



香港城市大學
City University of Hong Kong

Lecture 1

Public cloud

EE4221 Cloud Computing Systems

Cloud services

Levels of cloud computing

Positioning some players

Business models

Deployment models

Selecting a cloud provider

Dr. Eric Wong and Dr. Anna Ruokonen
City University of Hong Kong



香港城市大學
City University of Hong Kong

- 1 Cloud services
- 2 Levels of cloud computing
- 3 Positioning some players
- 4 Business models
- 5 Deployment models
- 6 Selecting a cloud provider

Cloud services

Levels of cloud computing

Positioning some players

Business models

Deployment models

Selecting a cloud provider

What is cloud computing?



Stefan Tai, Karlsruhe Institute of Technology:

- cloud computing provides scalable, network-centric, abstracted IT infrastructure, platforms, and applications as on-demand services that are billed by consumption.
- Three important viewpoints
 - Business opportunities
 - Faster time-to-market and cost efficient innovation processes
 - Dynamically established open service and business networks
 - Internet scale service computing
 - Offering and using sophisticated platforms, infrastructure and business applications as modular (Web) services
 - Efficient management and utilization of systems
 - E.g. assigning resources based on demand and charging based on actual usage

Cloud services

Levels of cloud computing

Positioning some players

Business models

Deployment models

Selecting a cloud provider

Cloud services

Public cloud



香港城市大學
City University of Hong Kong

Cloud services

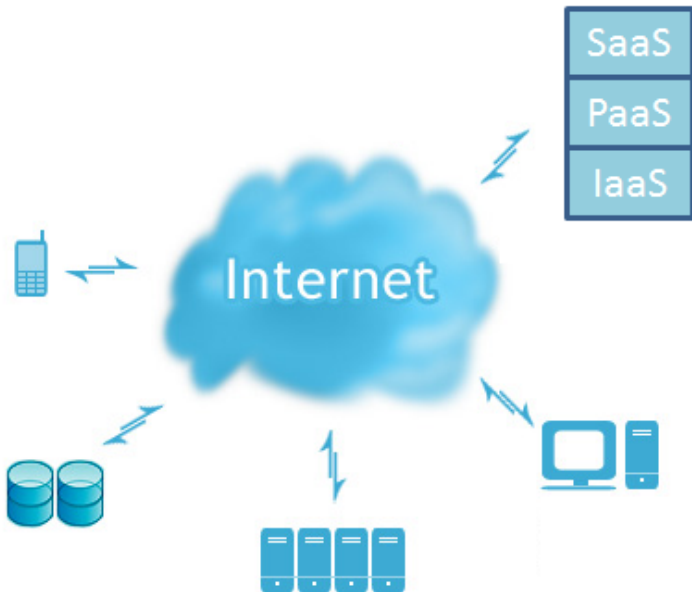
Levels of cloud computing

Positioning some players

Business models

Deployment models

Selecting a cloud provider



E.g., Amazon Web Services global infrastructure

Public cloud



香港城市大學
City University of Hong Kong

Cloud services

Levels of cloud computing

Positioning some players

Business models

Deployment models

Selecting a cloud provider





香港城市大學
City University of Hong Kong

Cloud services

Levels of cloud computing

Positioning some players

Business models

Deployment models

Selecting a cloud provider

- According to NIST (National Institute of Standards and Technology), the key characteristics of cloud computing are:
 - on-demand self-service,
 - instant access to services, i.e. cloud computing is ready on-demand
 - broad network access,
 - Cloud computing requires broad network access and services are location (and device) independent
 - Computing power is available to all kinds of (remote) clients through networking
 - resource pooling,
 - resources are shared among multiple users
 - resource allocation is done by self services accessible to the end users; users can choose an amount of resources or a service level on which they operate.



- According to NIST the key characteristics are: (cont.)
 - rapid elasticity and measured service
 - Requires good scalability
 - The amount of resources a user consumes can be changed rapidly
 - Depending on the cloud architecture, the user can either reserve an amount of resources to be consumed or just consume the resources when they are needed
 - Measured services is a key element that allows a user to be charged according to the consumption
- Cloud computing offers instantaneous, location independent, high quality, computing power to users and charges according to consumption.



