Cloud Computing Fundamentals
IN401 – M1S1 – 5 Credits

Course Grading

- Partial Exam
- Final Exam
- Second Session

Course Contents

- Chapter 1: Introduction to Cloud Computing
- Chapter 2: Infrastructure as a Service (IaaS)
- Chapter 3: Platform as a Service (PaaS)
- Chapter 4: Software as a Service (SaaS)
- Chapter 5: Cloud Native Technologies & Architectures
- Chapter 6: Cloud Security
- Chapter 7: Practical Case Studies

Chapter 1:

Introduction to Cloud Computing

Chap.1: Introduction to Cloud Computing

1.1 Existing Computing paradigms

1.2 Evolution (On-premise → Cloud)

1.3 Definitions

1.4 Characteristics

1.5 Cloud deployment models: Public, Private, Hybrid, Multi-Cloud

1.6 Cloud service models: IaaS, PaaS, SaaS, XaaS

1.7 Advantages and challenges, SLA basics

1.1 Existing Computing paradigms

Green Computing

Personal computing

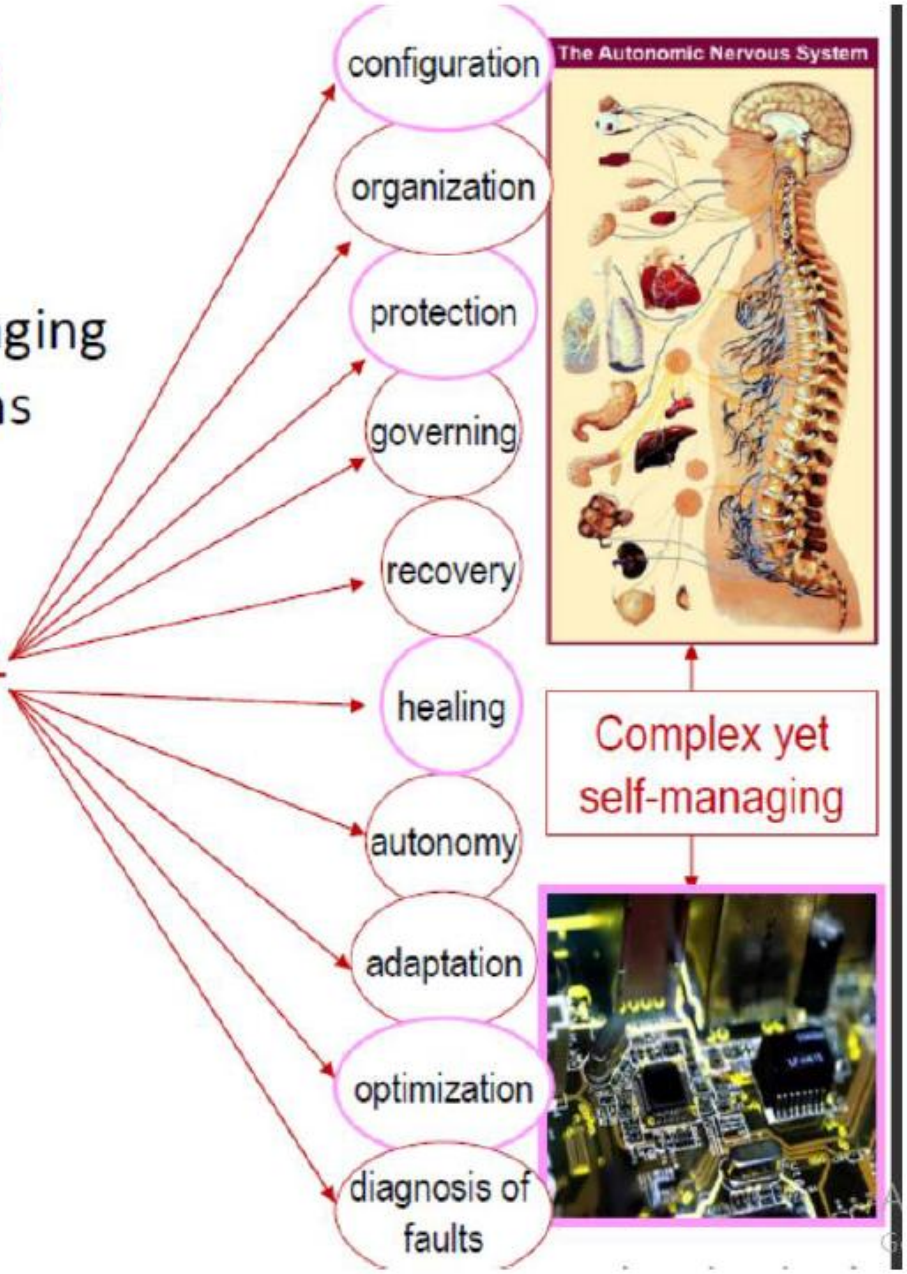
- Local Installation and maintenance
- Customizable to user needs
- Very low utilization
- High up-front cost

Reconfigurable computing

- Field programmable gate arrays (FPGA)
- Reprogrammable hardware
- Parallel code
- Power hungry

Autonomic Computing

- Motivation: rapidly growing complexity of integrating, managing and operating computer systems
- introduced by IBM in 2001
- Inspired by Human ANS
- Self-management includes: self-



Mobile Computing

- Mobile Communication, Hardware, Software.
- **Advantages**: Portability, Increased Productivity, Real-time Access.
- **Challenges and Limitations**: Security Risks, Connectivity Issues, Battery Life, Bandwidth.





