THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

Cloud Computing (INFS3208)

Lecture 1: Introduction to Cloud Computing

Lecturer: Dr Sen Wang

School of Electrical Engineering and Computer Science (EECS)

Faculty of Engineering, Architecture and Information Technology

The University of Queensland

Acknowledgement of **Country**

The University of Queensland (UQ) acknowledges the Traditional Owners and their custodianship of the lands on which we meet.

We pay our respects to their Ancestors and their descendants, who continue cultural and spiritual connections to the Country.

We recognise their valuable contributions to Australian and global society.

Outline

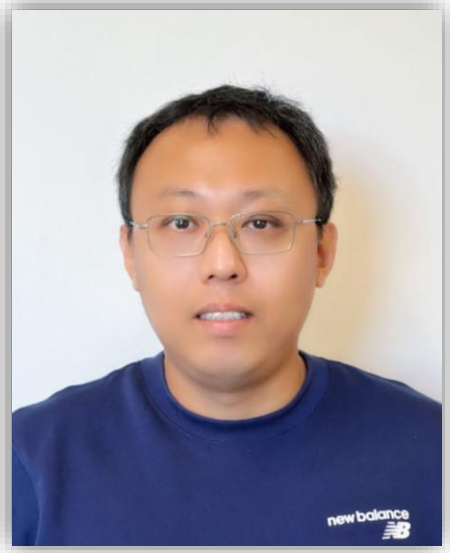
I. Course Introduction

- Teaching Team
- Course Website & Blackboard
- Course Overview – Lectures, Tutorials & Practicals
- Course Assessment & Policy
- Learning Materials

II. Introduction to Cloud Computing

- History & Definitions
- Business Drivers
- Technology Innovations
- Cloud Characteristics
- Cloud Delivery Models & Cloud Deploy Models
- Cloud-enabling technologies: Broadband Networks and Internet Architecture, Virtualisation Technology (VT), Data Centre Technology, Web Technology, and Multitenant Technology
- Goals and Benefits
- Risks and Challenges
- Cloud-based Applications in the World

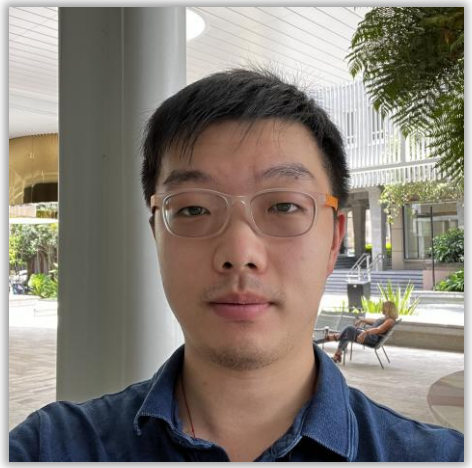
Teaching Team – Course Coordinator & Lecturer



Dr. Sen Wang

- Associate Professor in CS (**INFS3202/7202**, **INFS3208**),
- ARC DECRA Fellowship (2020-2022)
- Program Convenor of MDataSc
- Homepage: http://csenw.github.io/
- Email: sen.wang@uq.edu.au
- Zoom: https://uqz.zoom.us/j/7751582216
- Office: Room 649, Building 78, St. Lucia Campus
  - Office Hour: By email appointment. Make an appointment before you come to my office in order to guarantee you can find me

Teaching Team – Lecturer (Heming)



Dr. Heming Du

- Associate Lecturer in Data Science (**INFS3208**)
- Research Interests: Computer Vision, Reinforcement Learning
- Email: heming.du@uq.edu.au
- Zoom: https://uqz.zoom.us/j/86247260673
- Office: Room 220, Building 78, St. Lucia Campus
  - Office Hour: By email appointment. Make an appointment before you come to my office in order to guarantee you can find me

Teaching Team – Tutors

Name	Office	Email
Mr. Jason Liang	78-624	xurong.liang@uq.edu.au
Ms. Lihong Wang	-	lihong.wang@uq.edu.au
Mr. Manjiang Yu	78-438	manjiang.yu@uq.edu.au
Ms. Vy Ho	-	vukhanhvy.ho@uq.edu.au
Mr. Fengqi Yu	78-438	fengqi.yu@uq.edu.au
Mr. John Chan	-	wilkinsonjohn.chan@student.uq.edu.au
Mr. Jiexu Xie	-	jiexu.xie@student.uq.edu.au

Information of tutors and labs can be found under "Course Staff" in Blackboard as well.

Lihong Wang

Manjiang Yu

Jason Liang

Vy Ho

Fengqi Yu

John Chan

Jiexu Xie

Teaching Team – Tutors

BYOD for Applied Classes (aka Tutorials) and Practicals are Welcome!

	Monday	Tuesday	Wendesday	Thursday	Friday	
08:00		App 3: 78-420 Lihong Wang		Prac 3: 78-116 Vu Ho		08:00
09:00		App 4: 78-420 Lihong Wang	App 12: 78-420 Manjiang Yu	Prac 4: 78-116 Vu Ho		09:00
10:00			App 13: 78-420 Manjiang Yu		Prac 10: 78-116 Fengqi Yu	10:00
11:00		App 5: 78-420 Lihong Wang	App 14: 78-420 Vu Ho		Prac 11: 78-116 Fengqi Yu	11:00
12:00		App 6: 78-420 Lihong Wang		Prac 5: 78-116 Manjiang Yu		12:00
13:00				Prac 6: 78-116 Manjiang Yu		13:00
14:00	50-T203	App 7: 78-420 Jiexu Xie	Prac 1: 78-116 Jason Liang		Prac 12: 78-116 John Chan	14:00
15:00		App 8: 78-420 Jiexu Xie	Prac 2: 78-116 Jason Liang		Prac 13: 78-116 John Chan	15:00
16:00	App 1: 78-420 John Chan	App 9: 78-420 Fengqi Yu		Prac 7: 78-116 Jiexu Xie	Prac 14: 78-116 John Chan	16:00
17:00	App 2: 78-420 John Chan	App 10: 78-420 Fengqi Yu		Prac 8: 78-116 Jiexu Xie		17:00
18:00		App 11: 78-420 Fengqi Yu		Prac 9: 78-116 Vu Ho		18:00

Practical

Applied Class

Information of tutors and labs can be found under “Course Staff” in Blackboard as well.

Laptop Loans

Students enrolled in the current semester can borrow a laptop from the Library using your UQ ID card. We offer loans of:

- **24 hours** at laptop lockers in Central Library and JK Murray Library
- **28 days** to undergraduate and coursework postgraduate students who do not have a computer or laptop you can use for your studies. Borrow a laptop from an AskUs service point at the **Dorothy Hill Engineering and Sciences** or Gatton libraries during AskUs desk hours.

By borrowing from the Library, you agree to comply with our *borrowing rules*.

For more details, please refer to the web page:
https://web.library.uq.edu.au/library-services/it/laptop-loans

Laptop lockers at Central Library.

Course Website

Address: https://learn.uq.edu.au/

You can find the following information on the web:

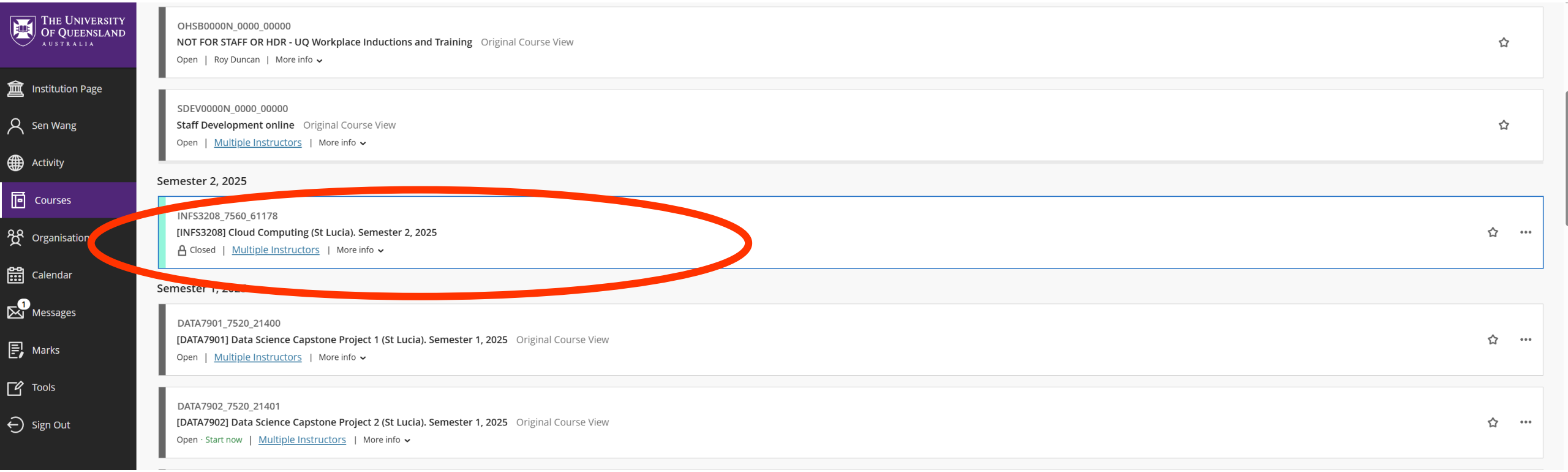
- Course Profile (ECP)
- Course Staff
- Announcements (latest news)

- Learning Resources (Teaching materials - Lecture notes, tutorials, and practicals)
- Assessment (assignments and project information, as well as final exam, and lodgement)
- Ed Discussion
- Discussion Board (not in use)
- My Grades
- Mock Exam (TBA)

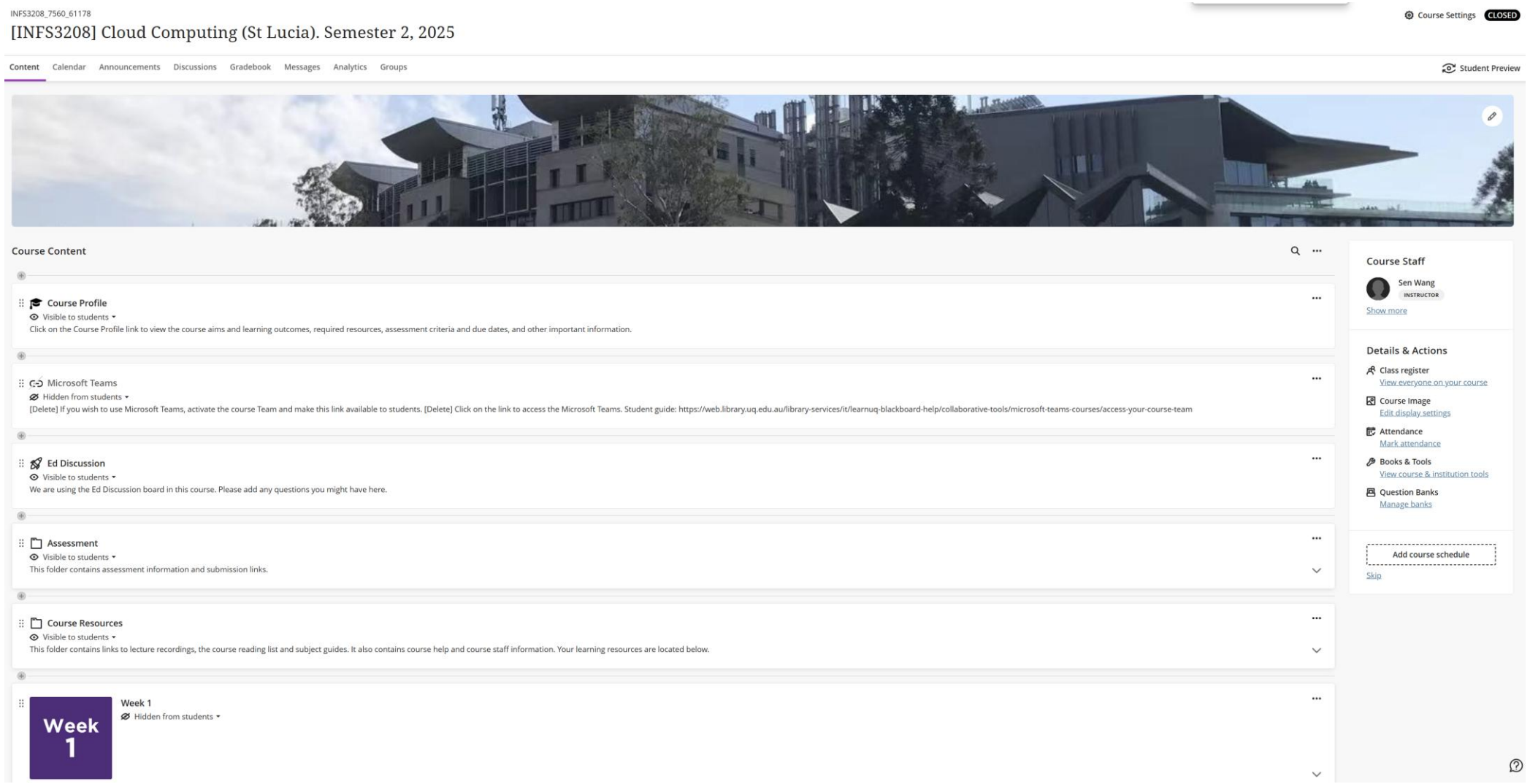
Note: *All students enrolled in this course will be automatically enrolled in Ed Discussion Board after they click on the link and sign in*.

PLEASE CHECK THE WEBSITE / BLACKBORAD REGULARLY!

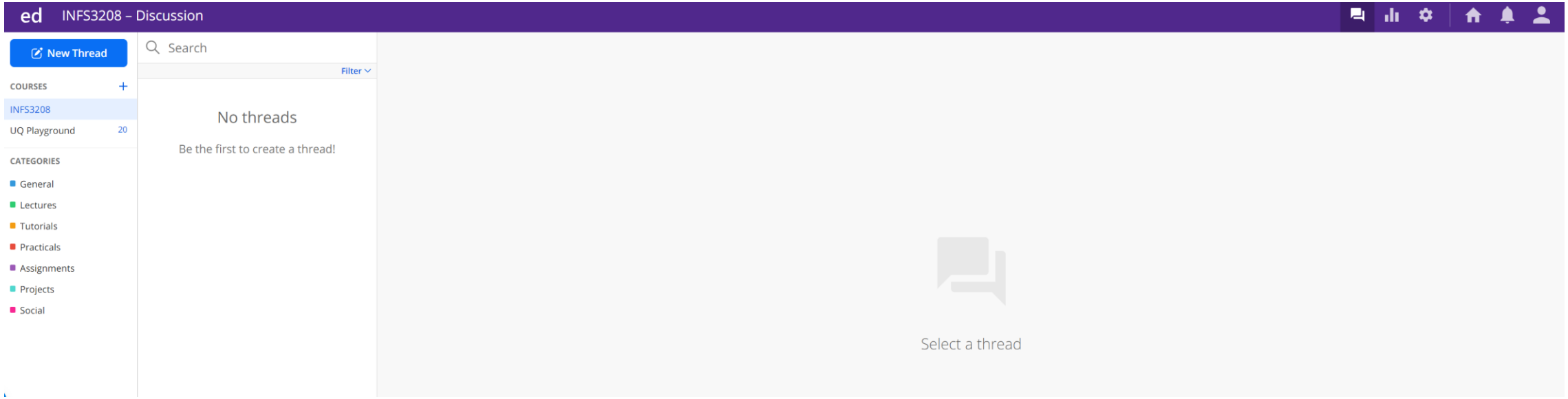
Blackboard – Course Website (New Ultra)



Blackboard – Course Website



Blackboard – Ed Discussion



Course Delivery Mode

On-campus mode only

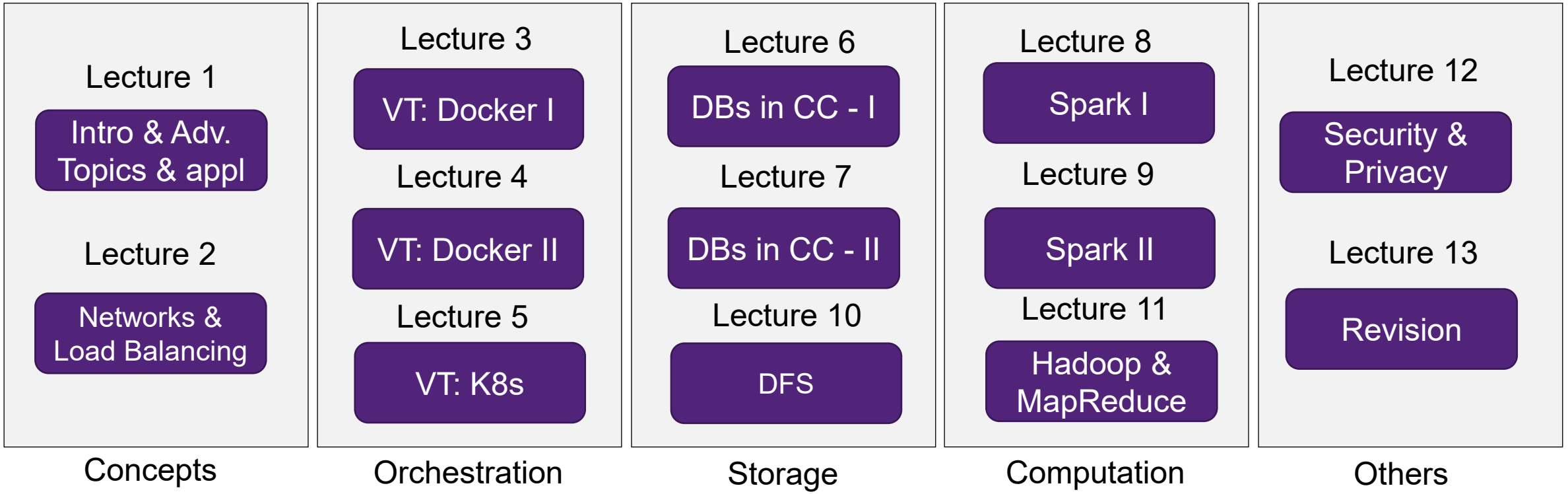
- On-campus lectures (recorded via Echo360 and available in BB);
- In-person Applied Class (aka Tutorial) and Practical sessions on campus
  - Different rooms for Practical (78-116) and Applied Class (78-420)
  - **No recordings for APPs and PRAs**

Weekly consultation and discussion

- Ed Discussion (Administrated by Jason & Fengqi)
- Check and reply to questions on Ed Discussion on a daily basis.