香港城市大學
City University of Hong Kong

Cloud services

Levels of cloud computing

Positioning some players

Business models

Deployment models

Selecting a cloud provider

- Infrastructure as a Service (IaaS)
 - the lowest level of abstraction type of cloud computing models
 - offers a computer infrastructure - often a virtual hardware infrastructure - ready to use
- Platform as a Service (PaaS)
 - offers a software stack (that deals with parallelism and distribution etc.) and a complete abstraction to hardware
 - offers a computing platform and/or solution stack as a service, often consuming IaaS and sustaining SaaS applications
- Software as a Service (SaaS)
 - highest level of cloud computing
 - while IaaS and PaaS are aimed for SW developer, SaaS is often aimed directly to the end user
 - the services range from machine services to desktop applications

Infrastructure as a Service (IaaS)



- Type of service in which an organization offers the equipment used to support operations, such as storage, hardware, servers, and networking components.. i.e. computer infrastructure - often a virtual hardware infrastructure
- The service provider owns the equipment and is responsible for housing, running and maintaining it

IaaS: Some reasons why to use



- Scalability
 - Pay-as-you-go model, which allows you to scale up or down depending on your needs.
- Error Recovery
 - Your hardware and the data located on your IaaS provider and are housed in (hopefully) secure data centers that are mirrored and backed up.
- Time Back
 - With the infrastructure being hosted offsite, you can focus on value-added tasks such as planning and development.
- Efficient payment model
 - Instead of possibly unnecessarily big investments, e.g. in terms of not utilizing efficiently the hardware invested in, you only pay based on the actual use

Platform as a Service (PaaS)



- Adds a layer of abstraction over the actual infrastructure
- A PaaS platform at least provides the hosting infrastructure, development tools and deployment tools.
- Sandboxed, more locked-in access to interfaces and resources, but also more tasks handled by provider (automation, load balancing, billing etc.)
- Payment e.g. based on
 - Outgoing bandwidth
 - CPU time
 - Data storage space used (per time unit, e.g. month)
 - Recipients emailed

Platform as a Service (PaaS)



- Services build on the platform promoted in the PAAS providers store (e.g. Google Apps Marketplace) => PAAS revenue from both providing resources and helping to bill/sell services
- PAAS customers can be also sources of innovation and targets for acquirement => it is easy to integrate services, which already utilize the same platform

PaaS may provide support for...

- Application design
- Application development
- Testing
- Deployment
- Hosting
- Team collaboration
- Web service integration



PaaS may provide support for...

- Database integration
- Security
- Scalability
- Storage
- Persistence
- State management
- Application versioning
- Application instrumentation



Some reasons to use PaaS

- Jump start development: No headache of setting up hardware or software environment, development skills are still however needed
- No maintenance cost: Host company maintains the hardware and development environment
- Lower risk factor: If your project fails, just free the reserved resources and pay the usage bill.





香港城市大學
City University of Hong Kong

Cloud services

Levels of cloud computing

Positioning some players

Business models

Deployment models

Selecting a cloud provider

- C.f. SOA services*
- The utilisers of IaaS and PaaS
- Applications provided and used over the web
- Activities managed from central locations
- A one-to-many model
- As many business models as there are companies
 - Pricing model depends on the service: subscriptions, pay-per-use, pay-per-seat, freemium model etc.
- Customer charges must cover the risks of service disruption and possible service level agreement (SLA) compensations.

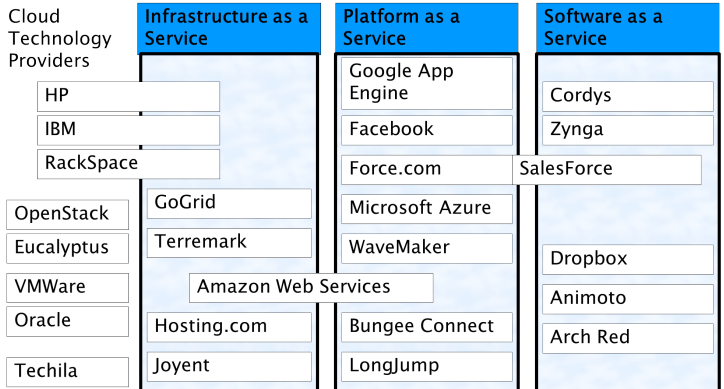
* <https://apprenda.com/library/architecture/soa-vs-saas-whats-the-difference/>

