INFS803
Cloud Computing

Introduction

My Background and Contact Details

- Dr Jian Yu
- Assoc. Prof. @AUT
- Lab Director: AUT UbiWeb Research Lab https://ubiweb.aut.ac.nz/
- PhD from Peking University
- Main teaching areas: web technologies, web and cloud computing, operating systems
- Research areas: Graph representation learning and graph neural networks, Recommender systems, Internet and cloud computing, Complex networks

- Office in WZ Level 11
- jian.yu@aut.ac.nz
- https://academics.aut.ac.nz/jian.yu

Course Assessment

- 50% on team project (team size: 3-4 people, to be released in Week 3)
- 50% on online theory test (Week 12)

Course Contents

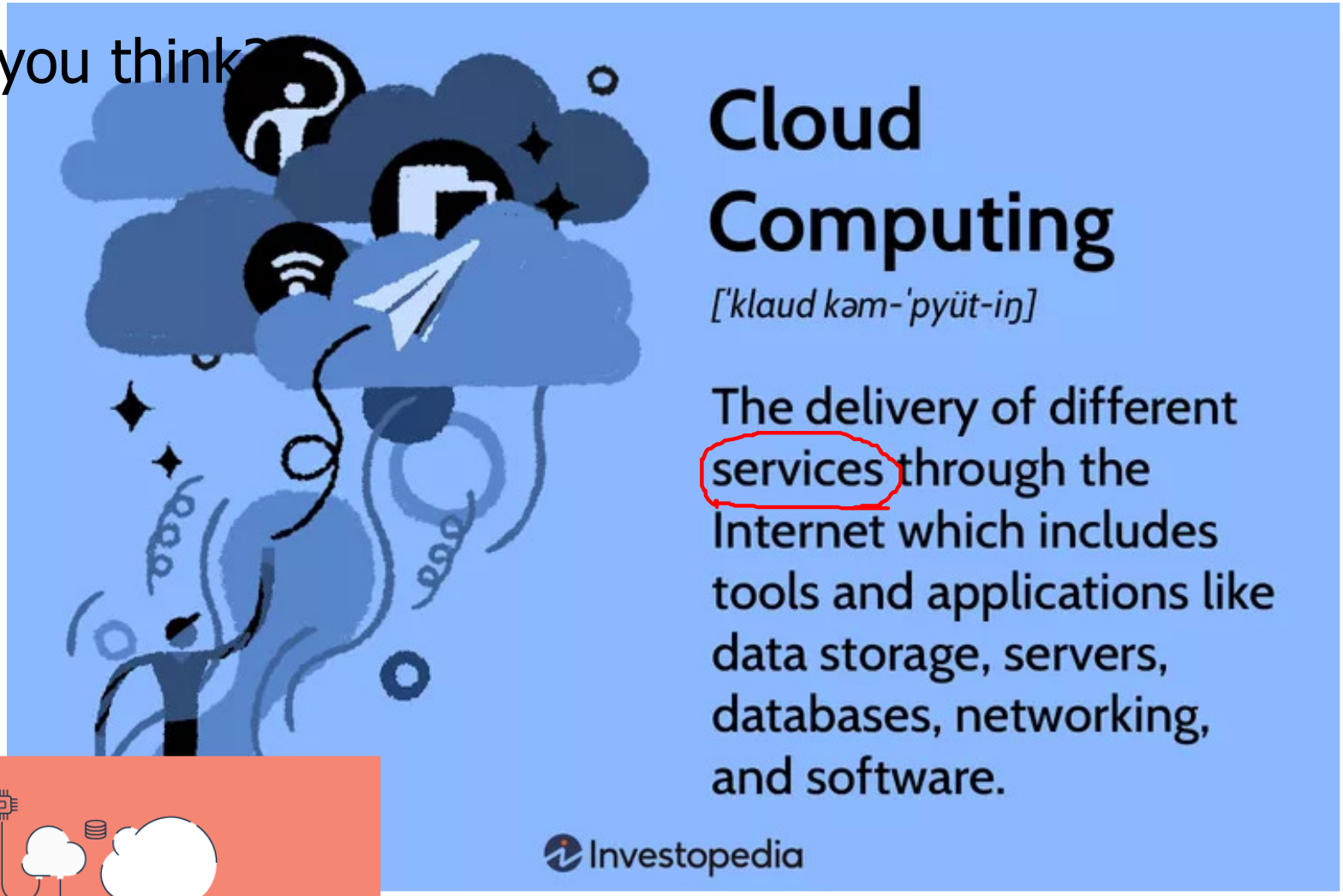
1. Introduction to Cloud Computing
2. Cloud and Distributed Computing Infrastructure
3. Cloud Web Service Protocols
4. Lab: Amazon Web Services
5. Complex Networks and the Cloud
6. Cloud Resource Virtualization
7. Cloud Service Workflow Management: BPMN and Petri Nets
8. Lab: Developing BPMN and Petri Nets Processes
9. Intelligence in the Cloud: Semantic Web, knowledge graph, Graph representation learning and Recommender Systems
10. Containerization and Security in the Cloud (Guest Lecture)

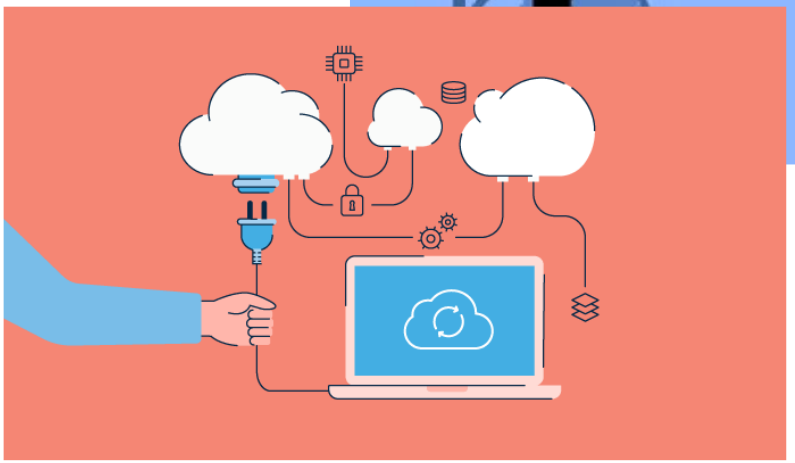
Lecture Contents

- Cloud Computing: concept and evolution history
- Attributes and benefits
- Delivery and deployment models
- Enabling technologies
  - Virtualization
  - Web service and service-oriented architecture
  - service flows and workflows
  - Web 2.0 and mashup
- Cloud computing challenges
- Parallel Computing and Distributed Systems (brief)

What is Cloud Computing?

- What do you think?





What is Cloud Computing?

- **Cloud computing**[1] is the on-demand availability of computer system resources, especially data storage (cloud storage) and computing power, without direct active management by the user.[2] Large clouds often have functions distributed over multiple locations, each of which is a data center. Cloud computing relies on sharing of resources to achieve coherence and typically uses a "pay as you go" model, which can help in reducing capital expenses but may also lead to unexpected operating expenses for users.[3] https://en.wikipedia.org/wiki/Cloud_computing

- *"Simply put, cloud computing is the delivery of computing **services** – servers, storage, databases, networking, software, analytics and more – over the Internet ("the cloud"). Companies offering these computing services are called cloud providers and typically charge for cloud computing services based on usage, similar to how you're billed for gas or electricity at home."* *https://azure.microsoft.com/en-gb/overview/what-is-cloud-computing/*