Course Overview – Lectures

This course includes 13 lectures and 10 tutorial/practical sessions



Basic Programming and Linux experiences are required!

Relationship between INFS3208 and INFS3202/7202

INFS3202/7202 Web Information Systems (WIS) and INFS3208 Cloud Computing are distinct yet interconnected.

- **Deployment Environment**: INFS3202/7202 aims to develop scripts that run on a server to generate dynamic web content before sending it to a client's web browser. INFS3208 teaches you how to deploy/host these scripts on cloud platforms considering scalability, reliability, and cost-effectiveness.
- **Microservices and Containerisation**: INFS3202/7202 teaches Three-layer (MVC) modelling for development, which separates the code logistically. INFS3208 introduces another architecture – microservices architecture widely used in the industry. This architecture enhances both development efficiency and application scalability.
- **Distributed Data Storage**: INFS3202/7202 involves interaction with databases (SQL or NoSQL) to store and retrieve data in MVC. INFS3208 focuses on big data storage solutions in the cloud environment.

In the roadmap of Backend Developers, Cloud Computing techniques are **core skills**.

Relationship between INFS3208 and DATA7201

DATA7201 Data Analytics at Scale and INFS3208 Cloud Computing offer big data analytic solutions with different focuses and perspectives.

- **Audience is different**: DATA7201, as one of the core courses in MDataSc, is only available to MDataSc students, while INFS3208 can be enrolled by any students with Python and SQL backgrounds.
- **Technical focus is different**: DATA7201 equips students with the skills to probe various types of data, often in substantial volumes, to gain insights through machine learning or data mining methods. INFS3208 technically focuses on big data analytical frameworks (e.g. MapReduce and Spark), aiming to solve data engineering issues (large-scale data queries and learning), hence a background in machine learning is not essential.
- **Distributed Data Storage**: DATA7201 introduces distributed file systems (HDFS and GFS). INFS3208 detailed the related topic and introduce NoSQL databases as an alternative data storage solution in the cloud context.

INFS3208 gives students a broader view of the infrastructure side of things.

Course Overview – App/Prac Sessions

Applied class and practical sessions start from Week 2

- Held in different rooms
- Seats and desktops are guaranteed to be available for enrolled students
- In extreme circumstances, if you want to change the session, please contact session tutors in the first place

Applied Class will focus on the contents delivered in lectures

- Concept understanding & question answering
- Assignment discussion & project discussion

Practical will be focusing on techniques introduced in lectures

- Google Faculty Grant is NO LONGER AVAILABLE
  - US$ 50 for each student (use it wisely – pricing calculator).
  - Instead, free trial US$300 for registered users (**recommended, but not compulsory**)
  - Use VPN (UQ VPN guideline) or VMs (Oracle VirtualBox)
- GCP compute engine – VM/Kubernetes
- Docker technologies – dockerfile/docker compose, KGE on GCP
- Spark programming – scala/RDD programming, DataProc on GCP
- DFS & DBs in Cloud
- Project implementation

Assessment

One online non-marked progress check (MCQs) will be available on BB during week 4 (feedback before Consensus Date).

30% Three Individual Programming Tasks in Programming Assignment (10% each)

- On-line submission via Blackboard Turn-it-In**.**
- No extension to anyone except to the medical or other unexpected serious situations.
- Submit the whole system file online in a single compressed file.

20% Individual Project (5% report & 15% implementation)

- Demonstration in the Lab (Tut & Prac) to tutors for evaluation after the deadline.
- Attendance is **compulsory.**
- No extension to anyone except to the medical or other unexpected serious situations.
- Submit the whole system file online in a single compressed file (one for each group).

50% Final Exam

- on-campus centrally administrated
- Double pass (at least 25 out of 50).

Best Project Demos

NOTE: These assessments evaluate students' abilities, skills and knowledge without the aid of Artificial Intelligence (AI). Students are advised that the use of AI technologies to develop responses is strictly prohibited and may constitute student misconduct under the Student Code of Conduct.

https://elearning.uq.edu.au/guides/group-peer-assessment/paf-formula-and-moderation-overview

Assessment Dues

Assessment summary

Category	Assessment task	Weight	Due date
Computer Code	Individual Programming Tasks	30%	Programming Task 1 23/08/2024 3:00 pm Programming Task 2 13/09/2024 3:00 pm Programming Task 3 11/10/2024 3:00 pm
Project	Individual Project Implementation	20%	18/10/2024 3:00 pm
Examination	Final Examination ⚠ Hurdle	50%	End of Semester Exam Period 2/11/2024 - 16/11/2024

Final Exam (50%)

Closed-book exam – 120 minutes on-campus (internal) & online invigilated via ProctorU (external)

Cover all materials in this course

Mainly on concepts and understanding

The final grade is determined by the total marks from the four assessments and the final exam

You must pass 50% of the Final Exam

Graduate Students will take different Final Exam questions

- Final mark — Grade

Final mark	Grade
85-100	7
75-84	6
65-74	5
50-64	4
45-49	3
20-44	2
0-19	1




INFS3208

Late Submission and Extension Policy

Deferral or extension

You may be able to **apply for an extension**.

The maximum extension allowed is 7 days. Extensions are given in multiples of 24 hours.

Marked assignments with feedback and/or detailed solutions with feedback will be released to students within 14-21 days, where the earlier time frame applies if there are no extensions.

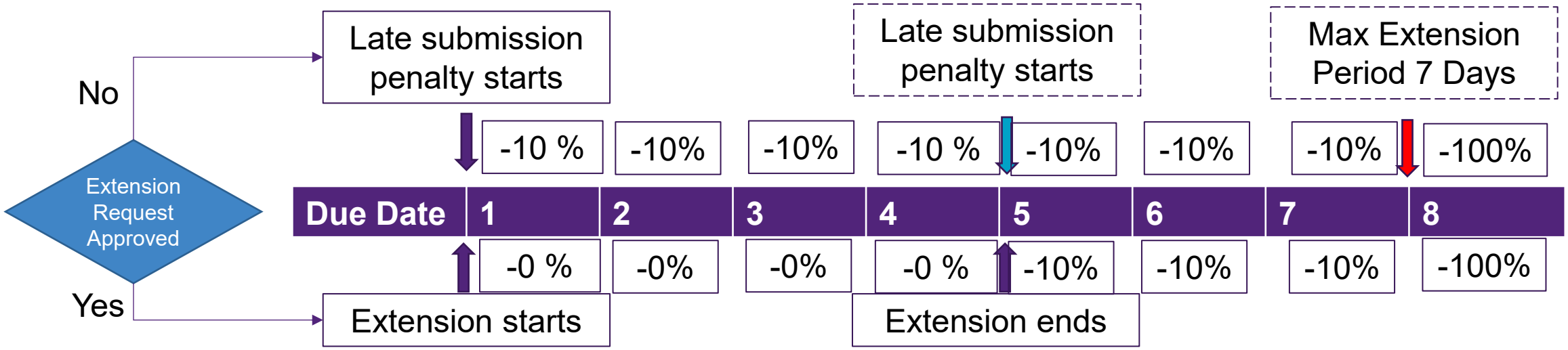
Late submission

A **penalty** of 10% of the maximum possible mark will be deducted per 24 hours from time submission is due for up to 7 days. After 7 days, you will receive a mark of 0.

Late Submission and Extension Policy

Late Submission After an Extension:

- If you have an extension, then late submission may not be possible. No submissions will be accepted after the total maximum extension length (7 calendar days) from the original due date. For example, if you have a 4-day extension, then per-day late penalties will apply after 4 days, but only up until 7 days from the original due date. After that time, no marks will be awarded for the item. If you receive an extension of the maximum length, then you must submit it by that time. No further extension is possible. An immediate 100% late penalty applies, i.e., late submission will receive zero marks.
  - For example, if you have a 4-day extension, then the **100% late penalty** will apply after 4 days.



Plagiarism & Generative AI

- Plagiarism is the act of misrepresenting as one's own original work the ideas, interpretations, words or creative works of another. These include published and unpublished documents, designs, music, sounds, images, photographs, computer codes and ideas gained through working in a group. These ideas, interpretations, words or works may be found in print and/or electronic media.
- Students are encouraged to read the UQ Student Integrity and Misconduct policy (http://ppl.app.uq.edu.au/content/3.60.04-student-integrity-and-misconduct) which makes a comprehensive statement about the University's approach to plagiarism, including the approved use of plagiarism detection software, the consequences of plagiarism and the principles associated with preventing plagiarism.
- All submitted works will be tested with electronic plagiarism check (iThenticate).

Use of AI

Individual Programming Tasks/Individual Project Implementation: This assessment task evaluates students' abilities, skills and knowledge without the aid of generative Artificial Intelligence (AI) or Machine Translation (MT). Students are advised that the use of AI technologies to develop responses is strictly prohibited and may constitute student misconduct under the Student Code of Conduct.

Academic integrity is:

Acting with the values of honesty, trust, fairness, respect and responsibility in learning, teaching and research.

(Universities Australia, 2017)

Academic integrity = Professional integrity = Personal integrity

Benefits of academic integrity

High standards of academic integrity protect you, the University and the community:

- You have the pride and confidence that comes with knowing you have developed your knowledge and learnt new skills
- You understand how new knowledge is created and how to apply that knowledge to your studies and future career
- You model the practices of integrity we want for society
- The community has faith in the value of a UQ qualification
- Your employer, your clients and your patients know you are knowledgeable and skilled.

UQ takes academic misconduct seriously

Incidences

Jan 2017–Jan 2019: over 1,400 cases of academic misconduct.

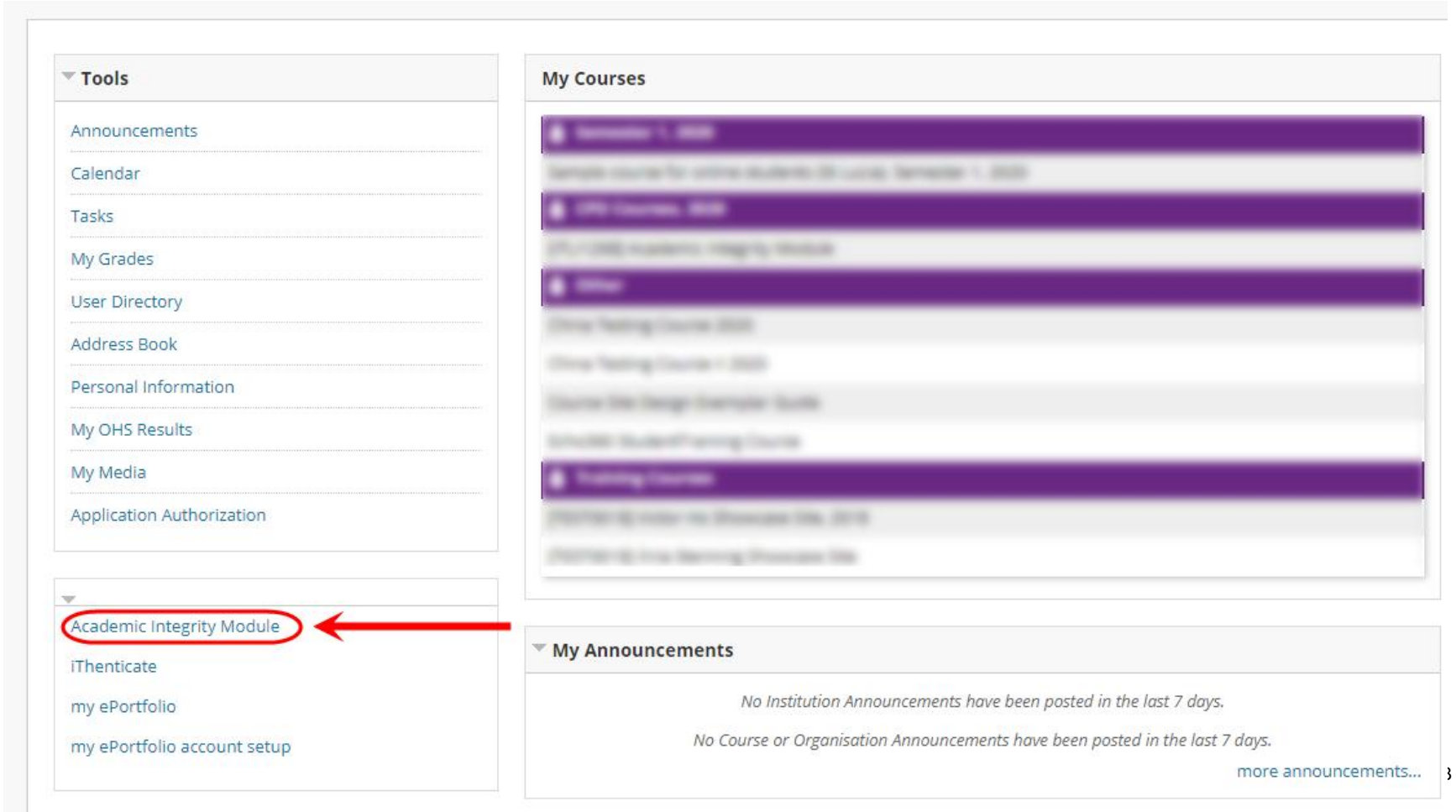
Penalties

Penalties include:

- re-submission of an assessment or automatic failure of a course (lower-end offences)
- suspension or expulsion (more serious or repeated misconduct).

Access the Academic Integrity module

1. Go to learn.uq.edu.au
2. Click on the Academic Integrity Module link



Uh, Mom?
Remember that paper
you wrote for me about nuclear
reactors? What does it mean
when the big, red “meltdown”
light is flashing?

AI Awareness Poster, Brigham Young University
https://live-academicintegrity.pantheonsite.io/wp-content/uploads/2017/12/462c19_1dc7602ba4ec4f4586d803e7c43904cc_mv2-1.gif

Outline

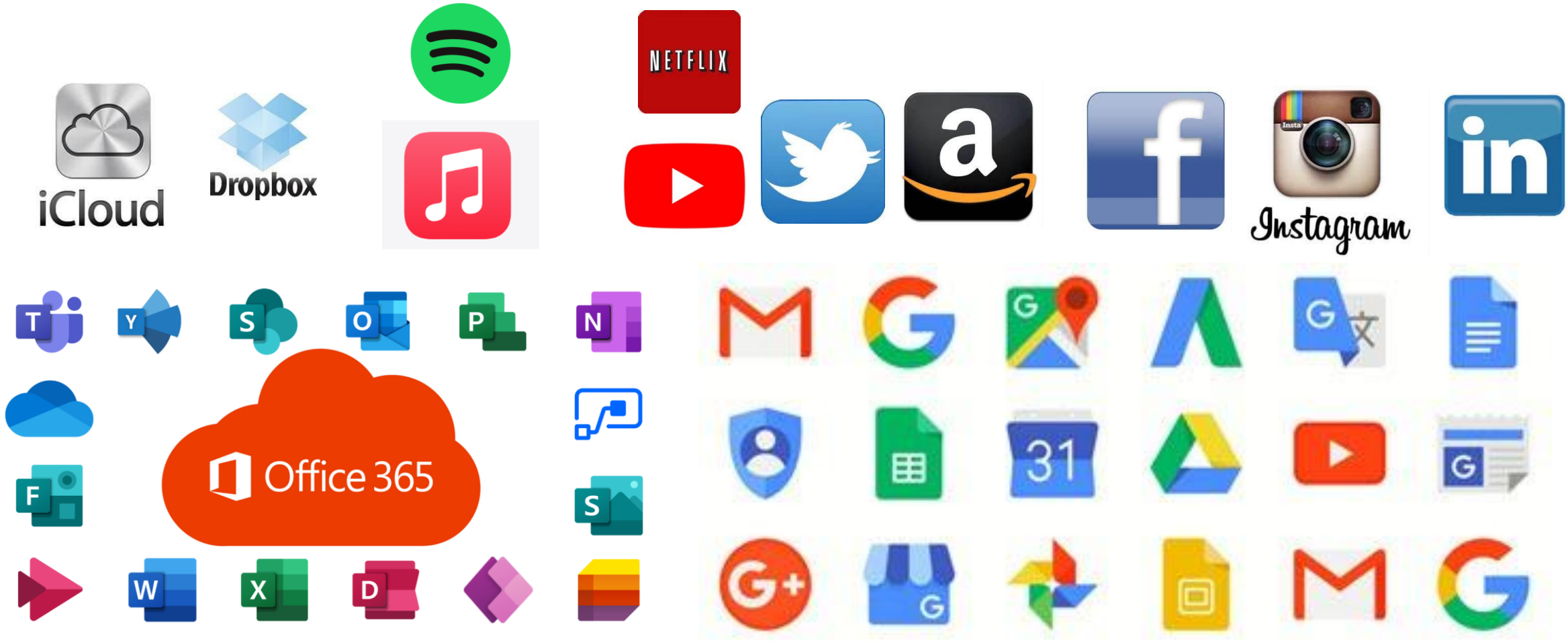
I. Course Introduction

- Teaching Team
- Course Website & Blackboard
- Course Overview – Lectures, Tutorials & Practicals
- Course Assessment & Policy
- Learning Materials

II. Introduction to Cloud Computing

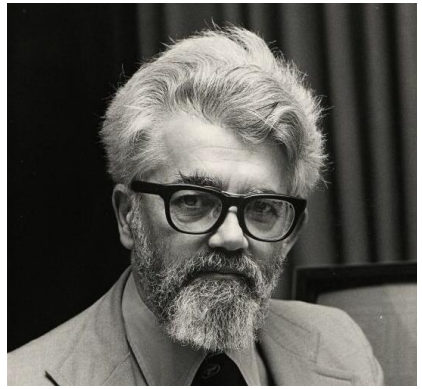
- History & Definitions
- Business Drivers
- Technology Innovations
- Cloud Characteristics
- Cloud Delivery Models & Cloud Deploy Models
- Cloud-enabling technologies: Broadband Networks and Internet Architecture, Virtualisation Technology (VT), Data Centre Technology, Web Technology, and Multitenant Technology
- Goals and Benefits
- Risks and Challenges
- Cloud-based Applications in the World

Do You Use the Cloud?