SaaS: some reasons why to use

- Scaling
- Costs
- Robustness
- Reliability
- Latency
- Promotion
- Distribution
- Economies of scale
- Marketing
- Exit strategy etc.
- Down sides:
 - no control over the operating system, hardware or network infrastructure,
 - difficult to change vendors

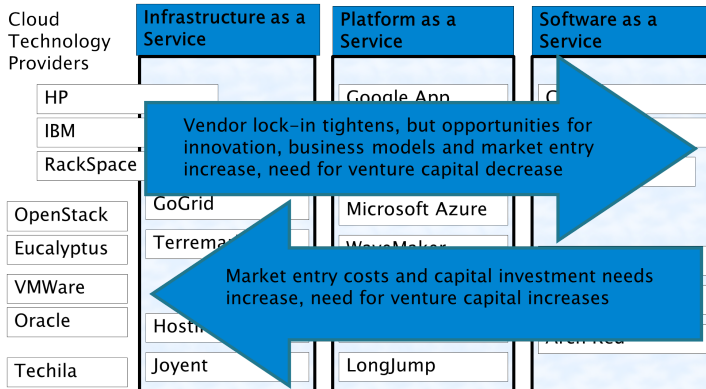


Positioning some players



[Huhtanen2010]

Positioning some players



[Huhtanen2010]



香港城市大學
City University of Hong Kong

Cloud services

Levels of cloud computing

Positioning some players

Business models

Deployment models

Selecting a cloud provider

- Many different business models
- Usually, charging is based on the resources and services used
 - Ingoing and outgoing bandwidth,
 - Transactions,
 - CPU time
 - Data storage space used (per time unit, e.g. month), etc.
- Custom units and different measuring methods make the comparison of the provider prices harder

Example: Amazon Web Services, EC2

Menu English Sign In Sign Up

PRODUCTS & SERVICES

- Amazon EC2 >
- Product Details >
- Instances >
- Pricing >
- Previous Generation Instances >
- Purchasing Options >

Amazon EC2 Spot Instances

Amazon EC2 Reserved Instances

Amazon EC2 Dedicated Instances

Developer Resources >

FAQs >

Amazon EC2 SLA >

AWS Management Portal for vCenter >

RELATED LINKS

- Windows Instances
- VM Import/Export
- Management Console
- Documentation
- Release Notes

Amazon EC2 Pricing

Pay only for what you use. There is no minimum fee. Estimate your monthly bill using [AWS Simple Monthly Calculator](#). The prices listed are based on the Region in which your instance is running. For a detailed comparison between On-Demand Instances, Reserved Instances and Spot Instances, see [Amazon EC2 Instance Purchasing Options](#).

Free Tier*

As part of AWS's [Free Usage Tier](#), new AWS customers can get started with Amazon EC2 for free. Upon sign-up, new AWS customers receive the following EC2 services each month for one year:

- 750 hours of EC2 running Linux, RHEL, or SLES t2.micro instance usage
- 750 hours of EC2 running Microsoft Windows Server t2.micro instance usage
- 750 hours of Elastic Load Balancing plus 15 GB data processing
- 30 GB of Amazon Elastic Block Storage in any combination of General Purpose (SSD) or Magnetic, plus 2 million I/Os (with Magnetic) and 1 GB of snapshot storage
- 15 GB of bandwidth out aggregated across all AWS services
- 1 GB of Regional Data Transfer

Get Started with AWS for Free

Create a Free Account

AWS Free Tier includes 750 hours of Linux and Windows t2.micro instances each month for one year. To stay within the Free Tier, use only EC2 Micro instances.