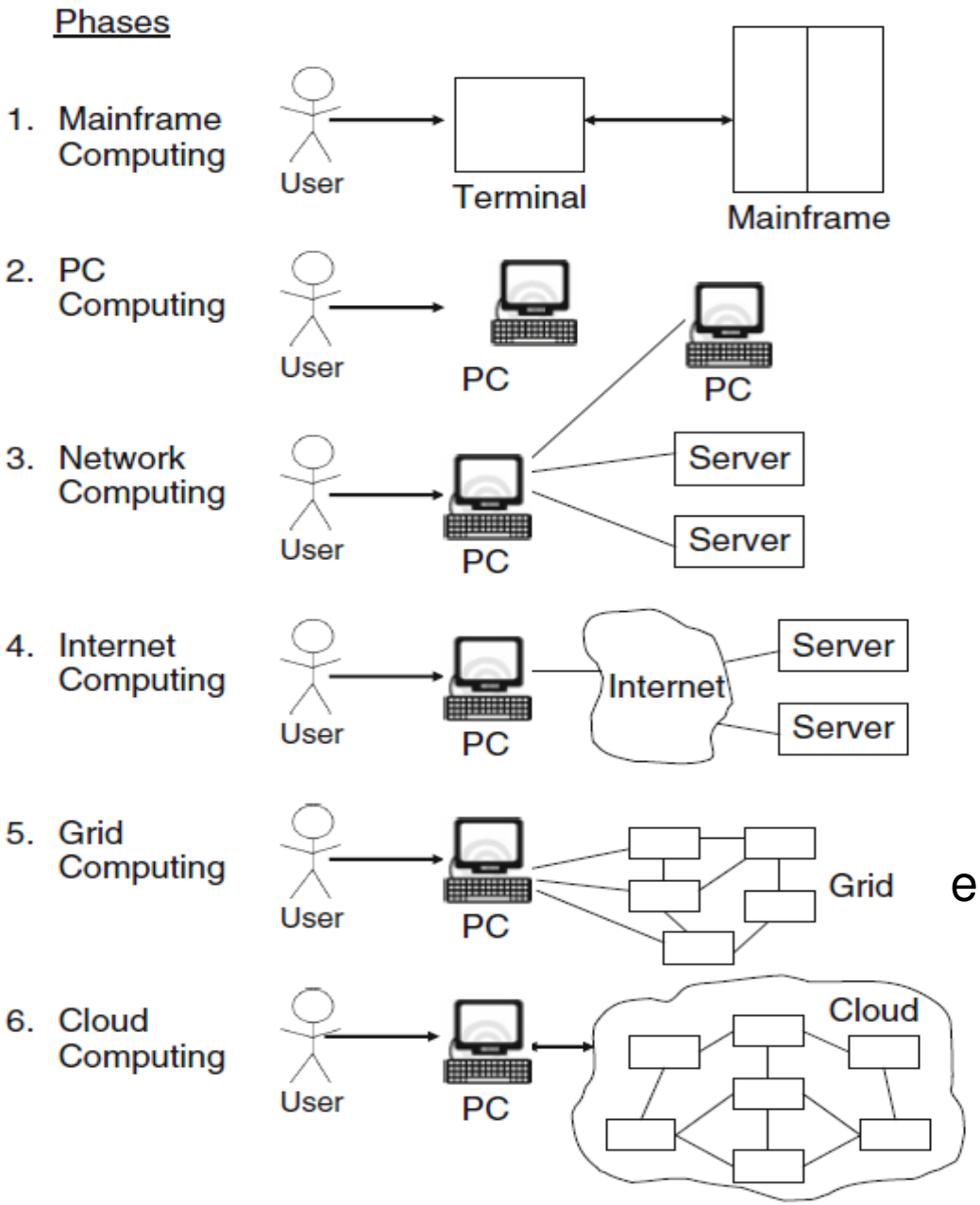
Key concepts: Service

- What is a service? What's the difference between a service and a product?

  Services are a means of delivering value to customers by facilitating outcomes customers want to achieve without the ownership of specific costs and risks.

- Cloud service/web service
  - Amazon cloud services: Commonwealth bank saves millions from cloud services IT budget on infrastructure now down from 75% to 26%

Computing phases: from mainframe to cloud

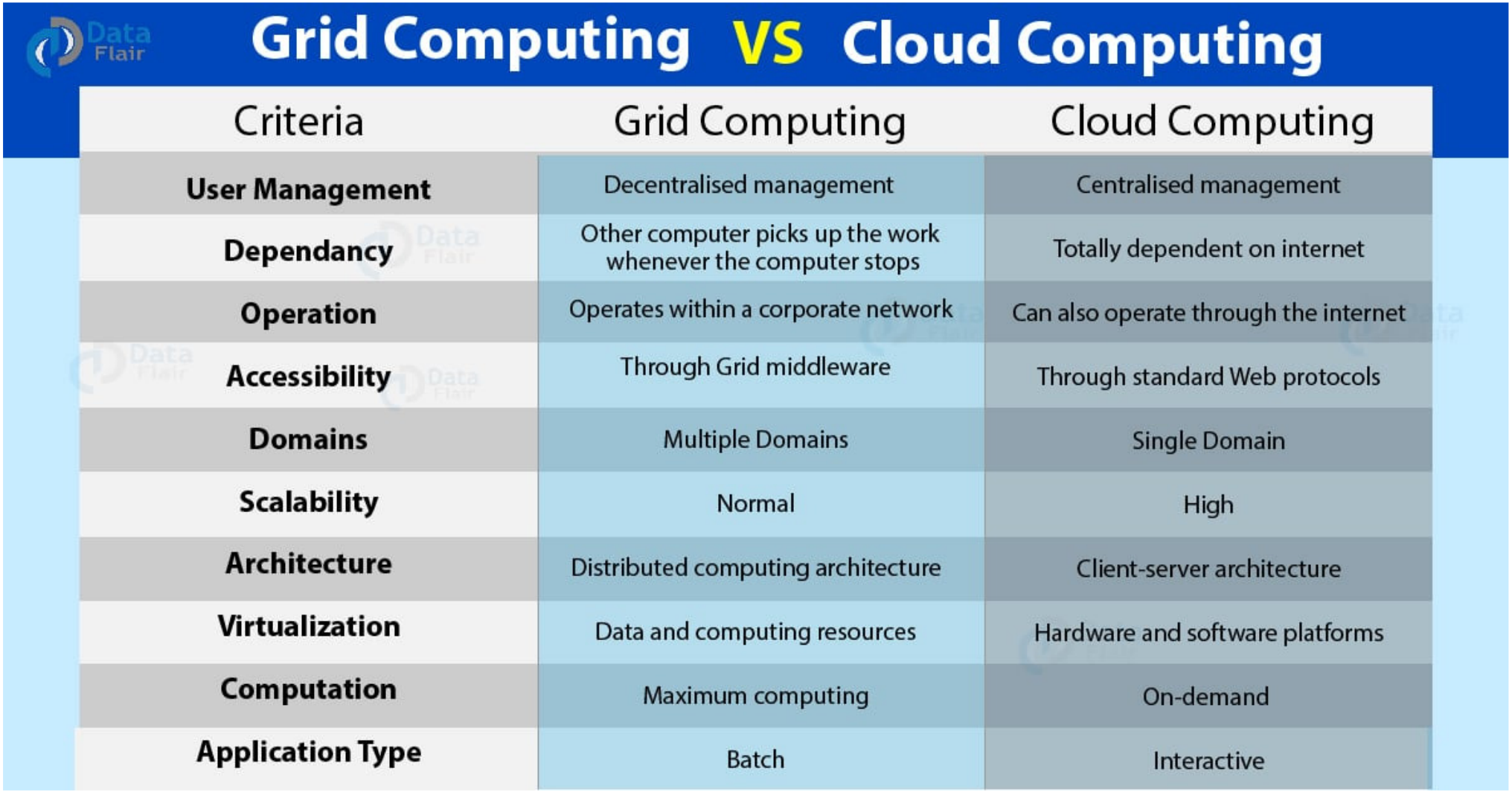


early 1990s

Around 2000s

A grid/cluster of computers (visible the users) collaborate and run independently to solve a scientific problem;
while a cloud is treated as a virtual computer.