History of Cloud Computing

"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 in **1961**



John McCarthy *
(1927 –2011)

Mid-1990s

YAHOO Google YouTube

2002 amazon web services™ aws

Provisioned storage
Computing resources
Business functionality

Remotely!!

2006 Amazon EC2

2008 Google App Engine

* https://en.wikipedia.org/wiki/John_McCarthy_(computer_scientist)

Definitions of Cloud Computing

Def 1: *"...a style of computing in which scalable and elastic IT-enabled capabilities are delivered as a service to external customers using Internet technologies." – by* ***Gartner Inc.*** [1]

Def 2: *"...a standardized IT capability (services, software, or infrastructure) delivered via Internet technologies in a pay-per-use, self-service way." – by* ***Forrester Research.*** [2]

Def 3: *"…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." – by* ***National Institute of Standards and Technology*** *(****NIST****).* [3]

Def 4: *"Cloud computing is a specialized form of distributed computing that introduces utilization models for remotely provisioning scalable and measured resources."* [4]

1. https://en.wikipedia.org/wiki/Gartner
2. https://en.wikipedia.org/wiki/Forrester_Research
3. https://en.wikipedia.org/wiki/National_Institute_of_Standards_and_Technology
4. Cloud Computing: Concepts, Technology & Architecture (The Prentice Hall Service Technology Series from Thomas Erl) 1st Edition

Outline

I. Course Introduction

- Teaching Team
- Course Website & Blackboard
- Course Overview – Lectures, Tutorials & Practicals
- Course Assessment & Policy
- Learning Materials

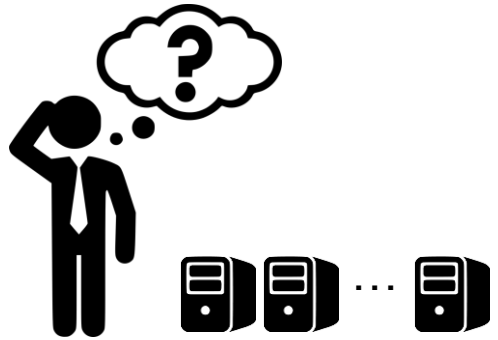
II. Introduction to Cloud Computing

- History & Definitions
- Business Drivers
- Technology Innovations
- Cloud Characteristics
- Cloud Delivery Models & Cloud Deploy Models
- Cloud-enabling technologies: Broadband Networks and Internet Architecture, Virtualisation Technology (VT), Data Centre Technology, Web Technology, and Multitenant Technology
- Goals and Benefits
- Risks and Challenges
- Cloud-based Applications in the World

Business Drivers of Cloud Computing – Capacity Planning

Capacity Planning (CP)

- CP is the process of determining and fulfilling future demands of an organization's IT resources, products, and services.
- discrepancy between the capacity of an IT resource and its demand
  - inefficient (over-provisioning). E.g. 100 TB vs 1TB (real demand)
  - unable to fulfil user needs (under-provisioning). E.g. 1 TB vs 100 TB (real demand)
- CP aims to minimize this discrepancy to achieve predictable efficiency and performance.
- 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 resource capacity in small increments as demand increases.
- CP can be challenging because it requires estimating usage load fluctuations.
- There is a constant need to balance peak usage requirements without unnecessary over-expenditure on infrastructure.



Business Drivers of Cloud Computing – Cost Reduction

Cost Reduction (IT resources)

- A direct alignment between IT costs and business performance can be difficult to maintain.
- Two costs need to be accounted for:

1. the cost of new infrastructure
   - technical personnel required to keep the environment operational
   - upgrades and patches that introduce additional testing and deployment cycles
   - utility bills and capital expense investments for power and cooling
   - security and access control measures that need to be maintained and enforced to protect infrastructure resources
   - administrative and accounts staff that may be required to keep track of licenses and support arrangements
2. the cost of its ongoing ownership.
   - can encompass burdensome responsibilities that impose compound impacts on corporate budgets e.g. an IT department, software license, etc.

Business Drivers of Cloud Computing – Organisational Agility

Organisational Agility:

- the measure of an organization's responsiveness to changes.
- Organisations need to respond to business change by scaling its IT resources.
- Organisations need more reliable IT resources than before.
  - Runtime exceptions and outages could happen.
  - lack of reliability controls and responsiveness to consumers
  - a business' overall continuity is threatened.
- Making organisations more responsive to changes become very essential
  - Keep up with market demands
  - Be competitive against peers
  - Fulfil strategic business goals

Outline

I. Course Introduction

- Teaching Team
- Course Website & Blackboard
- Course Overview – Lectures, Tutorials & Practicals
- Course Assessment & Policy
- Learning Materials

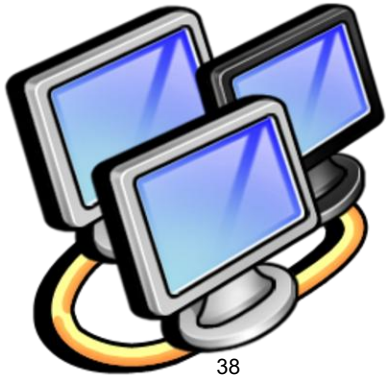
II. Introduction to Cloud Computing

- History & Definitions
- Business Drivers
- Technology Innovations
- Cloud Characteristics
- Cloud Delivery Models & Cloud Deploy Models
- Cloud-enabling technologies: Broadband Networks and Internet Architecture, Virtualisation Technology (VT), Data Centre Technology, Web Technology, and Multitenant Technology
- Goals and Benefits
- Risks and Challenges
- Cloud-based Applications in the World

Technology Innovations: Clustering

Clustering:

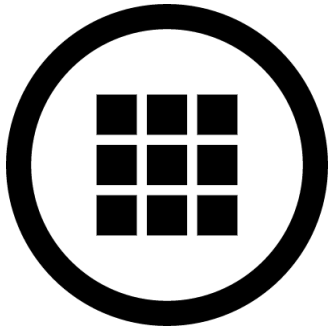
- A cluster is a group of independent IT resources that are interconnected and work as a single system.
- Cluster nodes have reasonably identical hardware and OS
  - provide similar performance levels when one failed component is to be replaced by another.
- Cluster nodes are synchronised through dedicated, high-speed communication links.
- System failure rates are reduced while availability and reliability are increased, since redundancy and failover features are inherent to the cluster.
- The basic concepts of built-in redundancy and failover are core to cloud platforms.



Technology Innovations: Grid Computing

Grid Computing:

- GC provides a platform in which computing resources are organized into one or more logical pools.
- these pools are collectively coordinated to provide a high-performance distributed grid (super virtual computer).
- GC differs from Clustering:
  - grid systems are much more loosely coupled and distributed.
  - grid computing systems can involve heterogeneous and geographically dispersed computing resources
  - grid computing systems are often connected through a relatively low-speed network (e.g. Internet)
- GC is based on a middleware layer that is deployed on computing resources in a grid pool
  - implements a series of workload distribution and coordination functions.
- an on-going research area in computing science since the early 1990s.
- Impacts of GC's research outcomes on cloud computing platforms and mechanisms:
  - networked access,
  - resource pooling,
  - scalability and resiliency.



Technology Innovations: Virtualisation Technology

Virtualisation:

- VT is used to create virtual instances of IT resources.
  - A layer of virtualization software allows physical IT resources to provide multiple virtual images of themselves
  - underlying processing capabilities can be shared by multiple users.
- before VT, software was coupled with static hardware environments.
- VT overcomes the software-hardware dependency
  - hardware requirements can be simulated by emulation software running in virtualized environments.
- VT inspires several cloud characteristics and cloud computing mechanisms.
- Evolutions of cloud computing advance modern virtualization technologies to overcome the performance, reliability, and scalability limitations.

Outline

I. Course Introduction

- Teaching Team
- Course Website & Blackboard
- Course Overview – Lectures, Tutorials & Practicals
- Course Assessment & Policy
- Learning Materials

II. Introduction to Cloud Computing

- History & Definitions
- Business Drivers
- Technology Innovations
- Basic Terms & Cloud Characteristics
- Cloud Delivery Models & Cloud Deploy Models
- Cloud-enabling technologies: Broadband Networks and Internet Architecture, Virtualisation Technology (VT), Data Centre Technology, Web Technology, and Multitenant Technology
- Goals and Benefits
- Risks and Challenges
- Cloud-based Applications in the World

Basic Concepts and Terminology

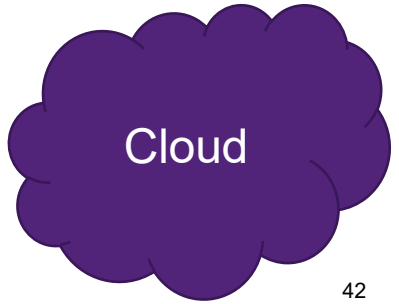
Cloud:

a distinct IT environment that is designed for the purpose of remotely provisioning *scalable and measured* IT resources.

Cloud VS Internet:

- Cloud has a clear and finite boundary
- Cloud is usually private and offers metered IT resources
- Cloud often provides back-end processing capabilities

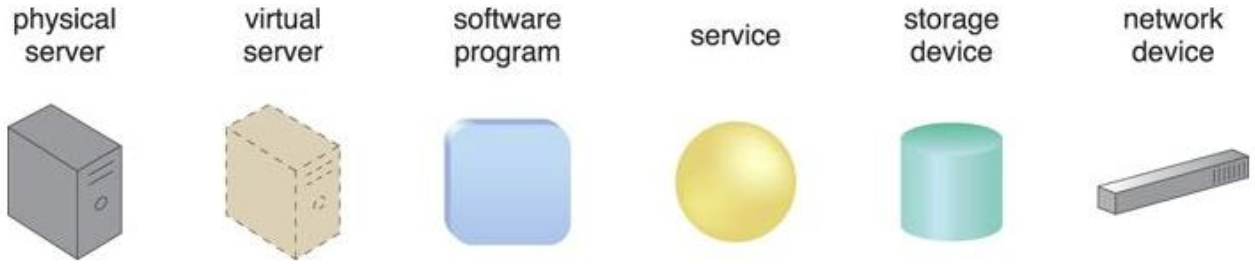
Internet Symbol in CC

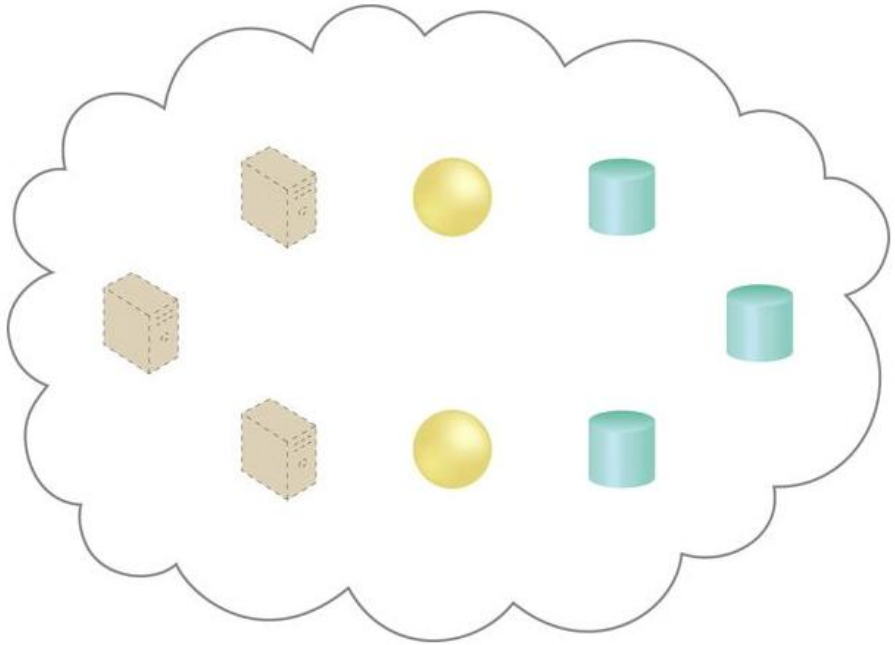


Basic Concepts and Terminology

IT Resource:

a physical or virtual IT-related artifact that can be either software-based or hardware-based.





the cloud symbol can be used to define a boundary for a cloud-based environment that hosts and provisions a set of IT resources.

A cloud is hosting eight IT resources: three virtual servers, two cloud services, and three storage devices.

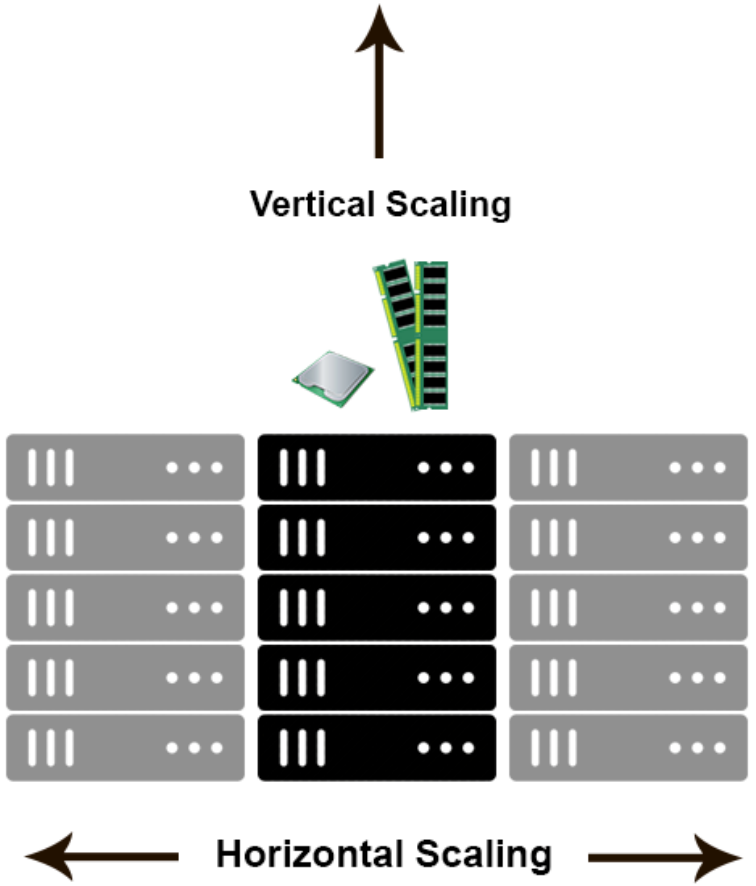
Basic Concepts and Terminology

Scaling:

the ability of the IT resource to handle increased or decreased usage demands.

Scaling Types:

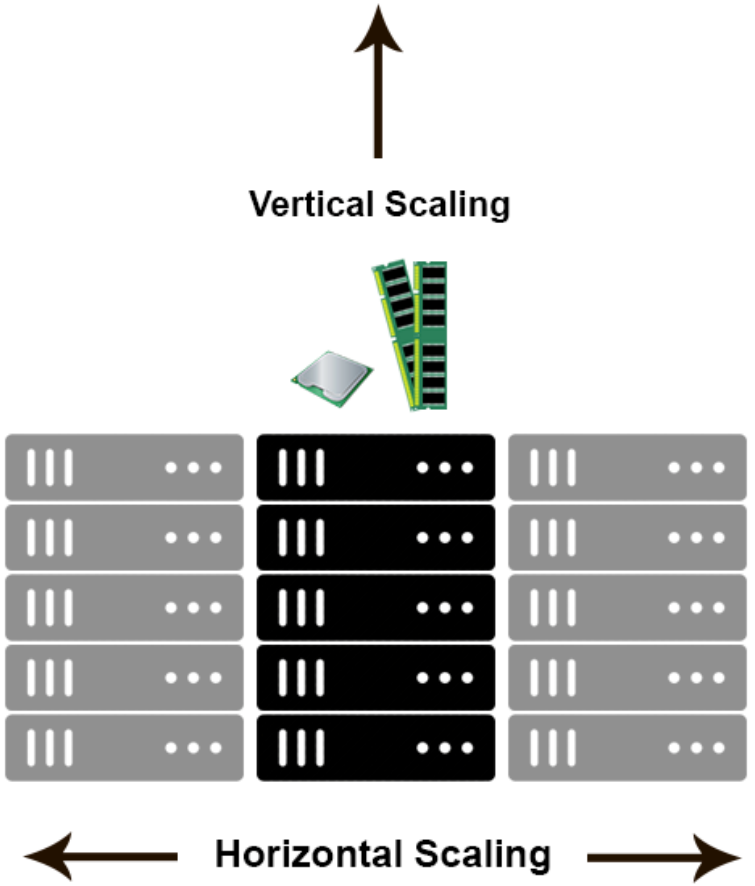
- Horizontal scaling: allocating or releasing of IT resources (same type)
- Vertical scaling: higher or lower capacity of the current IT resources (less common due to downtime)



Basic Concepts and Terminology

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

A comparison of horizontal and vertical scaling.



