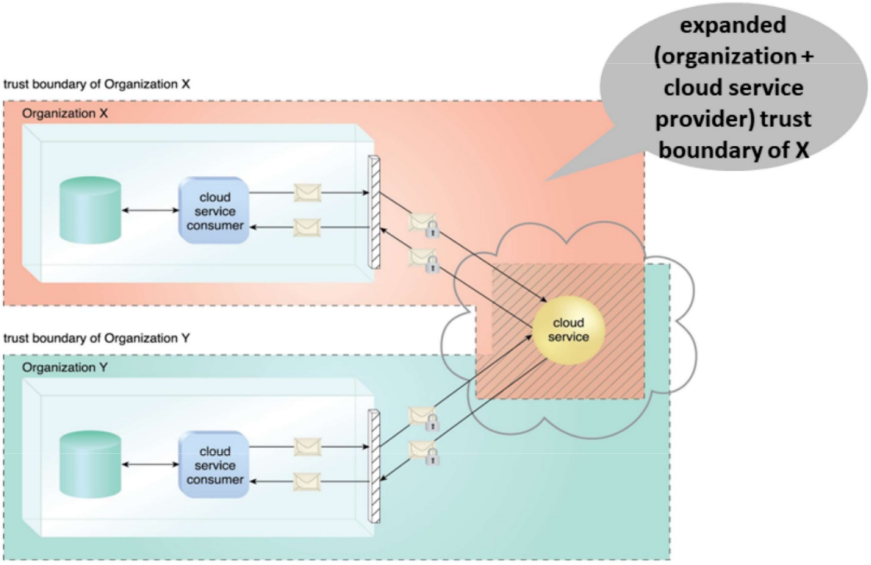
Responsibility over data security becomes shared with cloud providers

Expansion of trust boundary introduces new vulnerabilities

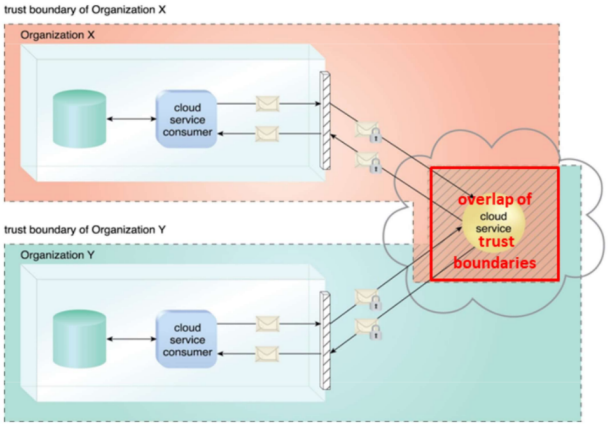
Shared IT resources across different cloud consumers introduce **overlapping trust boundaries** – opportunities for malicious cloud consumers to steal or damage data

NUS National University of Singapore

Expanded Trust Boundaries



Overlapping Trust Boundaries



Shaded area with diagonal lines indicates the overlap of trust boundaries between Organisation X and Organisation Y

NUS National University of Singapore

Reduced Operational Governance Control

For cloud consumers – different levels of IT resource governance between on-premise and cloud

Unreliable cloud provider may not maintain/meet SLA guarantees

Physical distance between provider and consumer introduces additional latency and bandwidth constraints

Privacy/Legal Issues

Data centers usually set up in affordable/convenient locations

Data Localization/Residency – Industry or government regulations on data privacy or storage policies (Very important!)

- E.g. Data belonging to Singapore citizens to be kept within Singapore

Legal issues on accessibility and disclosure of data, e.g., laws require data to be disclosed appropriately

Tension between personal rights (privacy) and society, e.g. location, proximity data, etc.

Note of Caution

The Rise Of Cloud Repatriation: Why Companies Are Bringing Data In-House

Forbes | Technology Council

Marcin Zgola Forbes Councils Member

Forbes Technology Council COUNCIL POST | Membership (Fee-Based)

https://www.forbes.com/councils/forbestechcouncil/2023/04/18/the-rise-of-cloud-repatriation-why-companies-are-bringing-data-in-house/

NUS National University of Singapore

Next Week

Things get serious!

Concepts and Models

References

Chapter 3, Cloud Computing: Concepts, Technology and Architecture, Thomas Erl, Zaigham Mahmood and Ricardo Puttini, Prentice-Hall, 2013.

Chapter 1, The Datacenter as a Computer – Designing Warehouse-Scale Machines, 3^{rd} Edition, Morgan and Claypool Publishers, 2019 (available online)

Above the Clouds: A Berkeley View of Cloud Computing, 2009

The NIST Definition of Cloud Computing, NIST Report, 2011

Continuous Feedback

https://forms.office.com/r/N3gWNzXVPZ

NUS National University of Singapore