Grid vs Cloud

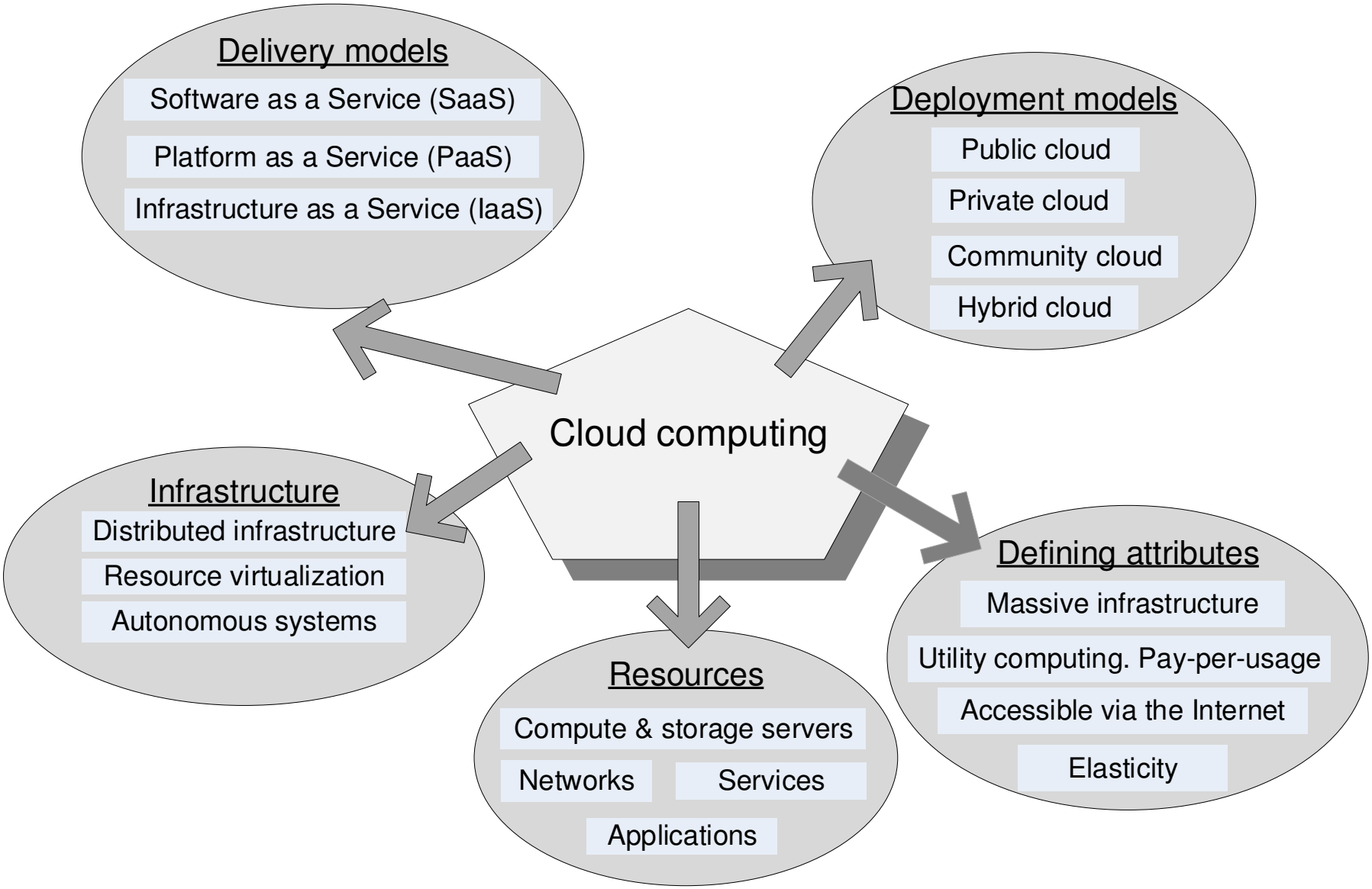


Criteria	Grid Computing	Cloud Computing
User Management	Decentralised management	Centralised management
Dependancy	Other computer picks up the work whenever the computer stops	Totally dependent on internet
Operation	Operates within a corporate network	Can also operate through the internet
Accessibility	Through Grid middleware	Through standard Web protocols
Domains	Multiple Domains	Single Domain
Scalability	Normal	High
Architecture	Distributed computing architecture	Client-server architecture
Virtualization	Data and computing resources	Hardware and software platforms
Computation	Maximum computing	On-demand
Application Type	Batch	Interactive

A grid/cluster of computers (visible the users) collaborate and run independently to solve a scientific problem; while a cloud is treated as a virtual computer.

https://data-flair.training/blogs/grid-computing-vs-cloud-computing/

Cloud Computing Attributes, Models, Resources



Cloud computing - Characteristics

"*Cloud Computing offers on-demand, scalable and elastic computing (and storage services). The resources used for these services can be metered and users are charged only for the resources used.*" from the Book

Shared Resources and Resource Management:

1. Cloud uses a shared pool of resources
2. Uses Internet techn. to offer **scalable** and **elastic** services.
3. The term "**elastic computing**" refers to the ability of **dynamically** and **on-demand** acquiring computing resources and supporting a variable workload.
4. Resources are metered and users are charged accordingly.
5. It is more cost-effective due to **resource-multiplexing.** Lower costs for the cloud service provider are passed to the cloud users

Cloud computing (cont'd)

Data Storage:

6. Data is stored:
   - in the "cloud", in certain cases closer to the site where it is used.
   - appears to the users as if stored in a location-independent manner.
7. The data storage strategy can increase reliability, as well as security, and can lower communication costs.

Management:

8. The maintenance and security are operated by service providers.
9. The service providers can operate more efficiently due to specialisation and centralisation.

Cloud Computing Advantages

1. Resources, such as CPU cycles, storage, network bandwidth, are **shared.**

2. When multiple applications share a system, their peak demands for resources are not synchronised thus, **multiplexing** leads to a higher resource utilization.

3. Resources can be **aggregated** to support data-intensive applications.

4. Data sharing facilitates **collaborative** activities. Many applications require multiple types of analysis of shared data sets and multiple decisions carried out by groups scattered around the globe.

Cloud Computing Advantages

5. Eliminates the **initial investment costs** for a private computing infrastructure and the maintenance and operation costs.

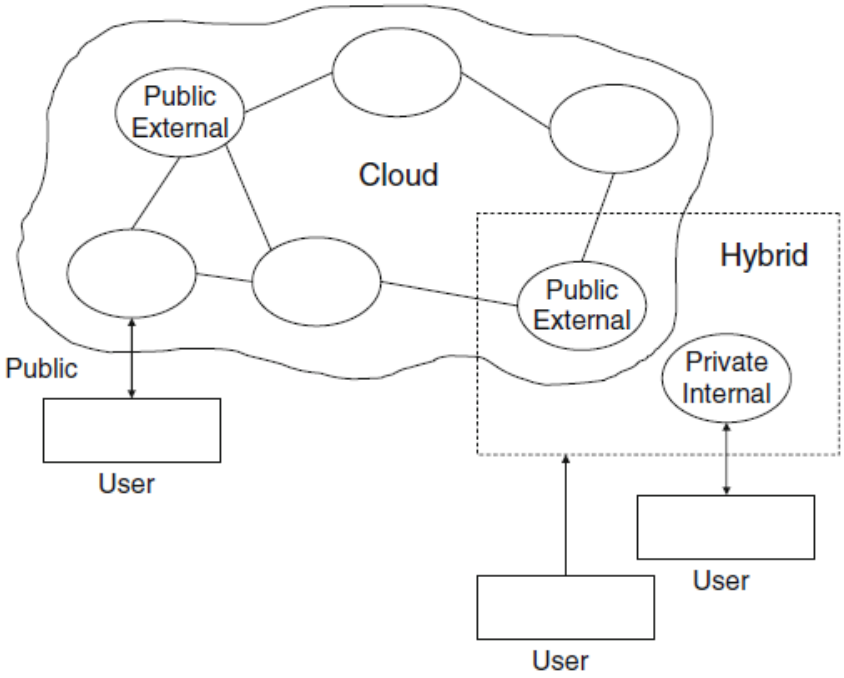
6. **Cost reduction**: concentration of resources creates the opportunity to pay as you go for computing.

7. **Elasticity**: the ability to accommodate workloads with very large peak-to-average ratios.

8. **User convenience**: virtualization allows users to operate in familiar environments rather than in idiosyncratic ones.

Types of clouds – deployment models

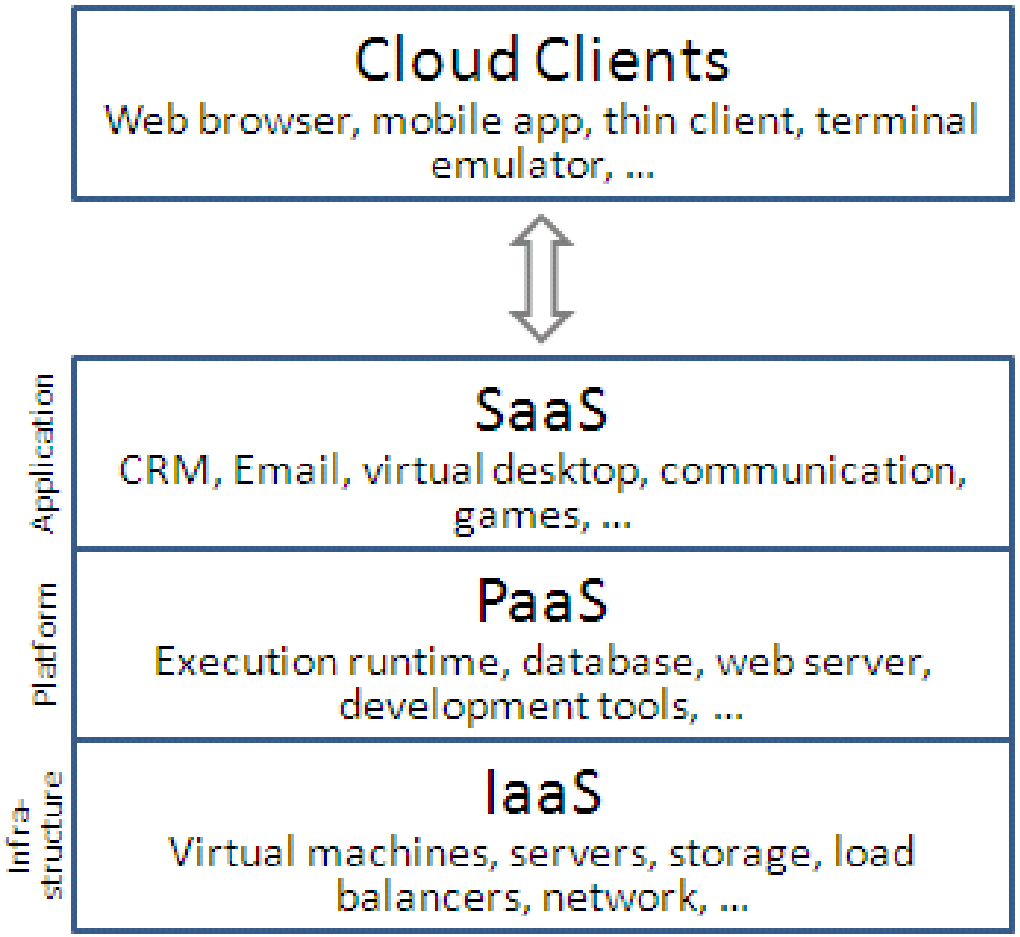
1. **Public Cloud** - the infrastructure is made available to the general public or a large industry group and is owned by the organization selling cloud services.
2. **Private Cloud** – the infrastructure is operated solely for an organization.
3. **Hybrid Cloud** - composition of two or more Clouds (public, private, or community) as unique entities but bound by a standardised technology that enables data and application portability.