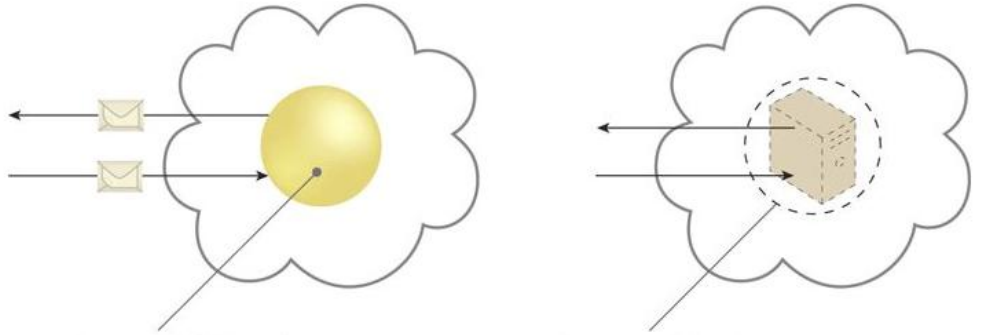
Basic Concepts and Terminology

Cloud Service:

any IT resource that is made remotely accessible via a cloud.

Broad context:

- a simple Web-based software program
- a remote access point for administrative tools
- larger environments
- etc



Driving Motivation -- to provide IT resources as services

- types of cloud services can be labelled with the "as-a-service" suffix
- Examples: Infrastructure-as-a-service (IaaS), PaaS, and SaaS

Cloud Characteristics

On-Demand Self-service

- Cloud consumer has freedom to self-provision cloud's IT resources
- Once configured, the usage can be automated without further involvement
- also known as "on-demand self-service usage"
- enables the service-based and usage-driven features found in mainstream clouds.

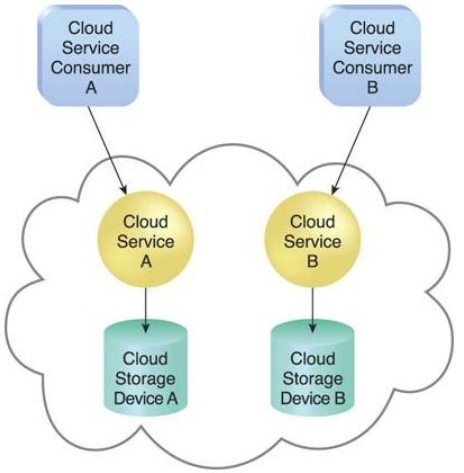
Ubiquitous Access

- Widely accessible
- Normally, accessed via Internet (requires support for a range of devices, transport protocols, interfaces, and security technologies)

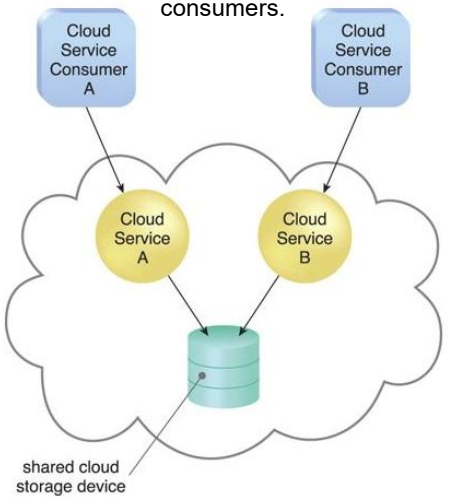
Multitenancy

- program instance can serve different consumers (e.g. database)
- good for dynamically assigned and reassigned resources (by virtualization technologies)

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 cloud storage device, serves multiple consumers.



Cloud Characteristics

Elasticity

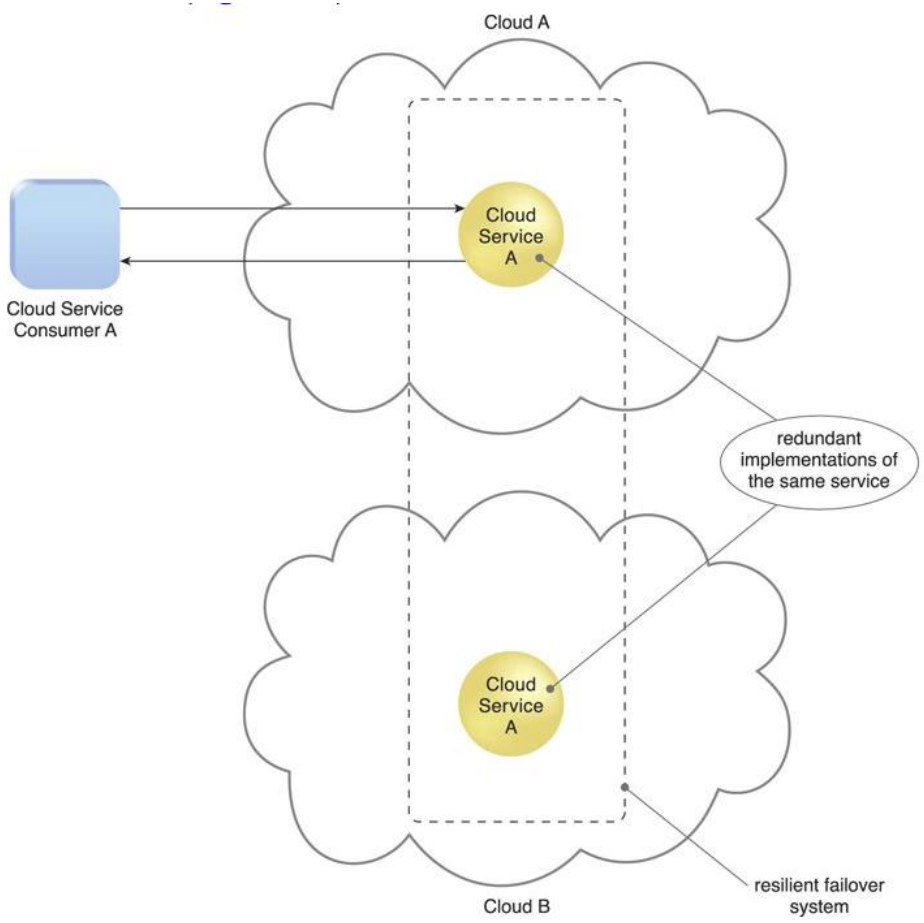
- is the ability to automatically scale IT resources
- is often considered a core justification for the adoption of cloud computing
- closely associates with the financial benefits.

Measured Usage

- ability to keep track of the usage of IT (by cloud consumers)
- is closely related to the on-demand characteristic
- billing purpose and usage report (for both cloud providers and consumers)

Resiliency

- is a form of failover that distributes redundant implementations of IT resources across physical locations.
- if one becomes deficient, processing is automatically handed over to another redundant implementation.
- increases reliability



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

Outline

I. Course Introduction

- Teaching Team
- Course Website & Blackboard
- Course Overview – Lectures, Tutorials & Practicals
- Course Assessment & Policy
- Learning Materials

II. Introduction to Cloud Computing

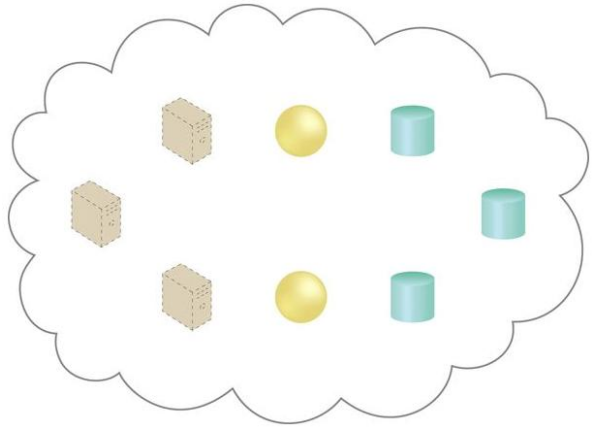
- History & Definitions
- Business Drivers
- Technology Innovations
- Basic Terms & Cloud Characteristics
- Cloud Delivery Models & Cloud Deploy Models
- Cloud-enabling technologies: Broadband Networks and Internet Architecture, Virtualisation Technology (VT), Data Centre Technology, Web Technology, and Multitenant Technology
- Goals and Benefits
- Risks and Challenges
- Cloud-based Applications in the World

Cloud Delivery Models

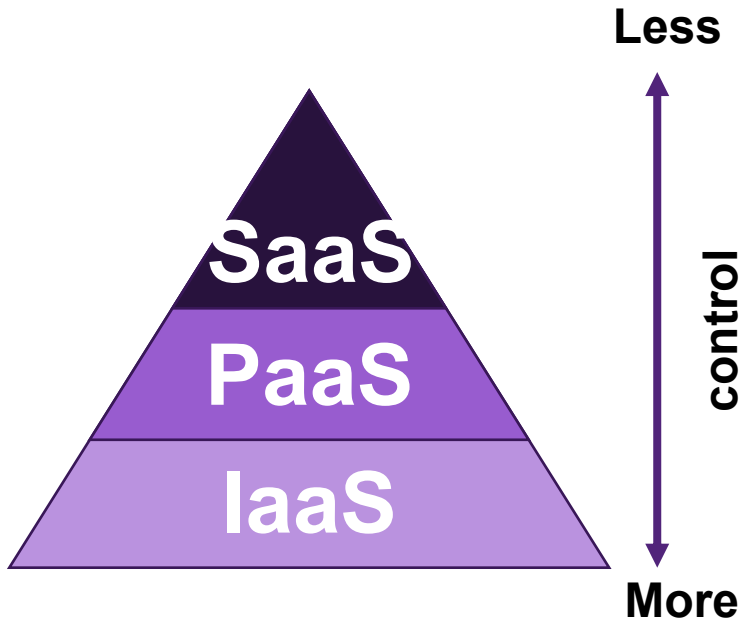
A *cloud delivery model* represents a specific, pre-packaged combination of IT resources offered by a cloud provider.

Three common cloud delivery models:

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



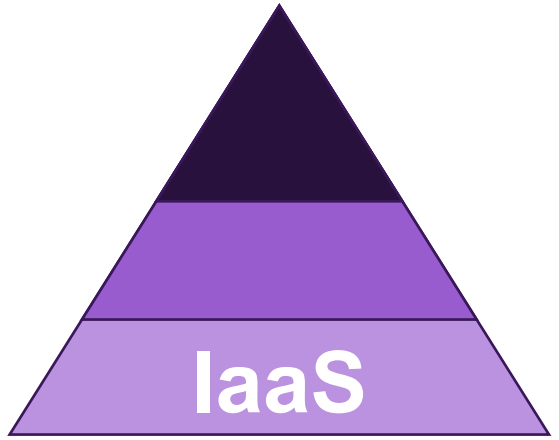
Cloud = IT Env (IT resources)

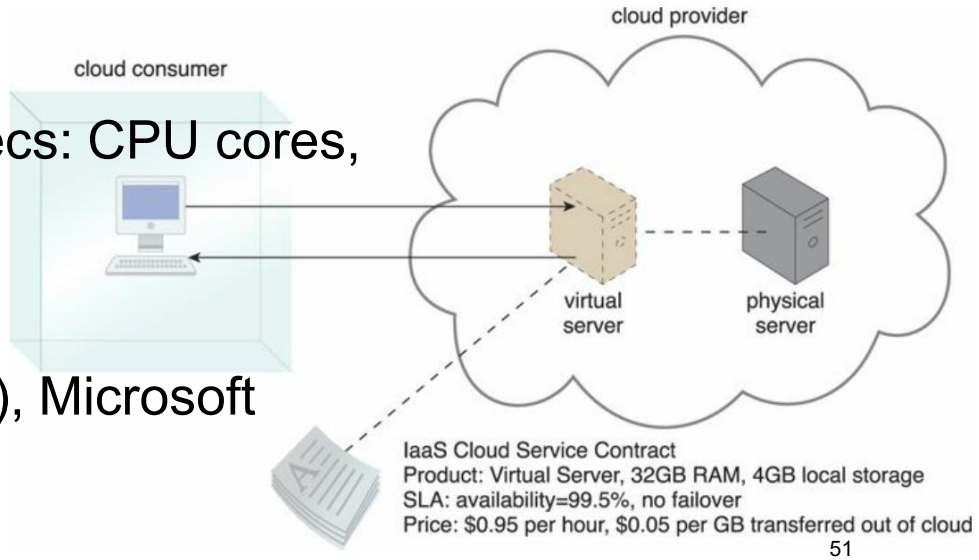


Cloud Delivery Models -- IaaS

Infrastructure-as-a-Service (IaaS):

- a self-contained virtual environment consists of infrastructure-centric IT resources – e.g. VMs on GCP
- IT resources can be accessed and managed via cloud service-based interfaces and tools
- can include hardware, network, connectivity, operating systems, and other "raw" IT resources
- provides a high level of control and responsibility over its configuration and utilization
- needs cloud consumers' administrative responsibility
- can be different by different cloud providers (different specs: CPU cores, RAM, storage, etc)
- is generally offered as freshly initialized virtual machines
- users: system admins
- examples: Amazon EC2, Google Compute Engine (GCE), Microsoft Azure

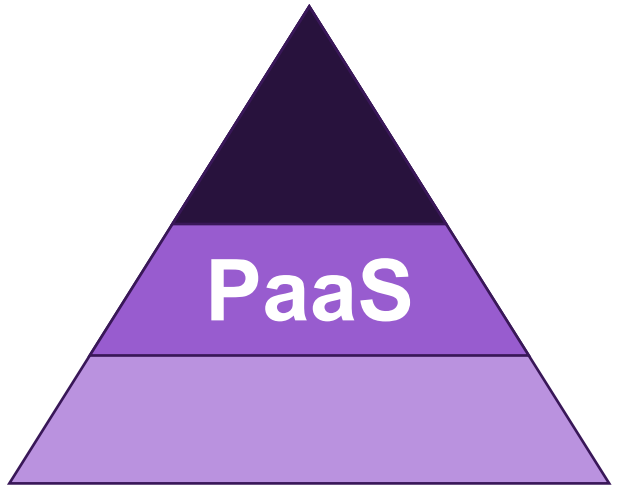


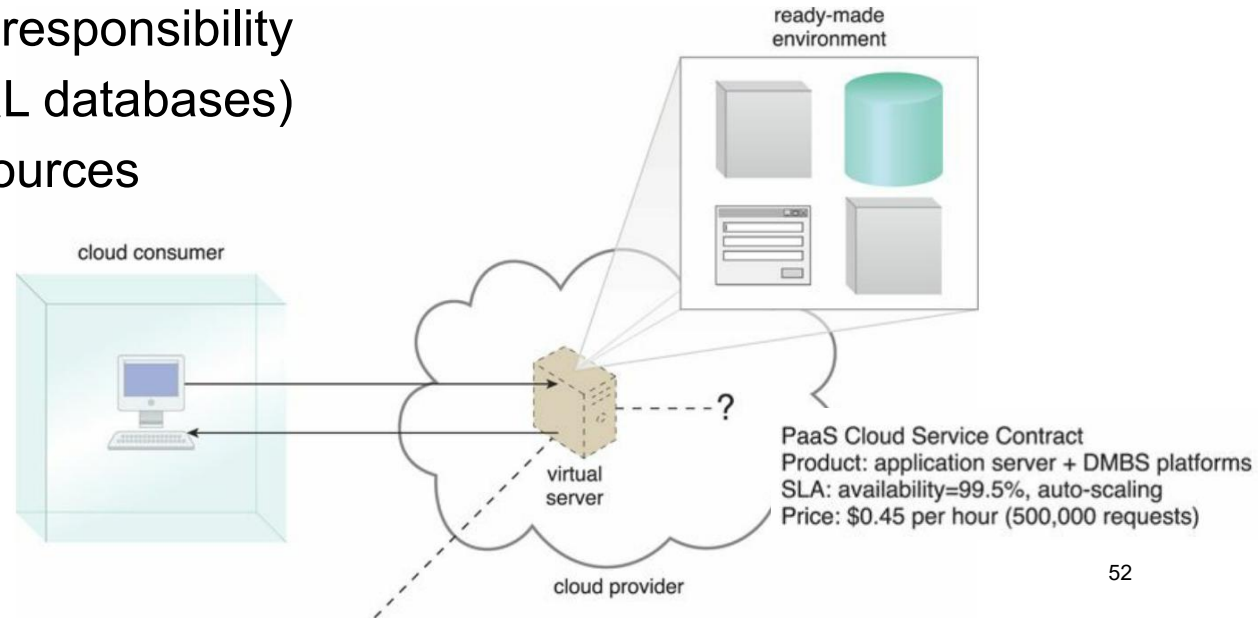


Cloud Delivery Models -- PaaS

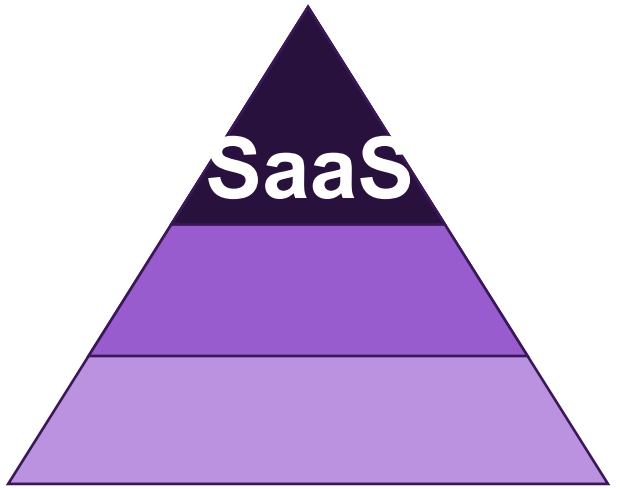
Platform-as-a-Service (PaaS):

- a pre-defined "ready-to-use" environment typically consists of already deployed and configured IT resources
- can include a programming language execution environment, an operating system, a web server, and a database.
- encapsulates an environment where users can build, compile, and run program without worrying about the infrastructure.
- no need to take administrative and maintaining responsibility
- needs users to manage their own data (e.g. SQL databases)
- lower level of control over the underlying IT resources
- users: developers
- examples:
  - AWS Elastic Beanstalk
  - Google app engine (GAE)