[View AWS Free Tier Details >](#)

- Amazon EC2 Spot Instances Amazon
- EC2 Reserved Instances Amazon
- EC2 On-Demand Instances

Example: Amazon Web Services, EC2



	Linux/UNIX Usage	Windows Usage
General Purpose - Current Generation		
m3.medium	\$0.008 per Hour	\$0.059 per Hour
m3.large	\$0.016 per Hour	\$0.117 per Hour
Compute Optimized - Current Generation		
c3.large	\$0.016 per Hour	\$0.104 per Hour
c3.xlarge	\$0.032 per Hour	\$0.198 per Hour
GPU Instances - Current Generation		
g2.2xlarge	\$0.0641 per Hour	\$0.1041 per Hour
Memory Optimized - Current Generation		
r3.large	\$0.016 per Hour	\$0.174 per Hour
r3.xlarge	\$0.032 per Hour	\$0.282 per Hour
Storage Optimized - Previous Generation		
hi1.4xlarge	\$0.129 per Hour	\$0.4811 per Hour
Micro Instances		
t1.micro	\$0.0031 per Hour	\$0.0061 per Hour

Example: Google Cloud Platform



Google Cloud Platform

[Why Google](#) [Products](#) [Solutions](#) [Customers](#) [Documentation](#) [Support](#) [Partners](#) [Blog](#)

Google Cloud Platform Pricing Calculator

Enter what you need to run in the cloud below, and we'll compute the cost.
Just want the pricing data instead? [Click Here](#).



Google Compute Engine

Servers

Number of servers * ⓘ

What are these servers for? ⓘ

Operating System ⓘ Free: Debian, CentOS, OpenSUSE, CoreOS ⌵

Instance type ⓘ f1-micro (Cores: shared, RAM: 0.60 GB) ⌵

Or

CPU Cores * ⓘ shared ⌵

Memory (RAM) * ⓘ 0.6 GB ⌵

Datacenter location ⓘ United States ⌵

Public cloud



香港城市大學
City University of Hong Kong

[Cloud services](#)

[Levels of cloud computing](#)

[Positioning some players](#)

[Business models](#)

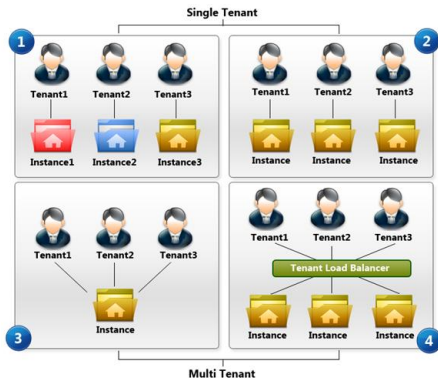
[Deployment models](#)

[Selecting a cloud provider](#)



- Public cloud
 - resources are dynamically provisioned on a fine-grained, self-service basis over the Internet
- Community cloud
 - pooled resource cloud that combines the resources of multiple community users
- Private cloud
 - owned by corporations which uses them internally may be
 - completely behind the firewall so that no external access is possible
- Hybrid cloud
 - consists of multiple internal and/or external providers
 - a combination of public and private clouds: utilizes public cloud for non-critical information handling and private cloud for dealing critical information
 - public cloud is also used to balance workload (high peaks are run on public cloud)

SAAS: Maturity levels



[<https://thecloudworld.wordpress.com/tag/maturity-model/>]

- Level 1: Ad-hoc or custom
- Level 2: Configurable
- Level 3: Configurable, multi-tenant-efficient
- Level 4: Scalable, configurable multitenant-efficient

How to select a cloud provider?



- Ensure that it can be configured to meet the changing needs
- Cloud vendor should provide high SLA guarantees for their services
- Learn about the cloud provider's key risks and performance indicators
- Learn how easily you can switch to another provider if needed
- Check the quality of customer service and support

Service Level Agreement (SLA)



- SLA is a contract between service provider and consumer
- Defines the service conditions and penalties if conditions are not met
- Used to define the level(s) of service being sold
- Usually given in plain language terms



Google Apps Service Level Agreement

Google Apps SLA. During the Term of the applicable Google Apps Agreement (the "Agreement"), the Customer does not meet the Google Apps SLA, and if Customer meets its obligations under this Google Apps SLA, Customer does not meet the Google Apps SLA.

Definitions. The following definitions shall apply to the Google Apps SLA.

"Downtime" means, for a domain, if there is more than a five per cent user error rate. Downtime

"Google Apps Covered Services" means the Gmail, Google Calendar, Google Talk, Google Docs, Google Drive, Google Mail Voice or Video Chat components of the Service.

"Monthly Uptime Percentage" means total number of minutes in a calendar month minus the number of minutes of downtime.