- **Other types:** e.g., Community/Federated Cloud - the infrastructure is shared by several organizations and supports a community that has shared concerns.

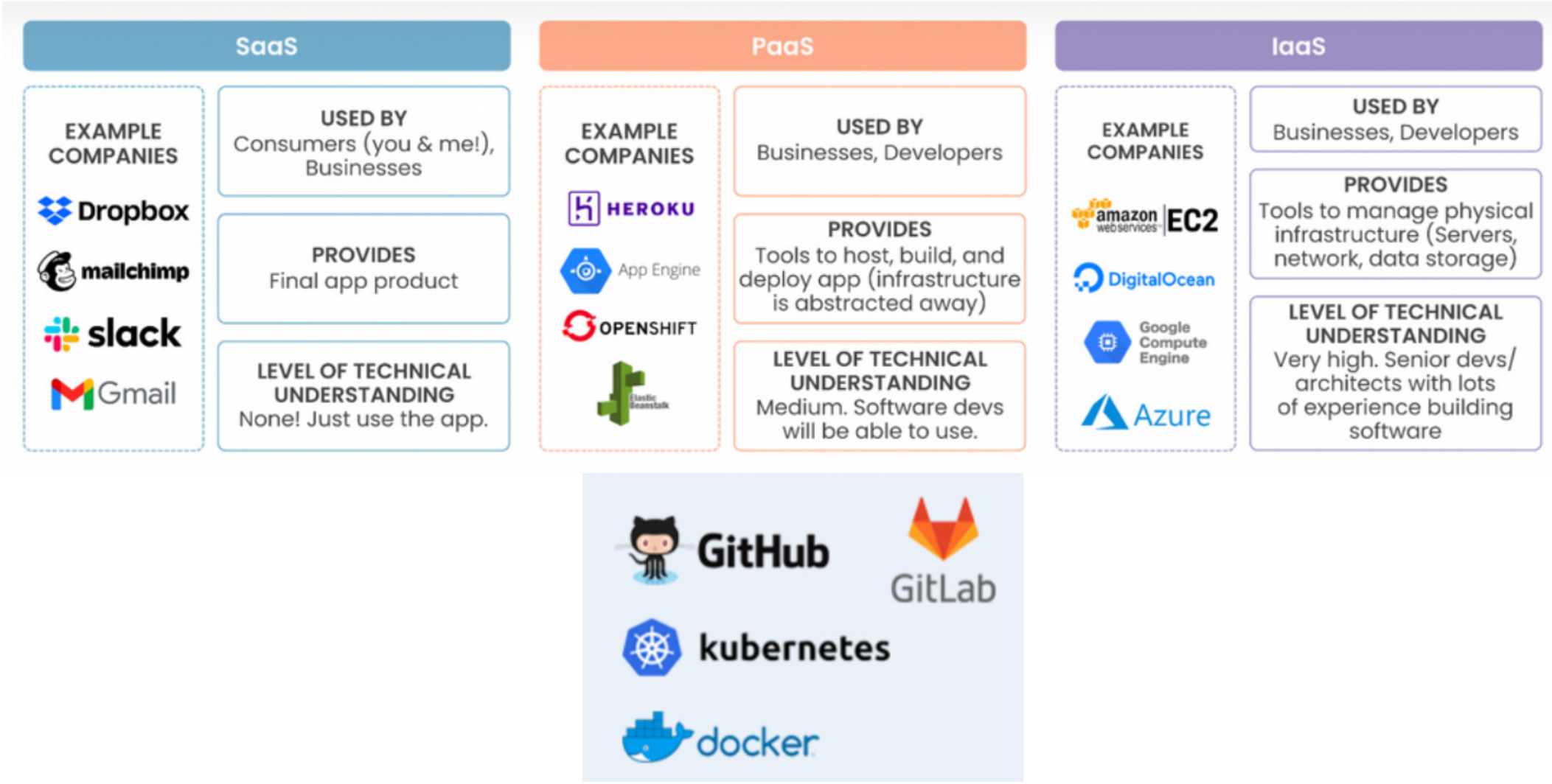
Cloud Delivery Models

1. **Software as a Service (SaaS)** (high level)
2. **Platform as a Service (PaaS)**
3. **Infrastructure as a Service (IaaS)** (low level)



What delivery model is Xbox Cloud Gaming/Streaming?

source Wikipedia



https://www.skiplevel.co/blog/paas-saas-iaas

Infrastructure-as-a-Service (IaaS)

- Infrastructure is compute resources, CPU, VMs, storage, etc

- The user is able to deploy and run arbitrary software, which can include operating systems and applications.
- The user does not manage or control the underlying Cloud infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of some networking components, e.g., host firewalls.

- Services offered by this delivery model include: storage, computing hardware, virtual instances, load balancing, Internet access, and bandwidth provisioning.

- Example: Amazon EC2

Platform-as-a-Service (PaaS)

- Allows a cloud user to deploy consumer-created or acquired applications using programming languages and tools supported by the service provider.

- The user:
  - Has control over the deployed applications and, possibly, application hosting environment configurations.
  - Does not manage or control the underlying Cloud infrastructure including network, servers, operating systems, or storage.

- Examples: docker, github, MongoDB, AWS Web hosting

Software-as-a-Service (SaaS)

- Applications are supplied by the service provider.
- The user does not manage or control the underlying Cloud infrastructure or individual application capabilities.

- Services offered include:
  - Enterprise services such as: workflow management, communications, digital signature, customer relationship management (CRM), desktop software, financial management, geo-spatial, and search.