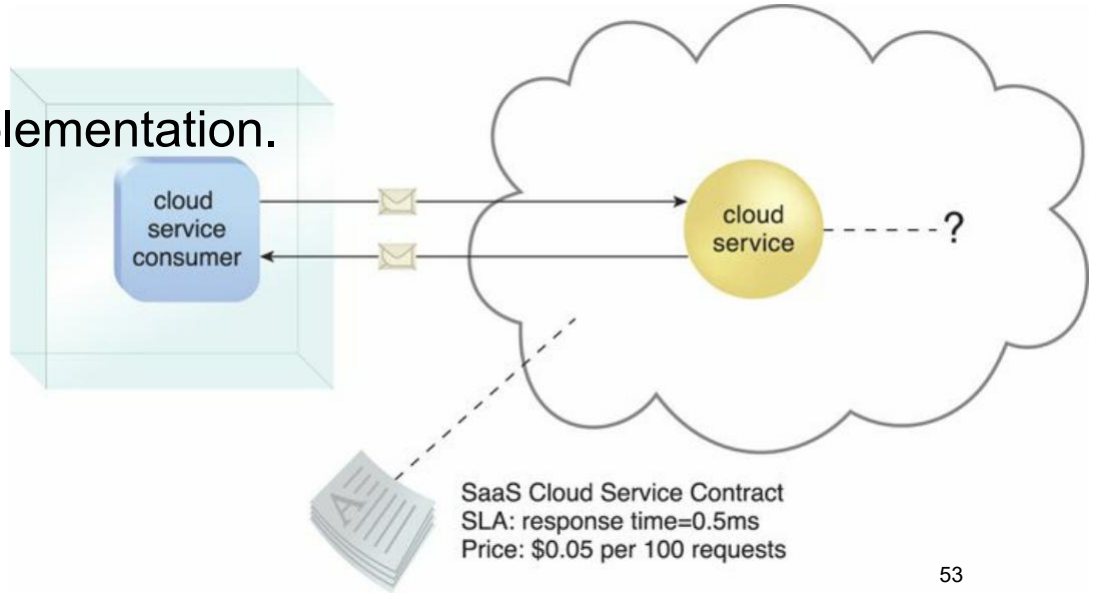


Cloud Delivery Models -- SaaS



Software-as-a-Service (SaaS):

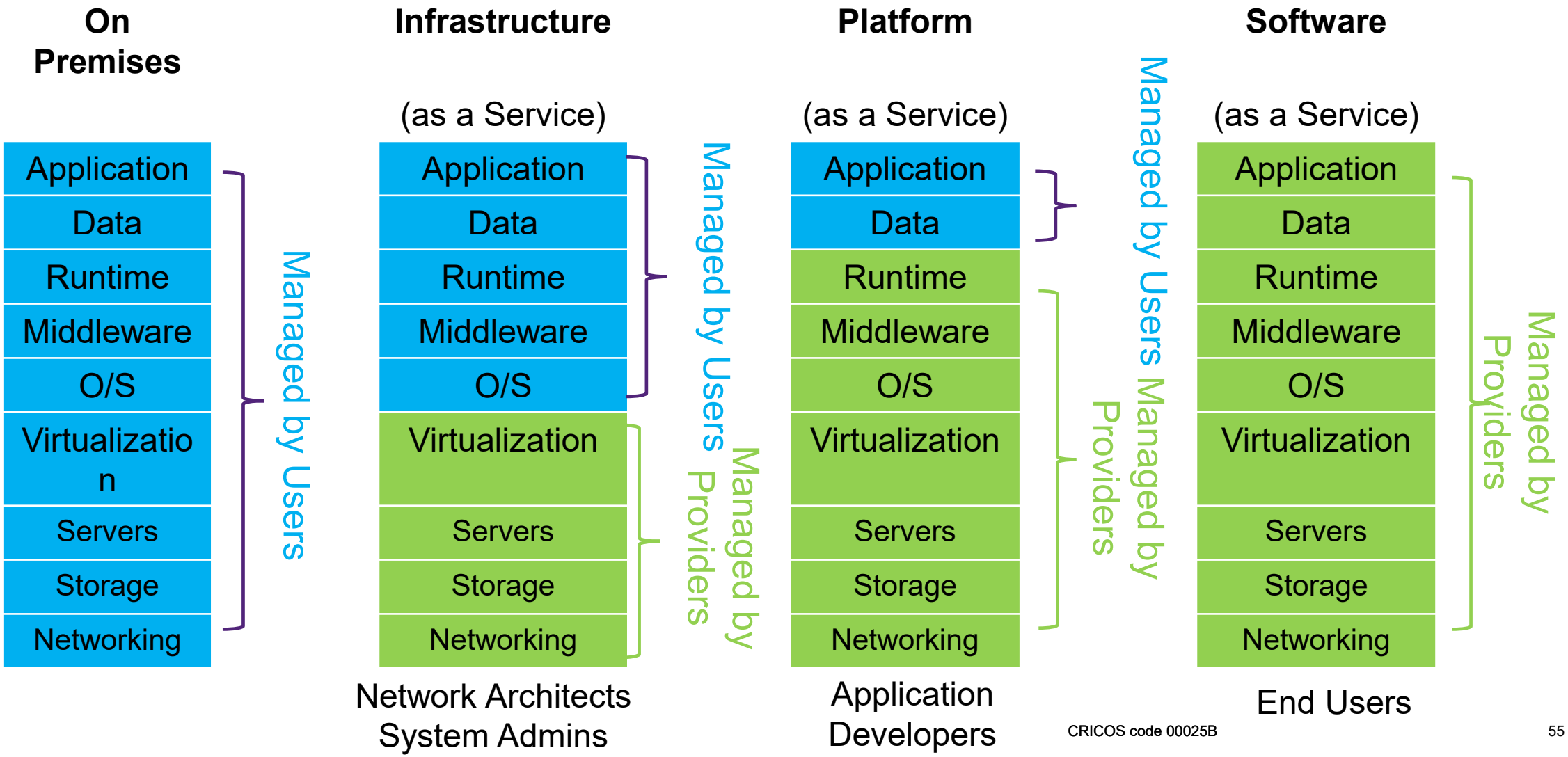
- a shared cloud service and made available as a "product"
- provides on-demand services
- no installation of the software on users' PCs
- accessible via a web browser or lightweight client apps.
- run a single instance of the software
- can be available for multiple users
- has very limited administrative control over a SaaS implementation.
- users: end-users
- examples:
  - Google G-Suite docs/sheets/mails/calendar/etc
  - Microsoft Office 365



Model Comparisons

Delivery Model	Control Level	Functionality	Consumer Activities	Provider Activities
SaaS	Usage and usage-related configuration	Access to front-end user-interface	Uses and configures cloud services	Implements, manages, and maintains cloud service Monitors usage by cloud consumers
PaaS	Limited administrative	Moderate level of administrative control over IT resources relevant to cloud consumer's usage of platform	Develops, tests, deploys, and manages cloud services and cloud-based solutions	Pre-configures platform and provisions underlying infrastructure, middleware, and other needed IT resources, as necessary Monitors usage by cloud consumers
IaaS	Full administrative	Full access to virtualized infrastructure-related IT resources and possibly to underlying physical IT resources	Sets up and configures bare infrastructure, and installs, manages, and monitors any needed software	Provisions and manages the physical processing, storage, networking, and hosting required Monitors usage by cloud consumers

Model Comparisons



New Adventures in the Cloud – Game-as-a-Service

STADIA

shut down on
January 18, 2023

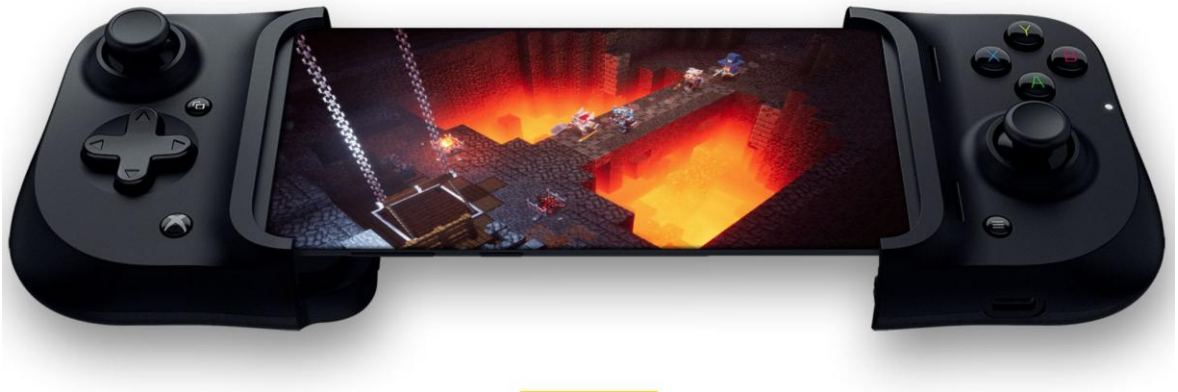
NVIDIA GEFORCE NOW

Nintendo

GAME PASS

PlayStation Now

PlayStation Now



COMING SEPTEMBER 15

Cloud gaming (Beta) with Xbox Game Pass Ultimate

Play over 100 games on your Android mobile phone or tablet from the cloud with Xbox Game Pass Ultimate.

https://www.cnbc.com/2020/08/04/microsoft-reveals-more-details-about-its-xcloud-game-streaming-service.html
https://www.xbox.com/en-US/xbox-game-pass/cloud-gaming
https://stadia.google.com/

Examples of Cloud Computing Usage

Amazon's AWS:

- leading company in Cloud Computing
- provides IaaS and PaaS
- famous for EC2

Google Cloud:

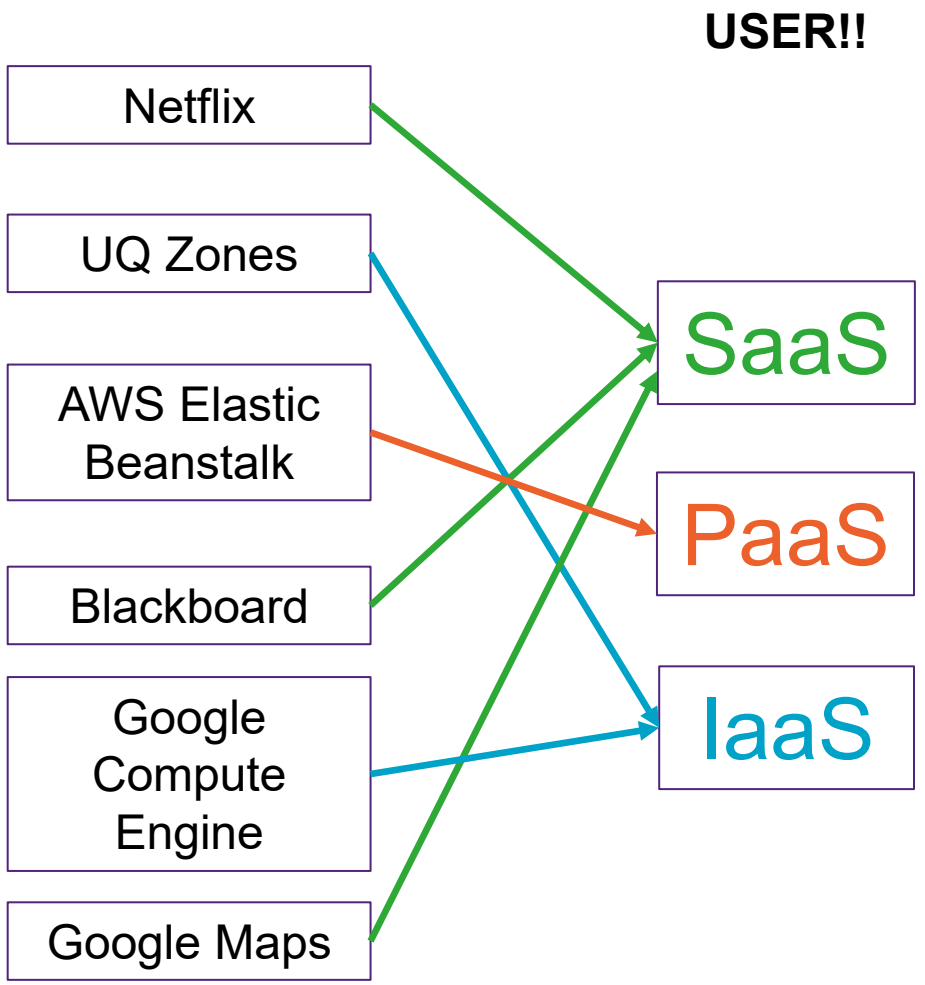
- offers IaaS, PaaS (GAE), and SaaS (Google docs/sheets/calendar/gmail)

Microsoft Azure:

- provides IaaS, PaaS, and SaaS (Office 365)

iCloud:

- majorly for Apple products (Macbook, iPad, iPhone, etc)
- store and backup users' documents online.
- Storage-as-a-service (aka STaaS)



Cloud Deployment Models

A cloud deployment model represents a specific type of cloud environment

In terms of ownership, size, and access, models can be divided into four common groups:

- Public cloud
  - a publicly accessible cloud environment owned by a third-party cloud provider
  - usually supplied via the delivery models and offered to consumers at a cost
  - is created and on-going maintained by the cloud provider.
  - typical examples: GCP, AWS, AZURE, etc.
- Community cloud
  - is similar to a public cloud except that its access is limited to a community of cloud consumers.
  - may be jointly owned by the community members or by a third-party cloud provider
  - cloud consumers of the community typically share the responsibility for defining and evolving the community cloud
  - Typical examples: Cloud for multiple governmental departments, e.g. FedRAMP.

Cloud Deployment Models

- Private cloud
  - is owned by a single organization and enables an organization to use CC technology to access to IT resources by different parts, locations, or departments.
  - actual administration may be carried out by internal or outsourced staff.
  - within a private cloud, the same organization is technically both the cloud consumer and provider
  - Typical examples: UQCloud (VMs) and UQRDM cloud (STaaS)
- Hybrid cloud
  - is a cloud environment comprised of two or more different cloud deployment models.
  - Example: private cloud (sensitive data) + public cloud (less sensitive cloud services)
  - can be complex and challenging to create and maintain due to the potential disparity in cloud environments

Outline

I. Course Introduction

- Teaching Team
- Course Website & Blackboard
- Course Overview – Lectures, Tutorials & Practicals
- Course Assessment & Policy
- Learning Materials

II. Introduction to Cloud Computing

- History & Definitions
- Business Drivers
- Technology Innovations
- Basic Terms & Cloud Characteristics
- Cloud Delivery Models & Cloud Deploy Models
- Cloud-enabling technologies: Broadband Networks and Internet Architecture, Virtualisation Technology (VT), Data Centre Technology, Web Technology, and Multitenant Technology
- Goals and Benefits
- Risks and Challenges
- Cloud-based Applications in the World

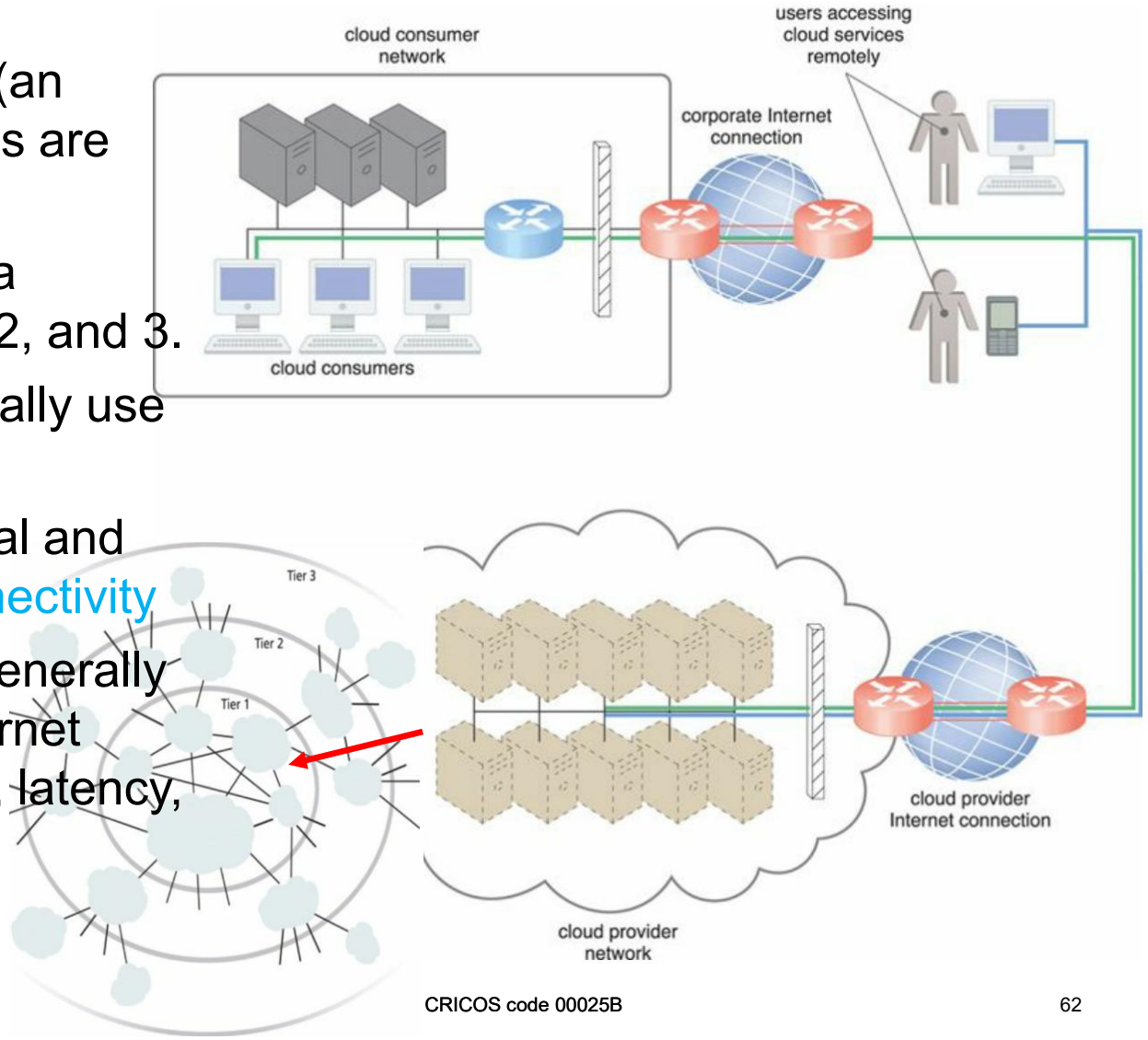
Cloud-Enabling Technology (CET)