"Service" means the Google Apps for Business service (also known as Google Apps Premier Edition), Google Apps Education Edition or Google Apps Vault (as applicable) provided by Google to Customer.

"Service Credit" means the following:

Monthly Uptime Percentage	Days of Service added to the end of the Service term (or monetary credit equal to the value of days of service for monthly postpay billing customers), at no charge to Customer
< 99.9% - >= 99.0%	3
< 99.0% - >= 95.0%	7
< 95.0%	15

Cloud services

Levels of cloud computing

Positioning some players

Business models

Deployment models

Selecting a cloud provider



香港城市大學
City University of Hong Kong

Cloud services

Levels of cloud computing

Positioning some players

Business models

Deployment models

Selecting a cloud provider

- Financial stability
 - Obtain records that can prove the financial stability of the cloud vendor.
- Make sure that the security policies and measures that the cloud vendor has established correspond with your security requirements.
- Find out who is allowed to access to your data.
- The IaaS vendor should allow you to assess the quality of the back end systems.
- Find out if cloud provider's data center infrastructure is highly ranked (Tier 2 or Tier 3 standards).



香港城市大學
City University of Hong Kong

Cloud services

Levels of cloud computing

Positioning some players

Business models

Deployment models

Selecting a cloud provider

- Ensure that the infrastructure can be configured to meet the changing needs.
- Cloud vendor should provide high Service Level Agreement (SLA) guarantees for their services.
- Learn about the cloud provider's key risks and performance indicators.
- Learn if you can easily switch to another provider if the need arises.
- Check the quality of customer service and support.

The Uptime data center tier standards



Tier level **Data center requirements**

- | | |
|---------------|---|
| <i>Tier 1</i> | <ol style="list-style-type: none">1) A single, non-redundant distribution path serving IT equipment.2) Non-redundant capacity components. |
| <i>Tier 2</i> | <ol style="list-style-type: none">1) All Tier 1 requirements.2) Redundant capacity components. |
| <i>Tier 3</i> | <ol style="list-style-type: none">1) All Tier 1 and 2 requirements.2) Multiple independent distribution paths serving IT equipment. Generally, only one distribution path serves equipment at any given time.3) All IT equipment is dual-powered and fully compatible within the topology of a site's architecture. |
| <i>Tier 4</i> | <ol style="list-style-type: none">1) All Tier 1, 2 and 3 requirements.2) The facility is fully fault-tolerant, through electrical, storage and distribution networks.3) All cooling equipment is independently dual-powered, including HVAC systems. |



香港城市大學
City University of Hong Kong

Cloud services

Levels of cloud computing

Positioning some players

Business models

Deployment models

Selecting a cloud provider

- PaaS adds an additional layer on top of IaaS
- Cloud vendor should provide high SLA guarantees for their services
- Learn about the cloud provider's key risks and performance indicators
- Learn how easily you can switch to another provider if needed
- Check the quality of customer service and support

- **Privileged user access** Sensitive data processed outside the enterprise brings a significant level of risk.
- **Regulatory compliance** Organizations are ultimately responsible for the security and integrity of their own data, even when it is held by a service provider.
- **Data location** When you use the cloud, you probably won't know exactly where your data is hosted.
- **Data segregation** Customers data in the cloud is typically stored in one shared environment.
- **Recovery** Even if you don't know where your data is, a cloud provider should tell you what will happen to your data and service in case of a disaster. Any offering that does not replicate the data and application infrastructure across multiple sites is vulnerable to a total failure.
- **Long-term viability** Be sure your data will remain available even if your cloud computing company doesn't exist any more.





香港城市大學
City University of Hong Kong

Cloud services

Levels of cloud computing

Positioning some players

Business models

Deployment models

Selecting a cloud provider

 Edited by C. Daniele and H. Giles
Assessing the Security Risks of Cloud Computing.
The European Network and Information Security Agency (ENISA), 2009.

 *Amazon Web Services, Amazon EC2.*
<http://aws.amazon.com/ec2/>.

 *Google Cloud Platform.*
<https://cloud.google.com/products/calculator/>.

 J. Rhoton and R. Haukioja
Cloud Computing Architected, Solution design handbook.
Recursive Press, 2013 edition.