- Examples: Gmail, Salesforce

The Three delivery models of Cloud Computing

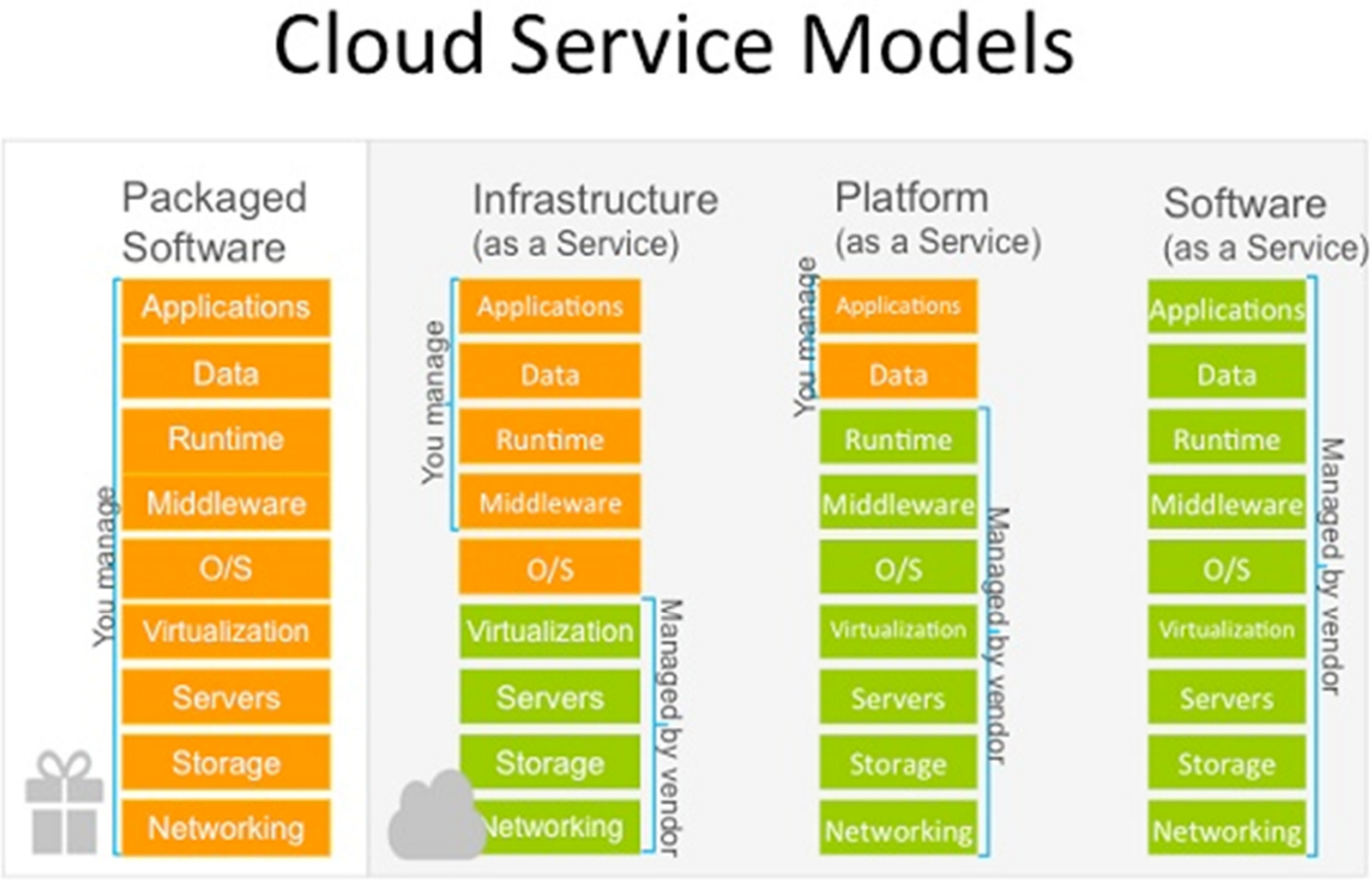


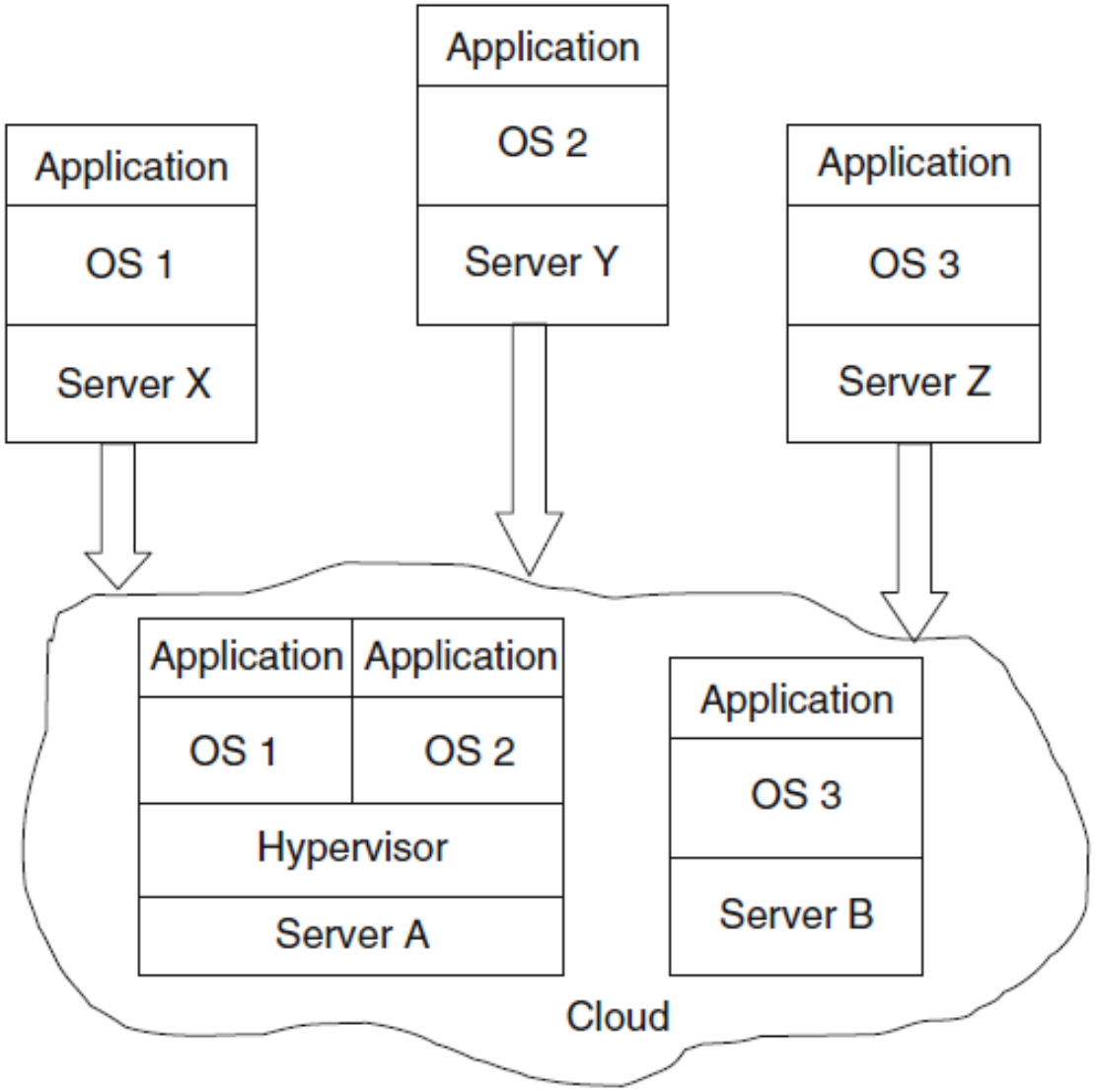
Figure 1. **Source: Microsoft Azure**

Enabling Technologies

- Virtualization
- Web service and Service-Oriented Architecture (SOA)
- service flows and workflows
- Web 2.0 and mashup

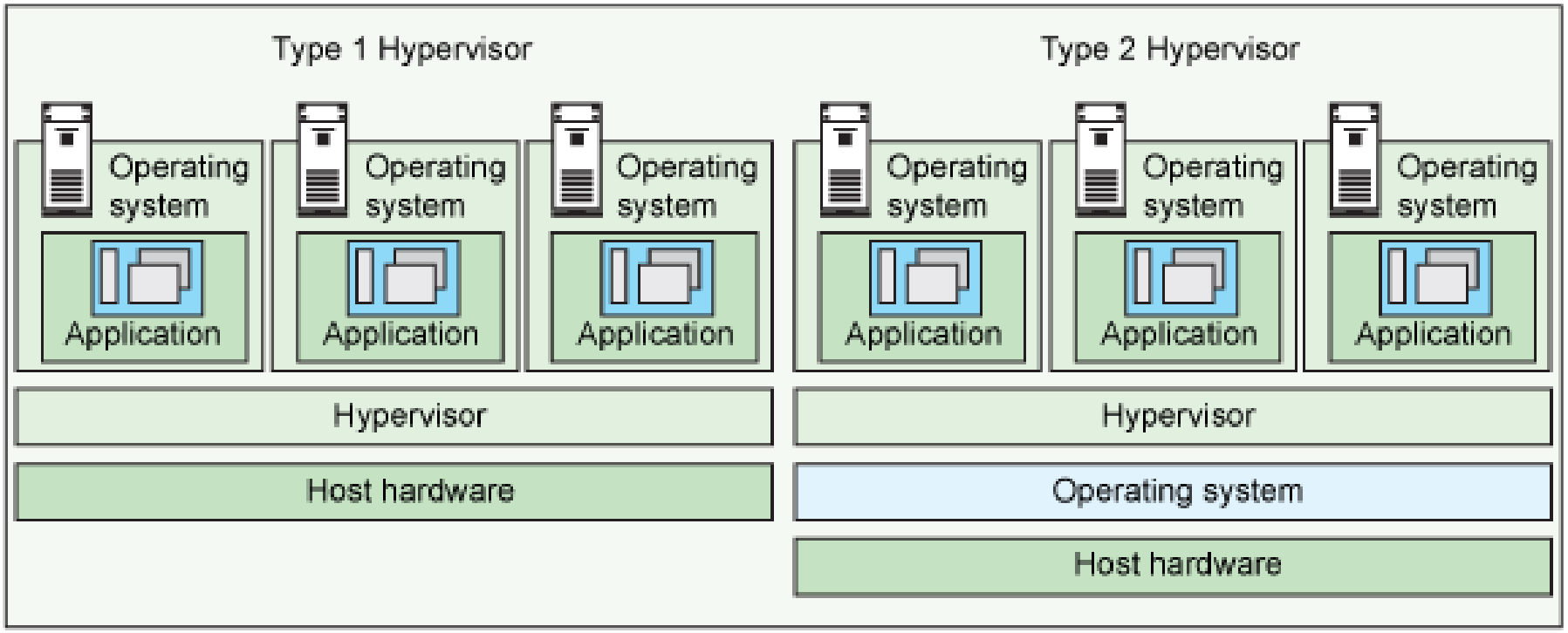
Virtualization

- Virtual memory
- Virtual machine
  - VMWare, Xen
- Virtual networks
  - VPN



Two types of hypervisors

- Type 1 hypervisor: hypervisors run directly on the system hardware – A "bare metal" embedded hypervisor,
- Type 2 hypervisor: hypervisors run on a host operating system that provides virtualization services, such as I/O device support and memory management.



https://vapour-apps.com/what-is-hypervisor/

Web Service and SOA

- Cloud services are typically designed as Web services

What are Web Services?

Web services perform encapsulated business functions such as:

1. a self-contained business task, such as a funds withdrawal or funds deposit service;
2. a full-fledged business process, such as the automated purchasing of office supplies;
3. an application, such as a life insurance application or demand forecasts and stock replenishment;
4. or a service-enabled resource, such as access to a particular back-end database containing patient medical records.