Modern-day clouds are underpinned by a set of primary technology components that collectively enable key features and characteristics associated with contemporary cloud computing:

- Broadband Networks and Internet Architecture
- Virtualisation Technology
- Data Centre Technology
- Web Technology
- Multitenant Technology

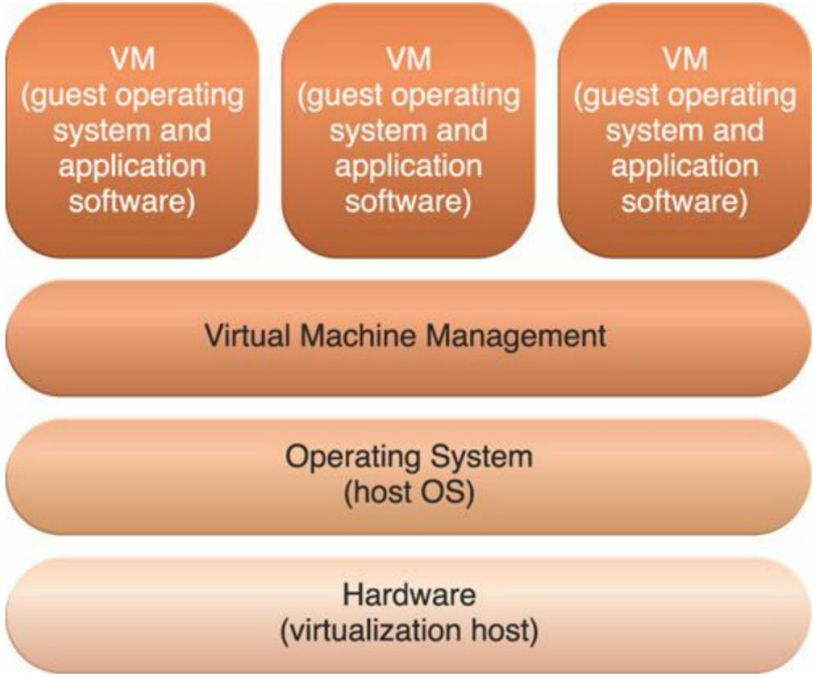
CET I – Broadband Networks and Internet Architecture

- All clouds must be connected to a network (an inherent dependency on WWW, most clouds are Internet-enabled).
- Worldwide connectivity is enabled through a hierarchical topology composed of Tiers 1, 2, and 3.
- Cloud consumers and cloud providers typically use the Internet to communicate.
- Easily configuration of IT resources (external and internal users via WWW) and superior connectivity
- The potential of cloud platforms therefore generally grows in parallel with advancements in Internet connectivity and service quality (bandwidth, latency, protocols, etc.)



CET II: Virtualisation Technology (VT)

- VT is the process of converting a physical IT resource into a virtual IT resource.
- Most types of IT resources can be virtualised, including:
  - *Servers* – A physical server can be abstracted into a virtual server.
  - *Storage* – A physical storage device can be abstracted into a virtual storage device or a virtual disk.
  - *Network* – Physical routers and switches can be abstracted into logical network fabrics, such as VLANs.
  - *Power* – A physical UPS and power distribution units can be abstracted into what are commonly referred to as virtual UPSs.
- A physical server is called a *host* or *physical host.*
- A software that manages VMs and hardware is called as *Virtual Machine Monitor (VMM)*, also known as *hypervisor* in cloud computing context.
- An operating system in a virtual machine is called as *guest OS*.



CET II: Virtualisation Technology (VT)

- Steps of creating a new virtual server through virtualisation software:
  I. the allocation of physical IT resources (e.g. specify #CPU, Mem, Storage in VirtualBox by Oracle);
  II. followed by the installation of an operating system (e.g. Install Ubuntu or Windows systems in VirtualBox).
- Virtual servers use their own guest operating systems, which are independent of the operating system in which they were created.
- Guest OS and the application software on the virtual server are unaware of the virtualisation process.

CET II: Virtualisation Technology (VT)

Hardware Independence

- VT can convert and translate IT hardware into emulated and standardised software-based copies.
- Due to HI, virtual servers can easily be moved to another virtualisation host
- Thus, cloning and manipulating virtual IT resources is much easier than duplicating physical hardware.

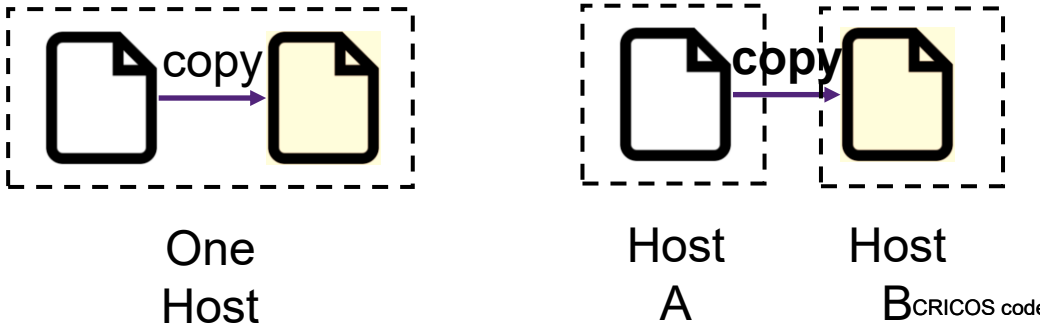
Server Consolidation

- VT enables different virtual servers to share one physical server, which is called *server consolidation*
- SC is commonly used to increase hardware utilisation, load balancing, and optimisation of available IT resources.
- The resulting flexibility: different virtual servers can run different guest operating systems on the same host.
- supports common cloud features, e.g. on-demand usage, resource pooling, elasticity, scalability, and resiliency.

CET II: Virtualisation Technology (VT)

Resource Replication

- Virtual servers are created as virtual disk images that contain binary file copies of hard disk content.
- Host's OS can access these disk images e.g. copy, move, and paste (replicate, migrate, and back up the virtual server).
- In this way, it enables:
  - Standard virtual machine creations with common configurations
  - Increased agility in the migration and deployment of a virtual machine's new instances
  - Backup & Roll back abilities



CET II: Virtualisation Technology (VT)

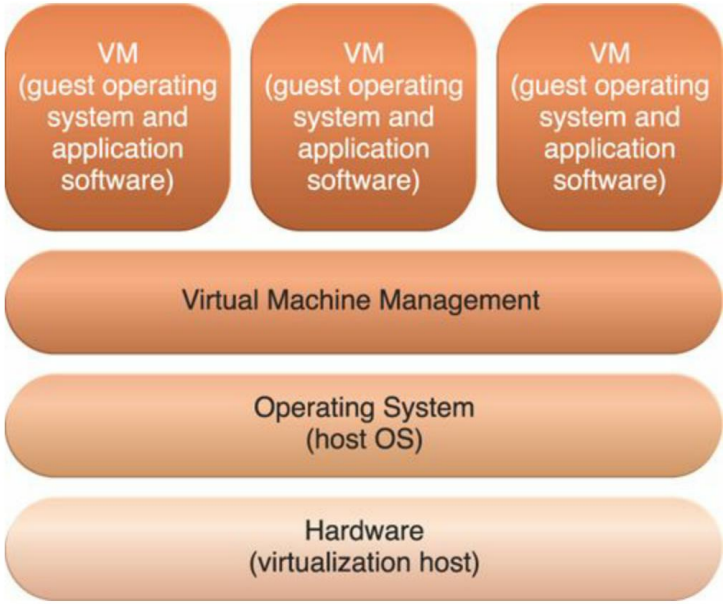
Operating System-Based Virtualisation

- Install virtualisation software in a pre-existing operating system (called the host operating system)
- example: Install ubuntu on Windows with VMware/VirtualBox
- processing overhead: virtualisation software and host OS.

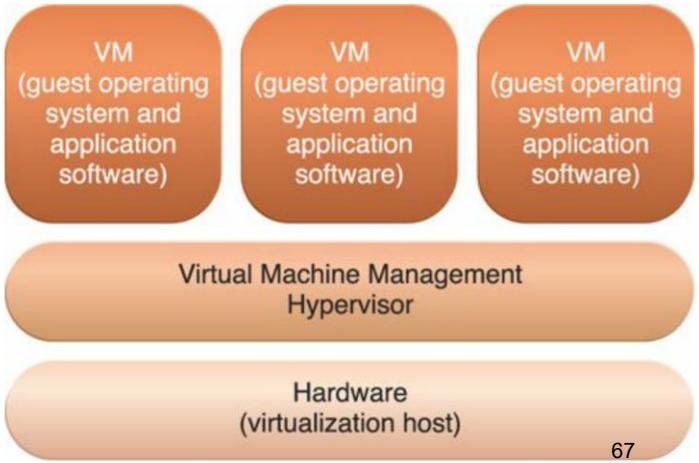
Hardware-Based Virtualisation

- Install virtualisation software directly on the physical host hardware bypassing the host OS
- example: Oracle VM Server for x86 (up to 384 CPUs and 6TB RAM)
- VMM is also named as Hypervisor
- more efficient (no hosting OS), but compatible issues.

operating system-based virtualisation



Hardware-based virtualisation



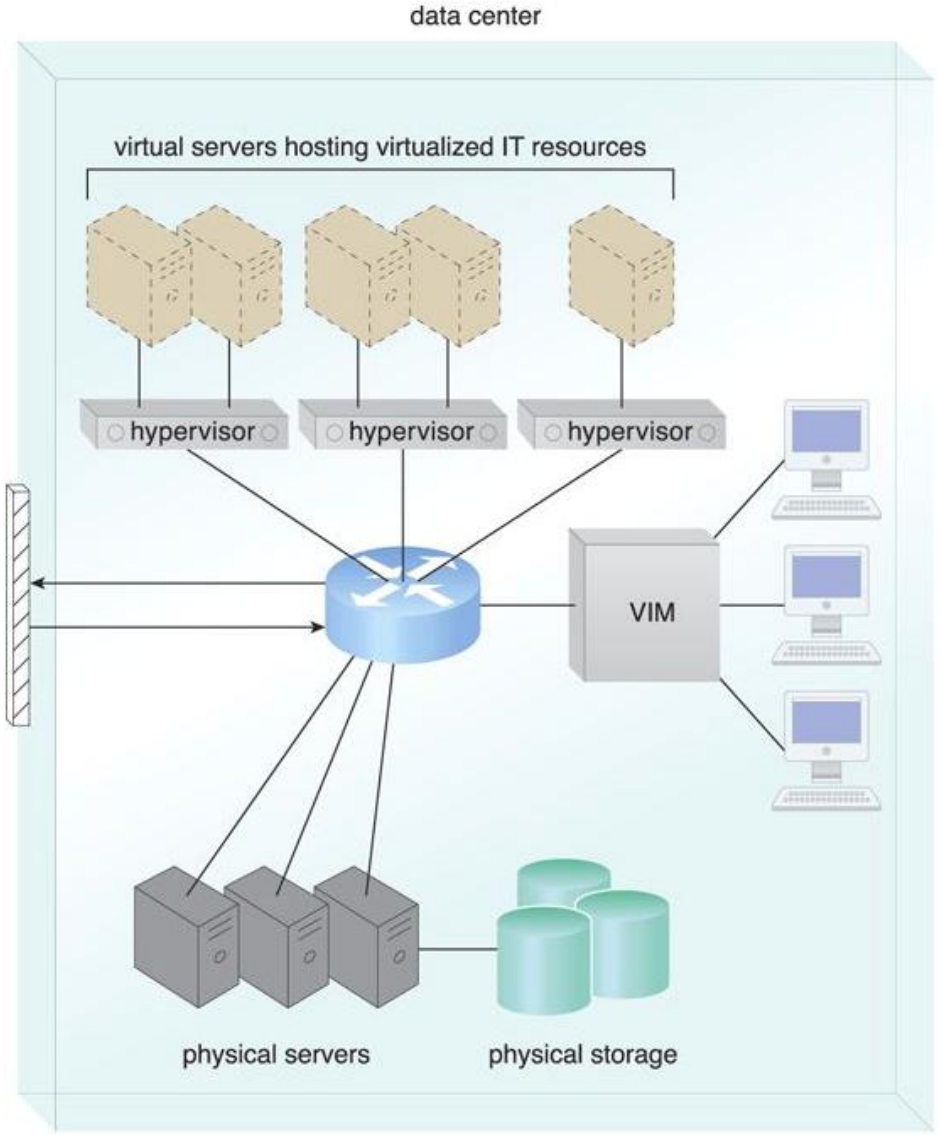
CET III – Data Centre Technology

A data centre is a specialised IT infrastructure that houses centralised IT resources

- Servers (rack in cabinet);
- Databases and software systems;
- Networking and telecommunication devices.

Typical technologies and components

- **Virtualisation**:
  - Data centres consist of both physical and virtualised IT resources.
  - The physical IT resource layer refers to the facility infrastructure that houses:
    - computing/networking systems and equipment,
    - hardware systems and their operating systems.



CET III – Data Centre Technology

Typical technologies and components of Data Centre:

- **Standardisation** and **Modularity**:
  - DCs are built upon standardised commodity hardware and designed with modular architectures.
  - reduce investment and operational costs.
- **Automation**:
  - DCs have specialised platforms that automate general management tasks such as provisioning, configuration, patching, and monitoring without supervision.
- **Remote Operation** and **Management**:
  - Remotely access via consoles and management systems: most of the tasks in DCs (e.g. operational and administrative tasks).
  - On-site jobs: highly specific tasks – equipment handling and cabling or hardware-level installation and maintenance.

CET III – Data Centre Technology

Typical technologies and components of Data Centre:

- **High Availability**:
  - Aiming to high-level availability, DCs usually have redundant, uninterruptable power supplies, cabling, and environmental control subsystems in anticipation of system failure, along with communication links and clustered hardware for load balancing.
- **Security-Aware Design, Operation and Management**:
  - Security requirements (e.g. physical and logical access controls and data recovery strategies) need to be comprehensive for DCs.
- **Facilities**:
  - Site: custom-designed locations that are outfitted with specialised computing, storage, and network equipment.
  - Layout: multiple functional areas
  - various power supplies, cabling, and environmental control stations that regulate heating, ventilation, air conditioning, fire protection, and other related subsystems.

CET III – Data Centre Technology

Hardware of Data Centres:

- Computing Hardware:
  - rackmount form factor server (multiple racks in a cabinet);
  - a power-efficient multi-core CPU architecture (many cores but low frequency, e.g. Xeon/EPYC CPUs);
  - redundant and hot-swappable components, such as hard disks, power supplies, network interfaces, and storage controller cards.
- Storage Hardware:
  - specialised storage systems that maintain enormous amounts of digital information in order to fulfill considerable storage capacity needs by using arrays of disks;
  - frequently used storage technologies: RAID, Hot-Swappable, Virtualisation, and Fast Data Replication Mechanisms.
- Network Hardware:
  - LAN fabric, high-performance switches & adaptors (up to 10 G/s), etc.

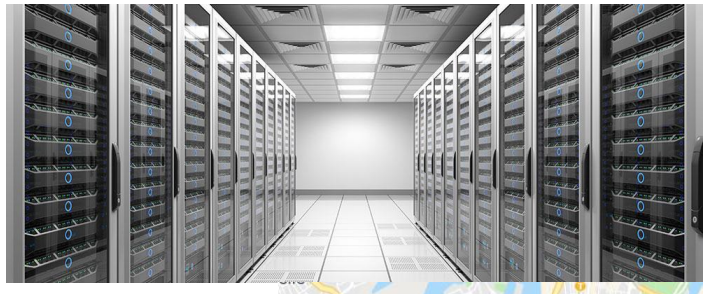
CET III – World-class Data Centres

Five Largest Data Centres in the world [1] (as of date Feb 2022)

1. China Telecom Data Centre (China)
2. China Mobile (Inner Mongolia, China)
3. Citadel Campus (Nevada, US)
4. CWL1 Data Centre (Newport, Wales)
5. Apple's Mesa Data Center (Arizona, US)

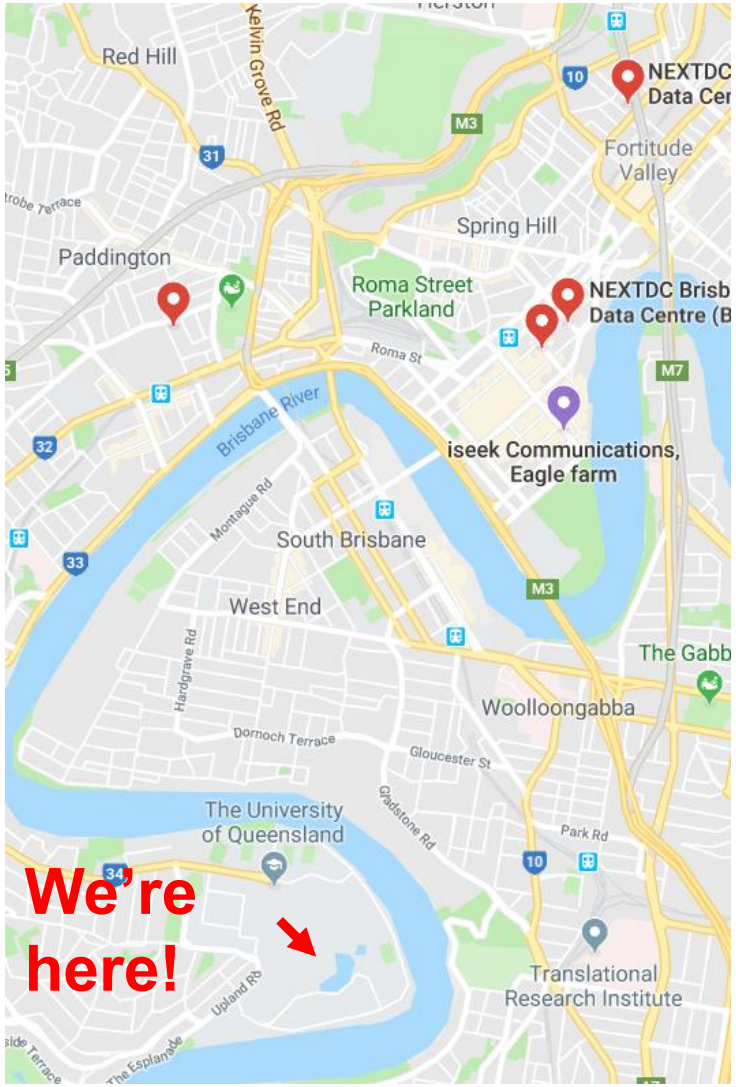
Largest Data Centre in Australia:

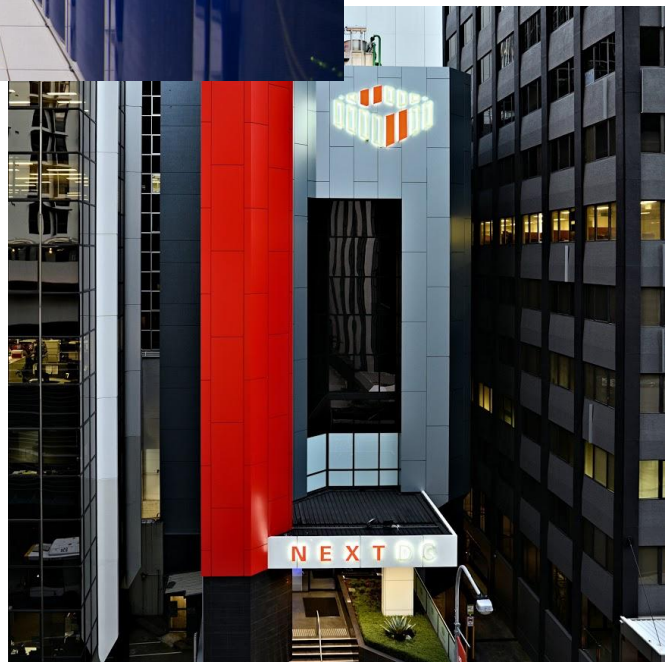
EQUINIX SY3 Data Centre (Sydney) [2]



1. https://www.rankred.com/largest-data-centers-in-the-world/
2. https://cloudscene.com/market/data-centers-in-australia/all

CET III – Data Centre in Brisbane





NEXTDC Brisbane Data Centre https://www.nextdc.com/

CET IV – Web Technology

- Web technology is very commonly used for cloud service implementations and for front-ends used to remotely manage cloud-based IT resources.
- Fundamental technologies of Web architecture:
  - ***Uniform Resource Locator (URL)*** – A standard syntax used for creating identifiers that point to Web-based resources, the URL is often structured using a logical network location.
  - ***Hypertext Transfer Protocol (HTTP)*** – This is the primary communications protocol used to exchange content and data throughout the World Wide Web. URLs are typically transmitted via HTTP.
  - ***Markup Languages (HTML, XML)*** – Markup languages provide a lightweight means of expressing Web-centric data and metadata: HTML (webpages) and XML (data).
- **Example**: a web browser can request to execute an action like **read, write, update, or delete** on a web resource on the Internet, and proceed to identify and locate the Web resource through its **URL**. The request is sent using **HTTP** to the resource host, which is also identified by a **URL**. The Web server locates the Web resource and performs the requested operation, which is followed by a response being sent back to the client. The response may be comprised of content that includes **HTML** and **XML** statements.

CET V – Multitenant Technology

- Multitenant application enables multiple users (tenants) to access the same application logic simultaneously.
- Each tenant has its own view of the application that it uses, administers, and customises as a dedicated instance of the software while remaining unaware of other tenants that are using the same application.
- Multitenant applications ensure that tenants do not have access to data and configuration information that is not their own.
- Tenants can individually customise features of the application:
  - User Interface – Tenants can define a specialised “look and feel” for their application interface.
  - Business Process – Tenants can customise the rules, logic, and workflows of the business processes that are implemented in the application.
  - Data Model – Tenants can extend the data schema of the application to include, exclude, or rename fields in the application data structures.
  - Access Control – Tenants can independently control the access rights for users and groups.

CET V – Multitenant Technology

- **Multitenant** application architecture is significantly more complex than that of **single-tenant** applications.
- Common **characteristics** of multitenant applications include:
  - **Usage Isolation** – individual behaviour does NOT affect the other tenants' behaviours.
  - **Data Security** – Tenants cannot access data that belongs to other tenants.
  - **Recovery** – Backup and restore procedures are separately executed for the data of each tenant.
  - **Application Upgrades** – Tenants are not negatively affected by the synchronous upgrading of shared software artifacts.
  - **Scalability** – The application can scale to accommodate increases in usage by existing tenants and/or increases in the number of tenants.
  - **Metered Usage** – Tenants are charged only for the application processing and features that are actually consumed.
  - **Data Tier Isolation** – Tenants can have individual databases, tables, and/or schemas isolated from other tenants. Alternatively, databases, tables, and/or schemas can be designed to be intentionally shared by tenants.

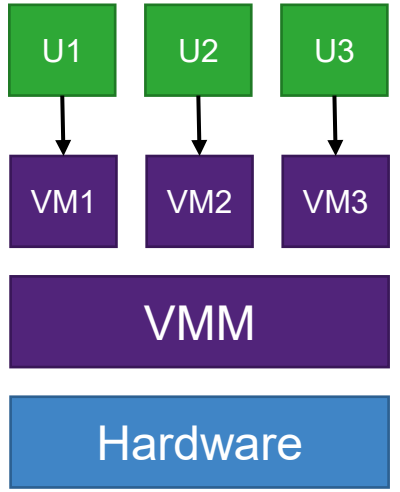
CET V – Multitenant Technology

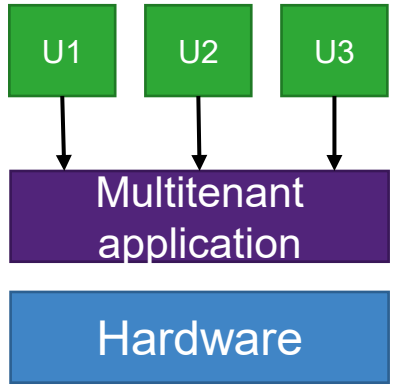
Multitenancy vs. Virtualisation

Multitenancy is sometimes mistaken for virtualisation because the concept of multiple tenants is similar to the concept of virtualised instances.

The differences lie in what is multiplied within a physical server acting as a host:

- **With virtualisation**:
  - Multiple virtual copies of the server environment can be hosted by a single physical server.
  - Each copy can be provided to different users, can be configured independently, and can contain its own operating systems and applications.
- **With multitenancy**:
  - A physical or virtual server hosting an application is designed to allow usage by multiple different users.
  - Each user feels as though they have exclusive usage of the application.





Outline

I. Course Introduction

- Teaching Team
- Course Website & Blackboard
- Course Overview – Lectures, Tutorials & Practicals
- Course Assessment & Policy
- Learning Materials

II. Introduction to Cloud Computing

- History & Definitions
- Business Drivers
- Technology Innovations
- Basic Terms & Cloud Characteristics
- Cloud Delivery Models & Cloud Deploy Models
- Cloud-enabling technologies: Broadband Networks and Internet Architecture, Virtualisation Technology (VT), Data Centre Technology, Web Technology, and Multitenant Technology
- Goals and Benefits
- Risks and Challenges
- Cloud-based Applications in the World

Goals and Benefits

Reduced Investments and Proportional Costs

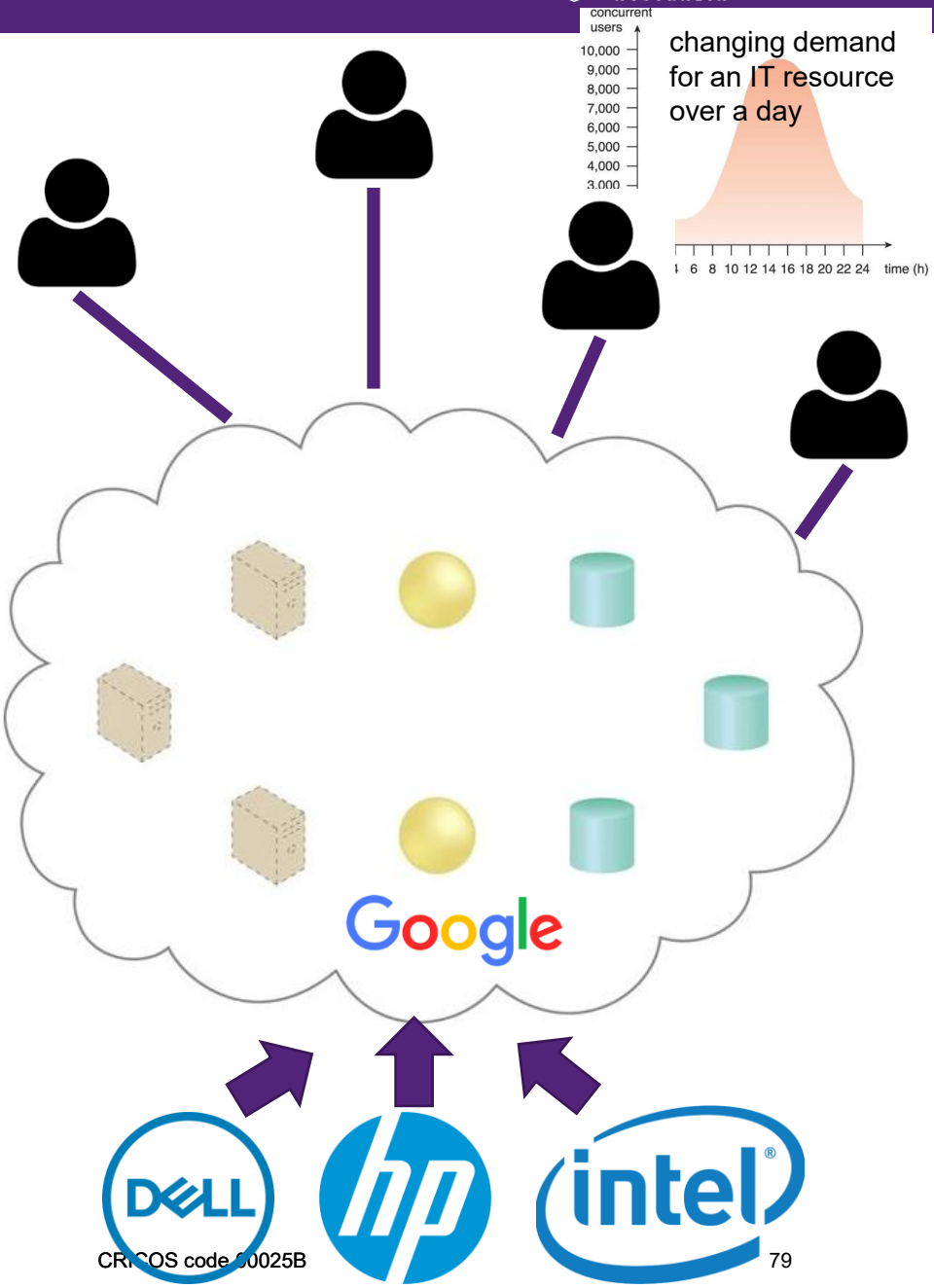
- Capital reduced by cloud provider (mass-inquisition, hardware/software sharing, and data centre deployments)
- Elimination or minimization of IT investments (focus on core business)
- Common measurable benefits, e.g. on-demand access to pay-as-you-go computing resources (CPU by hr)

Increased Scalability

- can instantly and dynamically allocate IT resources
- always meet and fulfil unpredictable demands avoids potential loss

Increased Availability and Reliability

- resilient IT resources
- failover support (recovery)



Outline

I. Course Introduction

- Teaching Team
- Course Website & Blackboard
- Course Overview – Lectures, Tutorials & Practicals
- Course Assessment & Policy
- Learning Materials

II. Introduction to Cloud Computing

- History & Definitions
- Business Drivers
- Technology Innovations
- Basic Terms & Cloud Characteristics
- Cloud Delivery Models & Cloud Deploy Models
- Cloud-enabling technologies: Broadband Networks and Internet Architecture, Virtualisation Technology (VT), Data Centre Technology, Web Technology, and Multitenant Technology
- Goals and Benefits
- Risks and Challenges
- Cloud-based Applications in the World

Risks and Challenges

- **Increased Security Vulnerabilities**
- **Reduced Operational Governance Control**
- **Limited Portability Between Cloud Providers**
- **Multi-regional Compliance and Legal Issues**

Outline

I. Course Introduction

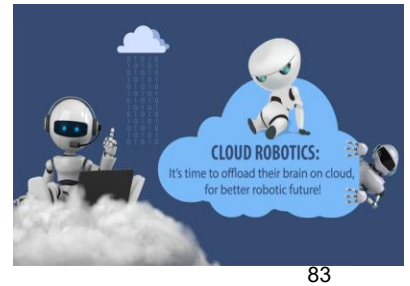
- Teaching Team
- Course Website & Blackboard
- Course Overview – Lectures, Tutorials & Practicals
- Course Assessment & Policy
- Learning Materials

II. Introduction to Cloud Computing

- History & Definitions
- Business Drivers
- Technology Innovations
- Basic Terms & Cloud Characteristics
- Cloud Delivery Models & Cloud Deploy Models
- Cloud-enabling technologies: Broadband Networks and Internet Architecture, Virtualisation Technology (VT), Data Centre Technology, Web Technology, and Multitenant Technology
- Goals and Benefits
- Risks and Challenges
- Cloud-based Applications in the World

Revisit: Importance of Cloud Computing

- It is projected that there will be **24 billion devices on the Internet** by 2020.
- The cloud will become more important as a controller of and resource provider fo
  Things (IoT).
- IDC predicts a similar market size of $210 billion in 2019, rising to $370 billion in 2022.
- Ten years ago, many people claimed that cloud computing was a fad that would never catch on. But today there can be no doubt that cloud computing is a very significant and growing computing trend.
- In the future, "***ubiquitous cities***" and "***smart homes***" can be built on cloud computing (cloud supported/controlled robotics).
  - Cloud robotics is an emerging field of robotics ingrained in cloud computing.
  - Cloud robotics provides a shared knowledge database.
  - Cloud robotics has the abilities about powerful computational, storage, and communications resources with cloud.
  - Cloud robotics offloads heavy computing tasks to the cloud.



https://www.explainingcomputers.com/cloud.html

https://www.gartner.com/en/newsroom/press-releases/2018-09-12-gartner-forecasts-worldwide-public-cloud-revenue-to-grow-17-percent-in-2019

https://www.idc.com/getdoc.jsp?containerId=prUS44891519

Who are using Google Cloud Platform (GCP)?

https://cloud.google.com/customers/#/

Examples of Cloud-based Applications - I

E-Health: Analysing breast cancer images faster and better with machine learning by ***ACS***

Background:

- Cancer is the **second most common** cause of death in the United States (**nearly ¼ deaths**).
- breast cancer is **the most commonly** diagnosed type of cancer and **the second leading cause of cancer death** in the United States.
- If detected early, breast cancer is one of the most survivable cancers: the **five- and ten-year** relative survival rates for women with invasive breast cancer are 90 percent and 83 percent, respectively.
- However, some molecular subtypes of breast cancer have **a poor prognosis** and there is limited understanding of these subtypes.
- Since 1992, the American Cancer Society has conducted the Cancer Prevention Study-II (CPS-II) Nutrition cohort, a prospective study of **more than 188,000** American men and women.
  - CPS-II provides valuable factors: height, weight, demographic characteristics, family history, use of medicines, etc.
  - CPS-II provides medical records and surgical tissue samples for approximately 1,700 CPS-II participants diagnosed with breast cancer

Aim to answer:

- What lifestyle, medical, and genetic factors are related to molecular subtypes of breast cancer?
- Do different features in the breast cancer tissue translate to a better survival rate?

https://cloud.google.com/customers/american-cancer-society/

https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2018/cancer-facts-and-figures-2018.pdf

Examples of Cloud-based Applications - I

American Cancer Society

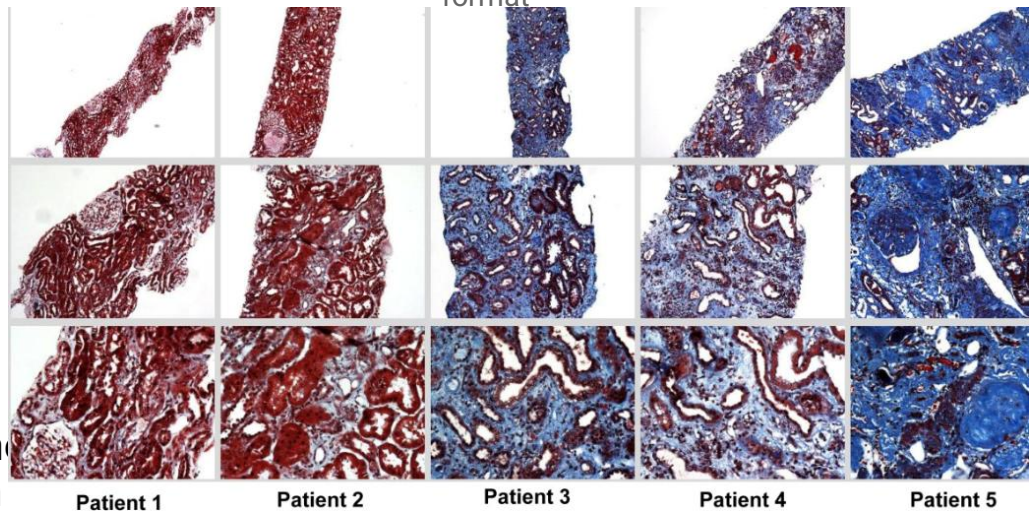
Challenges:

- The computing power to analyse the high-resolution tissue;
- Effective and efficient detection by human experts;
- Labour costs of novel pattern detection in the image data.

Solutions:

- Convert the tissue images into TIF format and store on **Cloud Storage (scalable and data security)**
- Pre-process Data (colour normalisation, etc)
- Run ML models on **Cloud ML Engine (ease of use and latest)**: unsupervised deep learning models – allow algorithms to determin the accuracy of their predictions and make adjustments without an engineer stepping in.

high-resolution tissue images – uncompressed, proprietary format



Results:

- 12x faster image analysis with ML for improved patient outcomes
- Understandable outcomes guide clinicians for more effective treatments.

Compute Engine
Cloud Storage

Cloud ML Engine

https://cloud.google.com/customers/american-cancer-society/

https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2018/cancer-facts-and-figures-2018.pdf

Examples of Cloud-based Applications - II

Social Media Analysis: Queensland University of Technology (QUT)

Background:

- Social media has significantly change our daily life and style
- 500 mil tweets per day (5,787 tw/sec)
- Public opinions and sentiments about events are becoming more available and useful
- A group of world-leading researcher and academics in QUT Digital Media Research Centre are collecting, analysing, and visualising Aussie tweets in the QUT Digital Observatory project.

Aims:

- analyses of the response to and interaction with particular events on digital platforms – short term
- tracking of public communication and consumption of content – long term
- support to research projects that explore social media activities

https://cloud.google.com/customers/qut/
https://www.qut.edu.au/institute-for-future-environments/facilities/digital-observatory

Examples of Cloud-based Applications - II

Challenges:

- Storage infrastructure to support such a huge and live dataset:
  - A collection of tweets from all identified Australian accounts, collected since 2006.
  - A live data storage: 2.4 billion tweets -- growing at about 1.3 mil tw/day.
  - User information: 3.7 mil Australian Twitter accounts & 140k daily active users.
- Processing capacity:
  - Online queries, analysis and visualisation

Solutions:

- Google BigQuery: an analytics data warehouse to capture fast-growing datasets
- Google Reporting and Visualisation Tools

Results:

- Support projects with industry partners, such as analyses of social media activity around natural crises.
- Help deliver insights about the sharing and consumption of news sources on Twitter

Examples of Cloud-based Applications - III

QSearch: Creating marketing opportunities with social media analytics

QSearch

Background:

- Opinions, thoughts, reviews, comments are usually posted and shared on social media
- Behaviours and demographics can be discovered by pattern recognition
- Social networks are a rich repository of user experiences and information for agencies, businesses, and government organisations.

Aims:

Analysing social networks for product, brand, or topic trends to

- identify opportunities,
- spend campaign dollars wisely,
- manage crises effectively.

https://cloud.google.com/customers/qsearch/

Examples of Cloud-based Applications - III

QSearch

Challenges:

- Big data storage and analytic power
- Data visualisation kits
- Infrastructure expense control for a start-up company with 10 employees

Solutions:

- BigQuery analytics data warehouse
- App Engine application development
- visualisation tools

Results:

- Records only one error per one million requests for short-URL customer tracking service
- Enables the business to connect to and integrate social media search and measurement tools
- Measures employee productivity and controls access to customer projects
- Analyses **43.2 billion records in 8 hours**, not days

https://cloud.google.com/customers/qsearch/

Examples of Cloud-based Applications - IV

E-Government: Improving the customer experience and worker productivity by *Service NSW*

Background:

- Service NSW is a whole of New South Wales Government Service access point that provides online, and in-person offices that handle more than 800 types of transactions including:
  - Drivers licences, photo cards and vehicle and boat registration services
  - Functions of NSW Registry of Births Deaths & Marriages
  - Obtaining Opal cards for public transport, Seniors Card
  - Etc.
- In 130 offices across the entire state, there are many customer-service kiosks serving 7.89 million people (as of 2018).

https://www.service.nsw.gov.au/service-centre

https://cloud.google.com/customers/service-nsw/ http://www.population.net.au/population-of-new-south-wales/

Examples of Cloud-based Applications – IV

Challenges:

- Old kiosks' operating system lacked remote service features and required time-consuming hard drive repairs
- Maintaining all the kiosks across the entire state is expensive

Solutions:

- Chrome devices for kiosks
- G Suite: Gmail, Hangouts, Calendar, and Google+ for communication; Drive for storage; Docs, Sheets, Slides, Forms, and Sites for collaboration.

Results:

- Chromebases require only 5 percent of support hours needed by Microsoft devices
- Cloud tools eliminate the need for costly private WAN networks
- Reduced annual IT operational costs by 46 percent
- Helps Service NSW meet goal to perform 70 percent of government transactions digitally by 2019
- Improves employee productivity, collaboration and data security

https://cloud.google.com/customers/service-nsw/

Examples of Cloud-based Applications - V

HKTaxi: Using Google to deliver an intelligent, reliable taxi booking service

Background:

- Hongkong is one of most crowded cities in the world – 7.492 million (2019) and density is 7,134 p/km^2 (3 for Australia)
- Hailing taxis in Hongkong can be a frustrating experience -- particularly during demand peaks.
- Mobile phones and apps are making the experience easier and more transparent.

Aims:

- Highly efficient booking service: enable drivers and users locate each other
- Effective routes to destinations during peaking hours
- Less administration, more app development

https://cloud.google.com/customers/hktaxi/
https://hktaxiapp.com/

Examples of Cloud-based Applications - V

Challenges:

- Support electronic payment services
- Service rating: enable users to rate drivers (better service)
- 24*7 connections of drivers and users
- Support staff to resolve any issues.

Solutions:

- Google Maps SDK for iOS/Android
- Cloud Machine Learning Engine: to predict the attractiveness of orders to drivers

Results:

- Stable service for both drivers and users
- More focus on app development, less administration cost

https://cloud.google.com/customers/hktaxi/
https://hktaxiapp.com/

Examples of Cloud-based Applications - VI

GO-JEK: Using Machine Learning for forecasting and dynamic pricing

Background:

- Traffic congestion is a fact of life for most Indonesian residents.
- The nation's roads and associated infrastructure strains to support the country's 260 million people, about 10 million of whom reside in the capital, Jakarta.
- To minimise delays, Indonesians rely heavily on motorcycles, including motorcycle taxis, to travel to and from work or personal engagements.

Aims:

- Spatial and temporal data collection
- Customer behaviour analysis
- Routes optimisation for food delivery, taxi, etc.
- Estimations of arrivals and pricing

https://cloud.google.com/customers/go-jek/

Examples of Cloud-based Applications - VI

Challenges:

- Manage over 1 mil drivers, or hundreds of thousands of active drivers concurrently online
- Manage over 300,000 merchants (restaurants, private sellers) for food deliver
- Ping each driver and customer every 10 seconds across the whole country – 6 million pings per minute and 8 billion pings per day
- Deal with customer interaction – generate about 4TB to 5TB of data every day
- Dynamically match the right driver with the right request (deliver person or food from A to B)

Solutions:

- Google Maps Platform: core components in the framework to find out optimised routes and estimated times of arrival
- Big Data Package: Cloud Dataflow, Cloud Bigtable, and BigQuery form the basis of the company's platform.

Results:

- Supports 1 million motorcycle drivers with rapid access to riders and optimised routes
- Enables demand forecasting and pricing adjustments
- Positions business for international expansion

Market Position in 2022

AWS

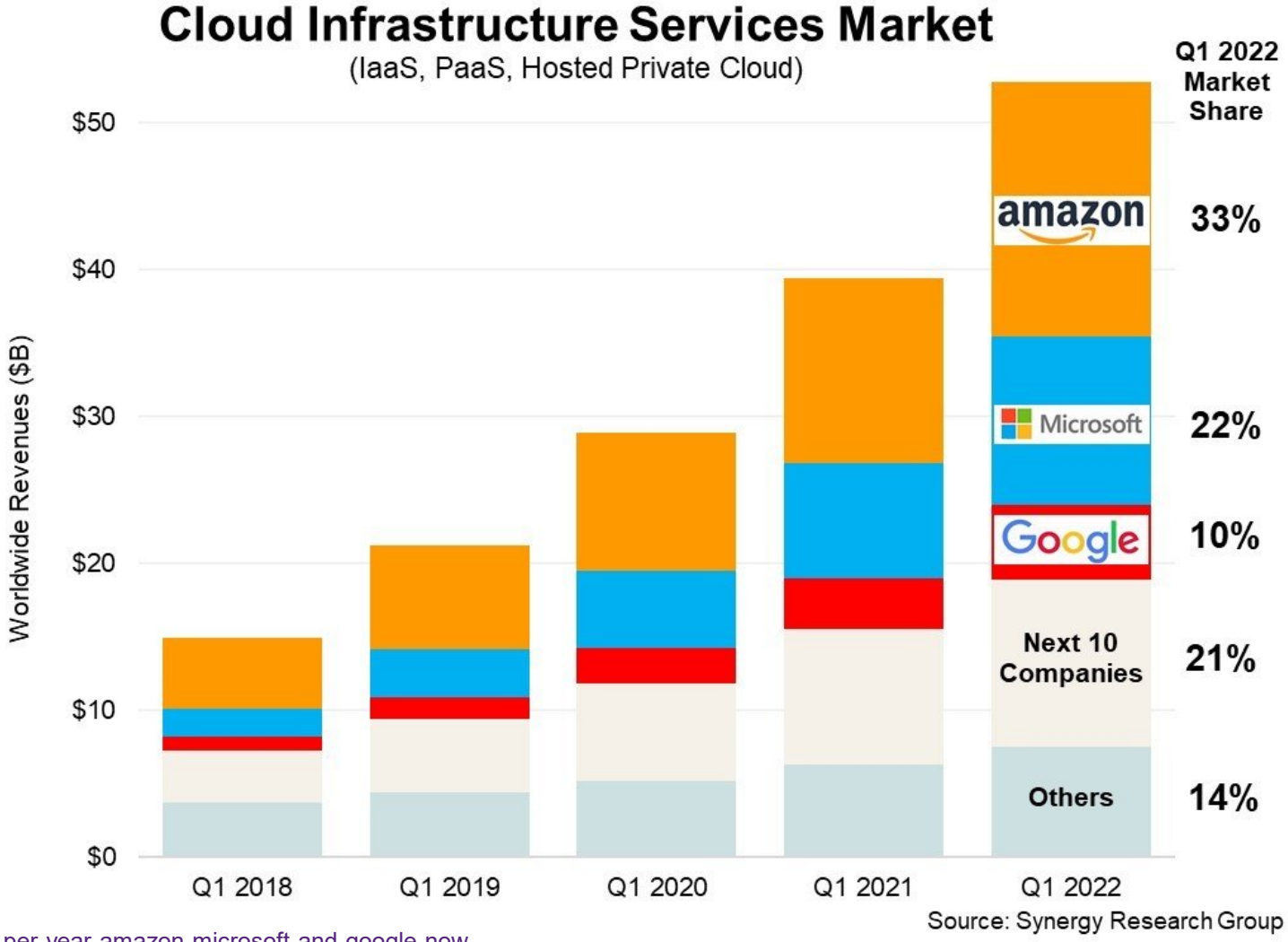
- More user-friendly to the developer
- Dominant position in CC.

MS Azure

- Strong connections with MS ecosystem (Office365, Windows, MS SQL, etc)
- Governments and Universities

Google GCP

- Scalable architect (K8s)
- Research-related products (ML, DS)



https://www.srgresearch.com/articles/huge-cloud-market-is-still-growing-at-34-per-year-amazon-microsoft-and-google-now-account-for-65-of-all-cloud-revenues

Products and Services under Google Cloud

There are over 90 products in nine categories:

Compute	Storage & Databases	Networking	Big Data	Cloud AI	Management Tools	Identity & Security	API Platform
App Engine	Cloud Storage	VPC - Virtual Private Cloud	BigQuery	Cloud AutoML	Stackdriver	Cloud Identity	Maps Platform
Compute Engine	Cloud SQL	Cloud Load Balancing	Cloud Dataflow	Cloud TPU	Cloud Deployment Manager	Cloud IAM	Apigee API Platform
Kubernetes Engine (GKE)	Cloud BigTable	Cloud Armor	Cloud Dataproc	Cloud Machine Learning Engine	Cloud Console	Cloud Identity-Aware Proxy	API Monetization

https://cloud.google.com/products/

Amazon Web Service v.s. Google Cloud Platform

Category	Service	AWS	GCP
Compute	IaaS	Amazon Elastic Compute Cloud	Compute Engine
	PaaS	AWS Elastic Beanstalk	App Engine
	Containers	Amazon Elastic Container Service	Google Kubernetes Engine
	Serverless Functions	AWS Lambda	Cloud Functions
	Managed Batch Computing	AWS Batch	N/A
Network	Virtual Networks	Amazon Virtual Private Cloud	Virtual Private Cloud
	Load Balancer	Elastic Load Balancer	Cloud Load Balancing
	Dedicated Interconnect	Direct Connect	Cloud Interconnect
	Domains and DNS	Amazon Route 53	Google Domains, Cloud DNS
	CDN	Amazon CloudFront	Cloud CDN

Amazon Web Service v.s. Google Cloud Platform

Category	Service	AWS	GCP
Storage	Object Storage	Amazon Simple Storage Service	Cloud Storage
	Block Storage	Amazon Elastic Block Store	Persistent Disk
	Reduced-availability Storage	Amazon S3 Standard-Infrequent Access, Amazon S3 One Zone-Infrequent Access	Cloud Storage Nearline
	Archival Storage	Amazon Glacier	Cloud Storage Coldline
	File Storage	Amazon Elastic File System	Cloud Filestore (beta)
Database	RDBMS	Amazon Relational Database Service, Amazon Aurora	Cloud SQL, Cloud Spanner
	NoSQL: Key-value	Amazon DynamoDB	Cloud Firestore, Cloud Bigtable
	NoSQL: Indexed	Amazon SimpleDB	Cloud Firestore
	Block Storage	Amazon Elastic Block Store	Persistent Disk

Amazon Web Service v.s. Google Cloud Platform

Category	Service	AWS	GCP
Big Data & Analytics	Batch Data Processing	Amazon Elastic MapReduce, AWS Batch	Cloud Dataproc, Cloud Dataflow
	Stream Data Processing	Amazon Kinesis	Cloud Dataflow
	Stream Data Ingest	Amazon Kinesis	Cloud Pub/Sub
	Analytics	Amazon Redshift, Amazon Athena	BigQuery
	Workflow Orchestration	Amazon Data Pipeline, AWS Glue	Cloud Composer
Application Services	Messaging	Amazon Simple Notification Service, Amazon Simple Queueing Service	Cloud Pub/Sub
Management Services	Monitoring	Amazon CloudWatch	Stackdriver Monitoring
	Logging	Amazon CloudWatch Logs	Stackdriver Logging
	Deployment	AWS CloudFormation	Cloud Deployment Manager

Amazon Web Service v.s. Google Cloud Platform

Category	Service	AWS	GCP
Machine Learning	Speech	Amazon Transcribe	Cloud Speech-to-Text
	Vision	Amazon Rekognition	Cloud Vision
	Natural Language Processing	Amazon Comprehend	Cloud Natural Language
	Translation	Amazon Translate	Cloud Translation
	Conversational Interface	Amazon Lex	Dialogflow Enterprise Edition
	Video Intelligence	Amazon Rekognition Video	Cloud Video Intelligence
	Auto-generated Models	N/A	Cloud AutoML (beta)
	Fully Managed ML	Amazon SageMaker	Cloud Machine Learning Engine

Summary

- **Business Drivers**:
  - Capacity Planning,
  - Cost Reduction,
  - Organizational Agility.
- **Technology Innovations:**
  - Clustering,
  - Grid Computing,
  - Virtualisation.
- **Concepts and Terminology:**
  - Cloud & Cloud Service
  - IT resource & Scaling
- **Cloud Characteristics:**
  - on-demand self-service usage,
  - ubiquitous access,
  - multitenancy,
  - elasticity,
  - measured usage,
  - resiliency.

- Cloud Delivery Models: SaaS, PaaS, and IaaS
- Cloud Deploy Models: Public/Private/Community/Hybrid Cloud
- Cloud-Enabling Technologies
  Broadband Networks and Internet Architecture
  Virtualisation Technology (VT)
  Data Centre Technology
  Web Technology
  Multitenant Technology
- Goals and Benefits
- Risks and Challenges
- Cloud-based Applications in the World

Reading Materials

1. “Cloud computing: concepts, technology & architecture”. Erl, Thomas, Ricardo Puttini, and Zaigham Mahmood. Pearson Education, 2013.
2. "The NIST Definition of Cloud Computing" (http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf). National Institute of Standards and Technology. Retrieved 24 July 2011.
3. "What is Cloud Computing?" (http://aws.amazon.com/what-is-cloud-computing/). Amazon Web Services. 2013-3-19. Retrieved 2013-3-20.
4. Qi Zhang, Lu Cheng, and Raouf Boutaba (2010) Cloud computing: state-of-the-art and research challenges, Journal of Internet Services and Applications, Springer, May 2010, Vol 1 Issue 1, pp 7-18.
5. Cynthia Harvey, 5 Things You Should Know About Hybrid Cloud, (https://www.informationweek.com/cloud/5-things-you-should-know-about-hybrid-cloud/d/d-id/1331818?) InformationWeek, 2018-05-17.
6. Hamdaqa, Mohammad. A Reference model for developing cloud applications (http://www.stargroup.uwaterloo.ca/~mhamdaqa/publications/A%20REFERENCEMODELFORDEVELOPINGCLOUD%20APPLICATIONS.pdf)
7. Amazon EC2 (http://aws.amazon.com/ec2/)

Tutorial Questions

1. What is the NIST (*National Institute of Standards and Technology*) definition of Cloud Computing?
2. What are the six essential characteristics of Cloud Computing? Moreover, for each characteristic, please make a brief introduction.
3. What are the four deployment models in cloud computing? Please read these reading materials and answer this question in detail.
4. Discuss the differences between cloud delivery models and make examples of delivery models.

Practical Activities

1. Get familiar with Google Cloud Platform
2. Initiate a VM instance and deploy web service