What are Web Services?

- **W3C definition:**
  - *[a Web service] has an interface described in a machine-processable format (specifically WSDL). Other systems interact with the Web service in a manner prescribed by its description using SOAP (Simple Object Access Protocol) messages, typically conveyed using HTTP with an XML serialization in conjunction with other Web-related standards.*
  - *REST-compliant Web services, in which the primary purpose of the service is to manipulate representations of Web resources using a uniform set of stateless operations.*

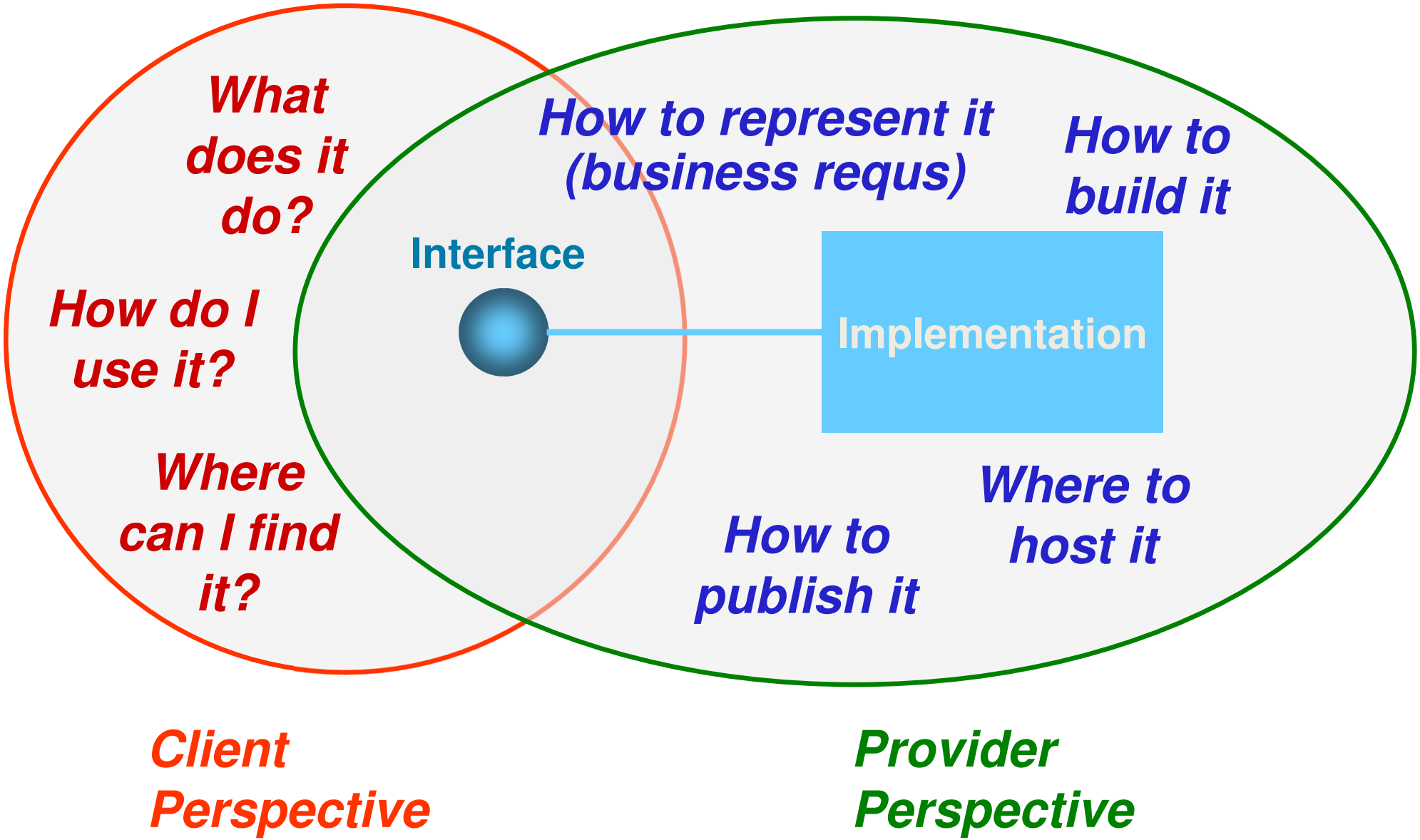
What are Web Services?

- **Can be mixed and matched to create a complete process:**
  - **Enable integration with few human interaction, e.g.,** create complete enterprise processes, such as supply chain management, procurement, logistics, etc.
  - **Both new & extensions to existing applications**
- **Available to a variety of clients (platform independent)**
- **Pricing (rating) model determines subscriber rates based on subscription and usage events, e.g., calculate charges for services based on the quality of the service (QoS).**
- **Billing model:**
  - **"pay per use", "lease it", "pay for it" basis.**

Service Interface & Implementation

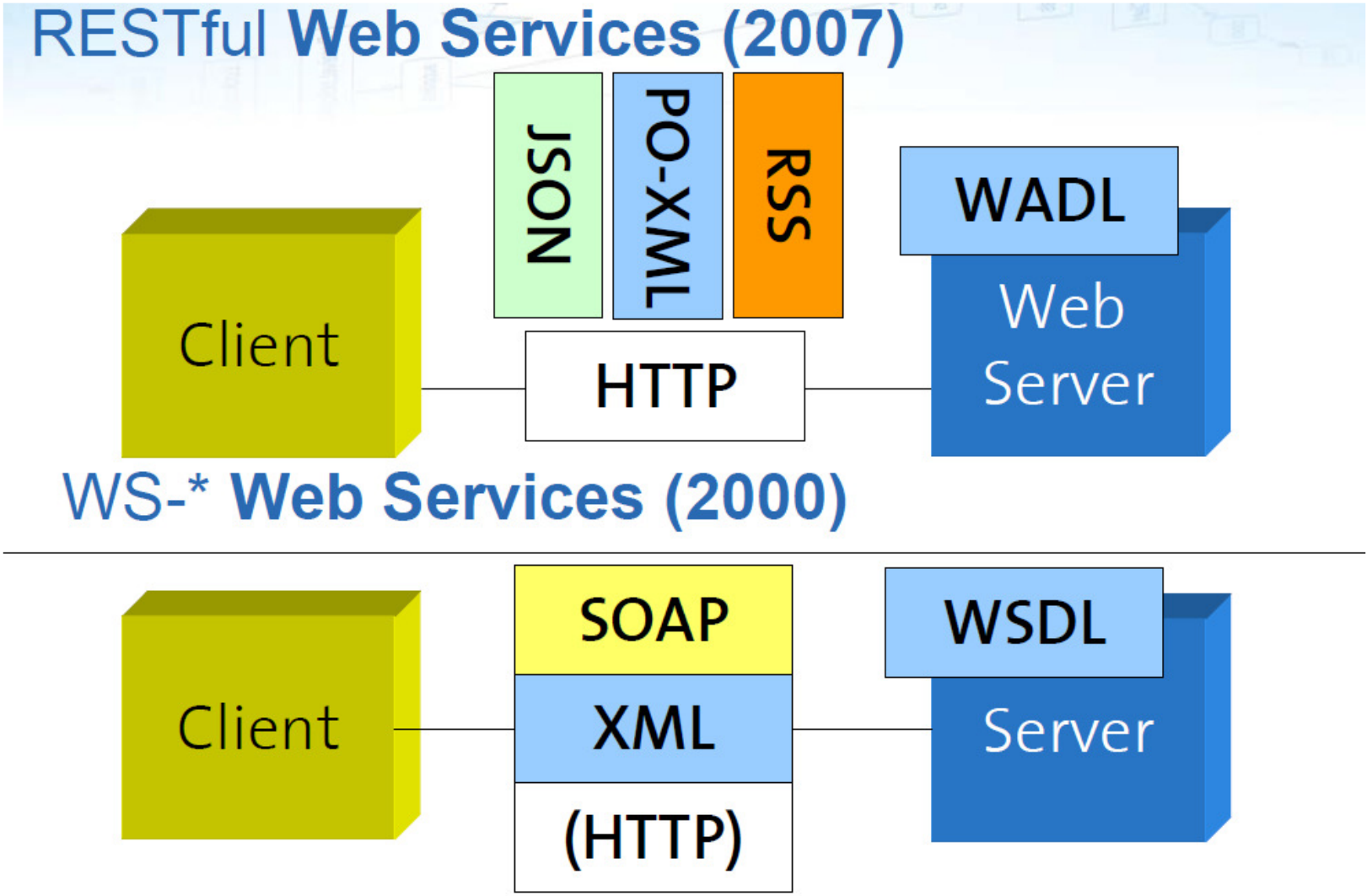
- The service interface defines service functionality visible to the external world and provides the means to access this functionality.
  - The service describes its own interface characteristics, i.e., the operations available, the parameters, data-typing and the access protocols, in a way that other software modules can determine what it does, how to invoke its functionality, & what result to expect in return.
- The service implementation realizes a specific service interface whose implementation details are hidden from its users.
  - Different service providers using any programming language of their choice may implement the same interface.

Service Interface & Implementation



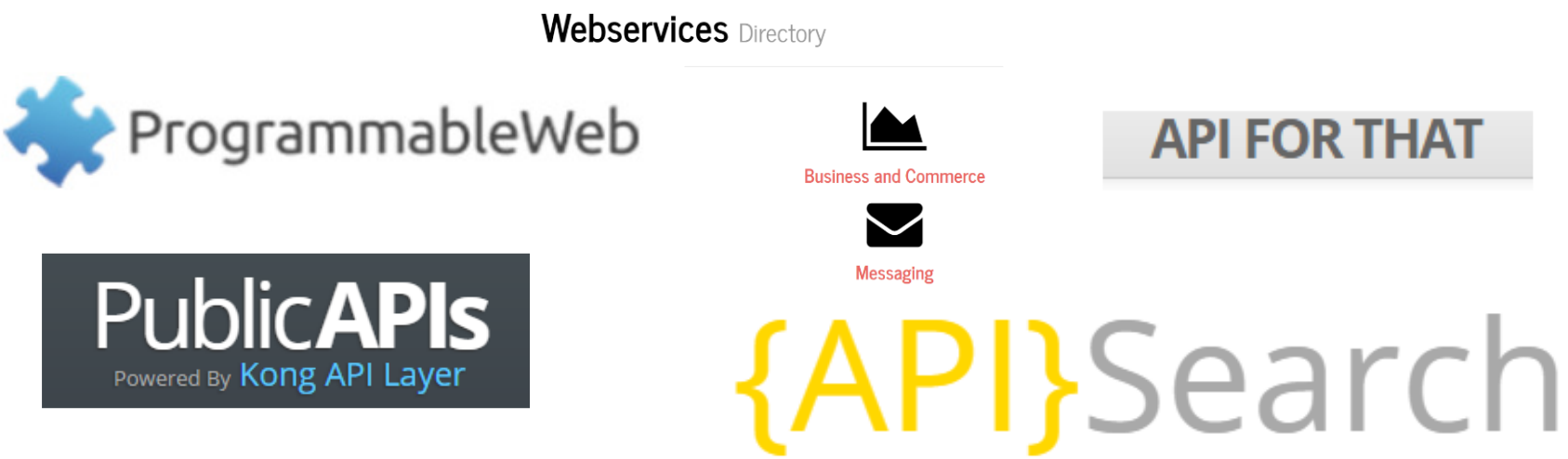
Two major classes of Web Service

- **XML SOAP Web services**
- **RESTful Web services**



Where to find web services?

- *http://www.programmableweb.com/*
- *http://www.apiforthat.com/*
- *https://www.publicapis.com/*
- *http://apis.io/*



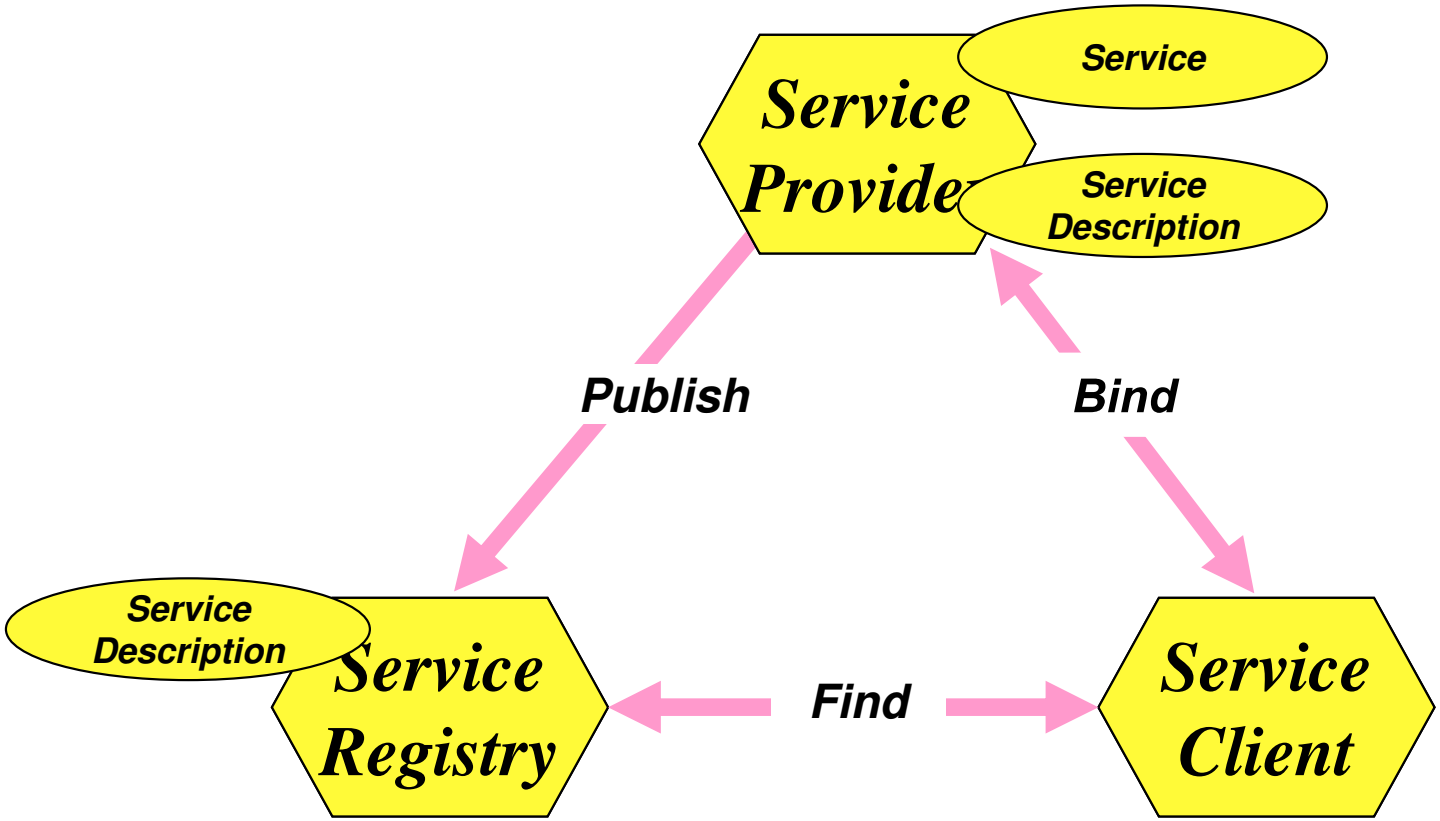
A simple weather web service

http://api.openweathermap.org/data/2.5/weather?q=Auckland&APPID=2f4d83e3d50672cf2009fc34611903f3

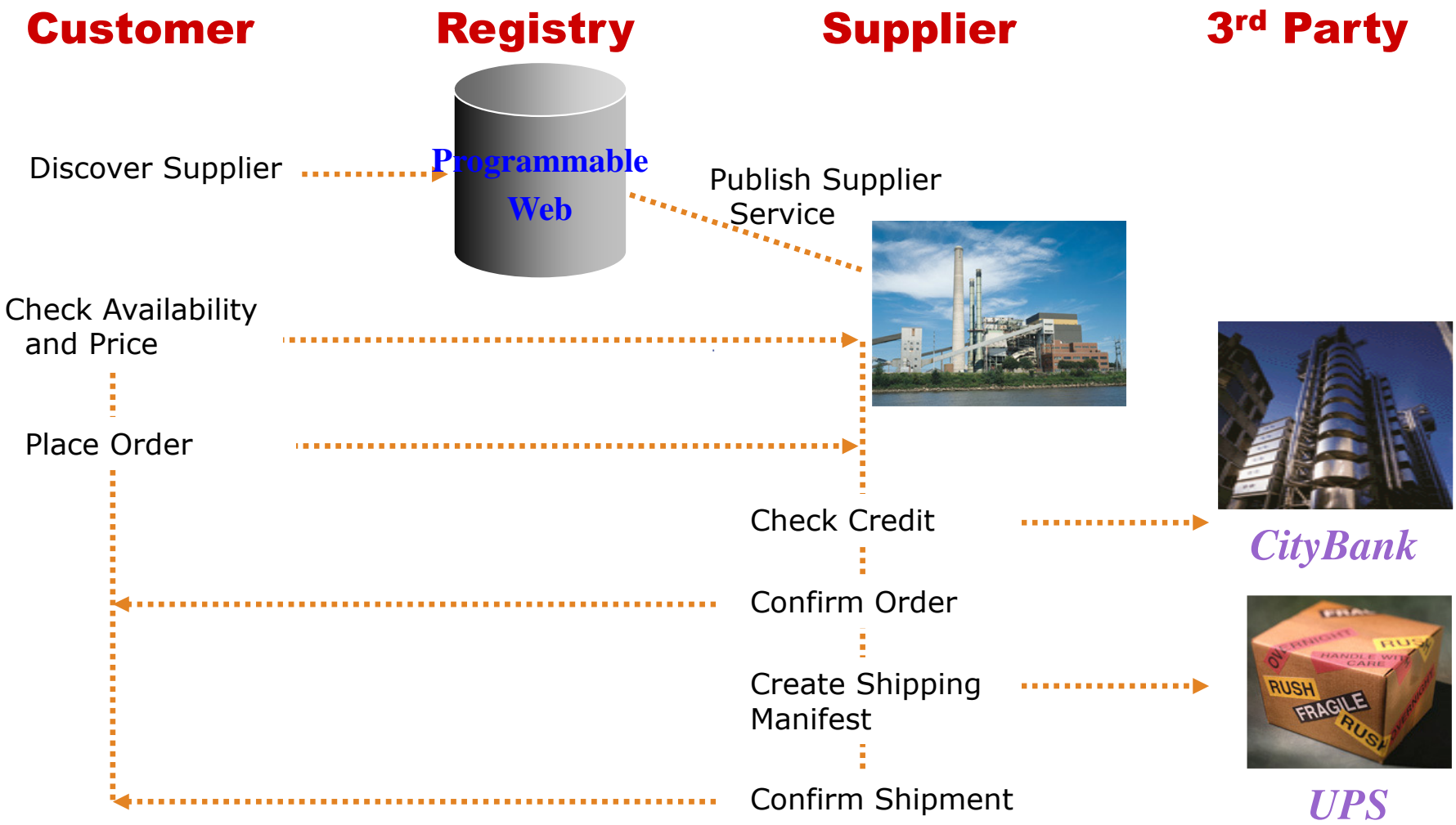

```
{    "coord":{"lon":174.77,"lat":-36.85},
     "weather":[{"id":804,"main":"Clouds","description":
     "overcast clouds","icon":"04d"}],
     "base":"stations",
     "main":{"temp":291.15,"pressure":1012,"humidity":63
     ,"temp_min":291.15,"temp_max":291.15},
     "visibility":10000,
     "wind":{"speed":4.6,"deg":270},
     "clouds":{"all":88},
     "dt":1526779800,
     "sys":{"type":1,"id":8276,"message":0.0038,
     "country":"NZ","sunrise":1526757347,"sunset":
     1526793526},
     "id":2193733,
     "name":"Auckland",
     "cod":200
}
```

Service Oriented Architecture

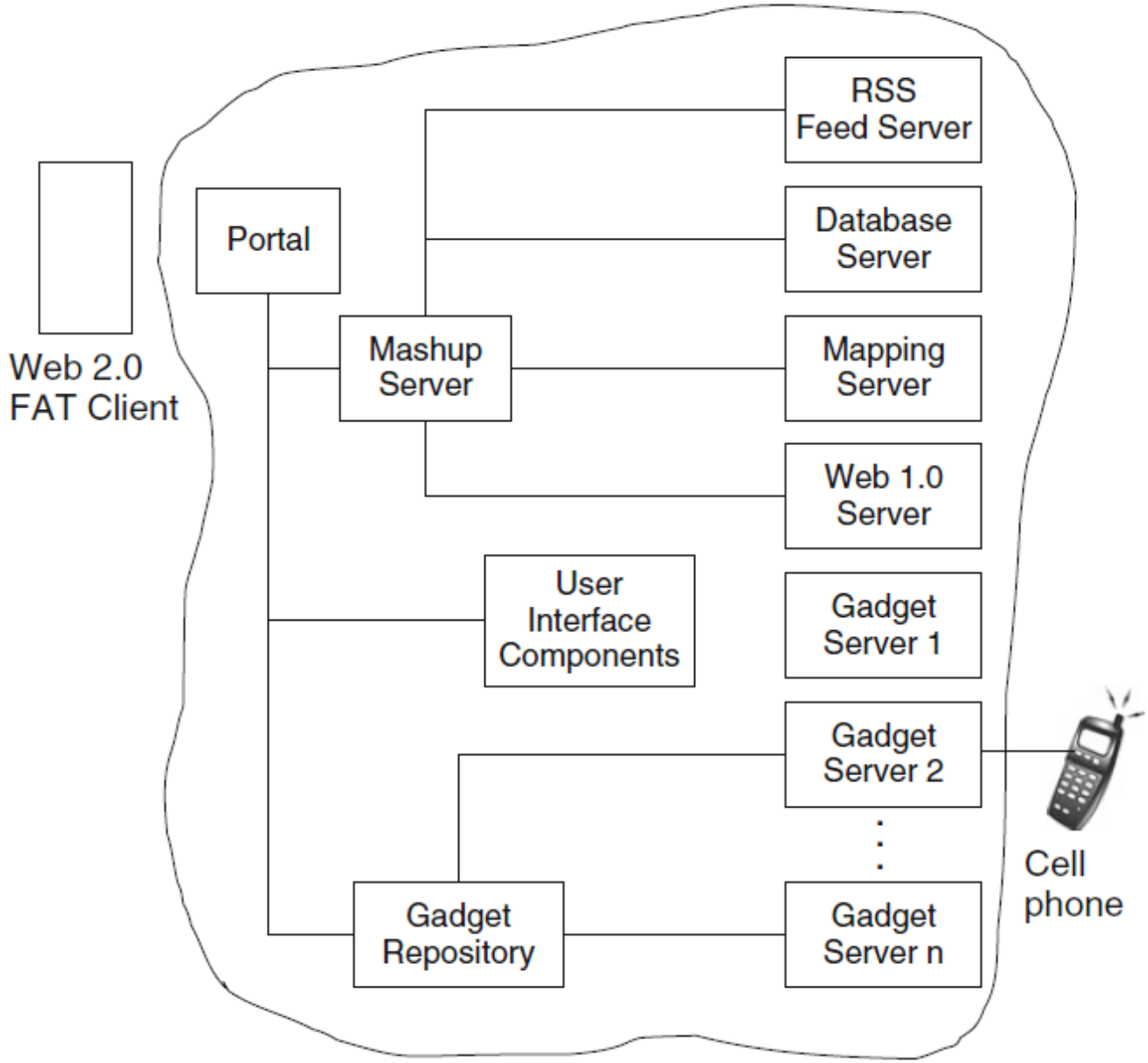
- SOA is a logical way of designing a software system to provide services to either end-user applications or to other services distributed in a network, via published and discoverable interfaces.



Service flow and workflow



Web 2.0 and mashup



Challenges for cloud computing

1. Availability of service: what happens when the service provider cannot deliver?

   in 2012 lightning caused a prolonged down time at Amazon.

2. Data confidentiality and auditability, a serious problem.

3. Diversity of services, data organization, user interfaces available at different service providers limit user mobility; once a customer is hooked to one provider it is hard to move to another.

4. Data transfer bottleneck; many applications are data-intensive.

More challenges

5. Performance unpredictability, one of the consequences of resource sharing.
   - How to use resource virtualization and performance isolation for QoS guarantees?
   - How to support elasticity, the ability to scale up and down quickly?

6. Resource management: It is a big challenge to manage different workloads running on large data centers. Are self-organization and self-management the solution?

7. Security and confidentiality: major concern for sensitive applications, e.g., healthcare applications.

   in 2009, Google was the target of a denial of service attack which took down Google News and Gmail for several days;

Addressing these challenges is on-going work!

Back to Basics -- Parallel Computing

- *"**Parallel computing** is a form of computation in which many calculations are carried out simultaneously, operating on the principles that large problems can often be divided into smaller ones, which are then solved concurrently (in parallel)."* Wikipedia

- Hardware and software systems allow us to:
  - Solve problems demanding resources not available on a single system.
  - **Reduce the time required to obtain a solution.**

Parallel Computing – Amdahl's Law

- The speedup *S* measures the effectiveness of parallelisation:

$$S(N) = T(1) / T(N)$$

  - *T(1)* → the execution time of the sequential computation.
  - *T(N)* → the execution time when N parallel computations are executed

- **Amdahl's Law:** if α is the fraction of running time a sequential program spends on non-parallelisable segments of the computation then:

$$S \cong 1 / \alpha$$

- This is a theoretical upper bound on the best speedup we can get from parallelising a certain program.

Back to Basics -- Distributed systems

- Collection of autonomous computers, connected through a network and distribution software (often) called middleware which enables computers to coordinate their activities and to share system resources for a common goal.
- Characteristics:
  1. The components are autonomous.
  2. Scheduling and other resource management and security policies are implemented by each system.
  3. There are multiple points of control and multiple points of failure.
  4. The resources may not be accessible at all times.
  5. Can be scaled by adding additional resources.
  6. Can be designed to maintain availability even at low levels of hardware/software/network reliability.

Summary

- Cloud Computing: concept and evolution history
- Attributes and benefits
- Delivery and deployment models
- Enabling technologies
  - Virtualization
  - Web service and service-oriented architecture
  - service flows and workflows
  - Web 2.0 and mashup
- Cloud computing challenges
- Parallel Computing and Distributed Systems (brief)

References

- Cambridge Uni Cloud Computing Course by Dr Eva Kalyvianaki
- Handbook of Cloud Computing, Springer