Utility Computing

- Water, gas, and electricity are provided to every home and business as commodity services
  - You get connected to the utility companies' "public" infrastructure
  - You get these utility services on-demand
  - And you pay-as-you use

- Utility computing is a service provisioning model that offers computing resources on an on-demand basis, charging users based on their actual consumption rather than a flat rate.

Ubiquitous/Pervasive Computing

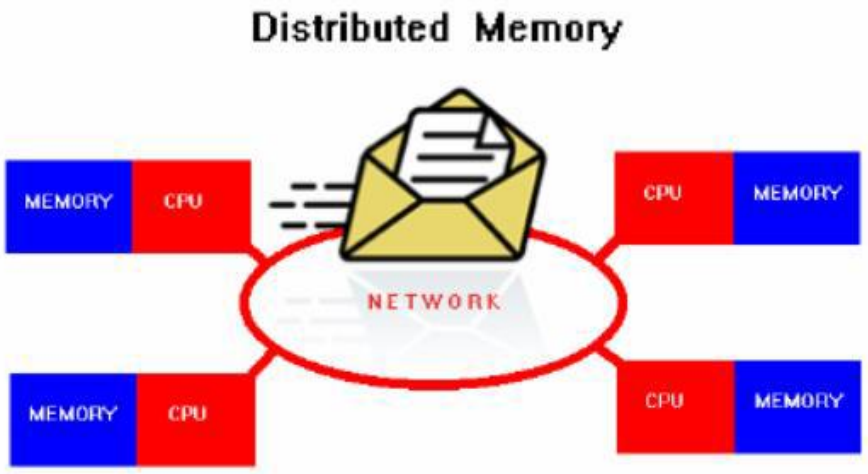
- Ubiquitous= "seeming to be in all places"
- Pervasive= "present or noticeable in every part of a thing or place"
- Information processing engaged in everyday's activities and objects.
- Term used since 1980s
- Different models but same vision:
  - Small, inexpensive, robust devices distributed throughout everyday's life

Green Computing

- Green computing refers to environmentally responsible computing practices that aim to:
- reduce energy consumption,
- carbon emissions,
- and electronic waste throughout the lifecycle of technology products.

Blue Group

- Distributed Computing
  - Using **distributed systems** to solve large problems.
  - **Distributed System:** multiple autonomous computers connected through a communication network
  - The system has a **distributed memory** where each processor has its private memory.
  - Information exchanged using communication models, ex: **MPI**

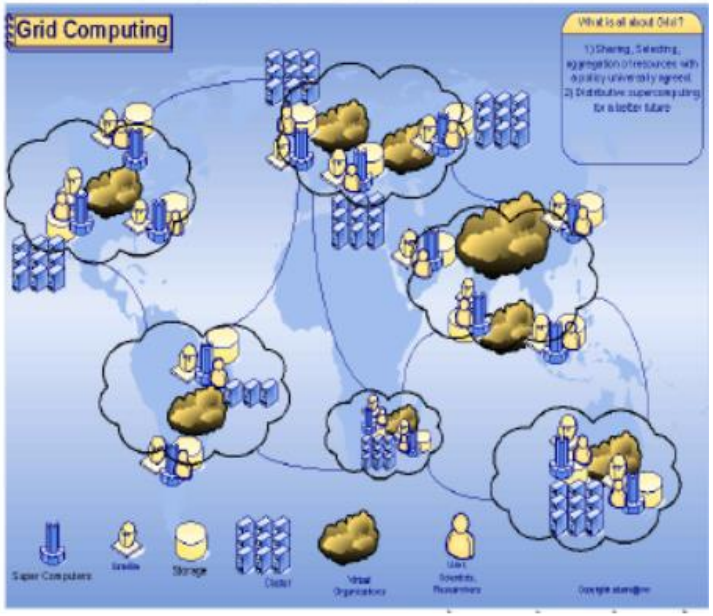


Blue Group

Distributed Computing

- **Cluster Computing:**
  - Characteristics:
    - tightly coupled computers
    - single system image
    - Centralized Job management & scheduling system
  - Better performance and availability and more cost-effectiveness over single computer with same capabilities
  - Since 1987

- **Grid Computing:**
  - According to Gartner, "a grid is a collection of resources owned by multiple organizations that is coordinated to allow them to solve a common problem."
  - Characteristics:
    - loosely coupled
    - no Single System Image
    - distributed Job Management & scheduling
  - Originated early 1990s



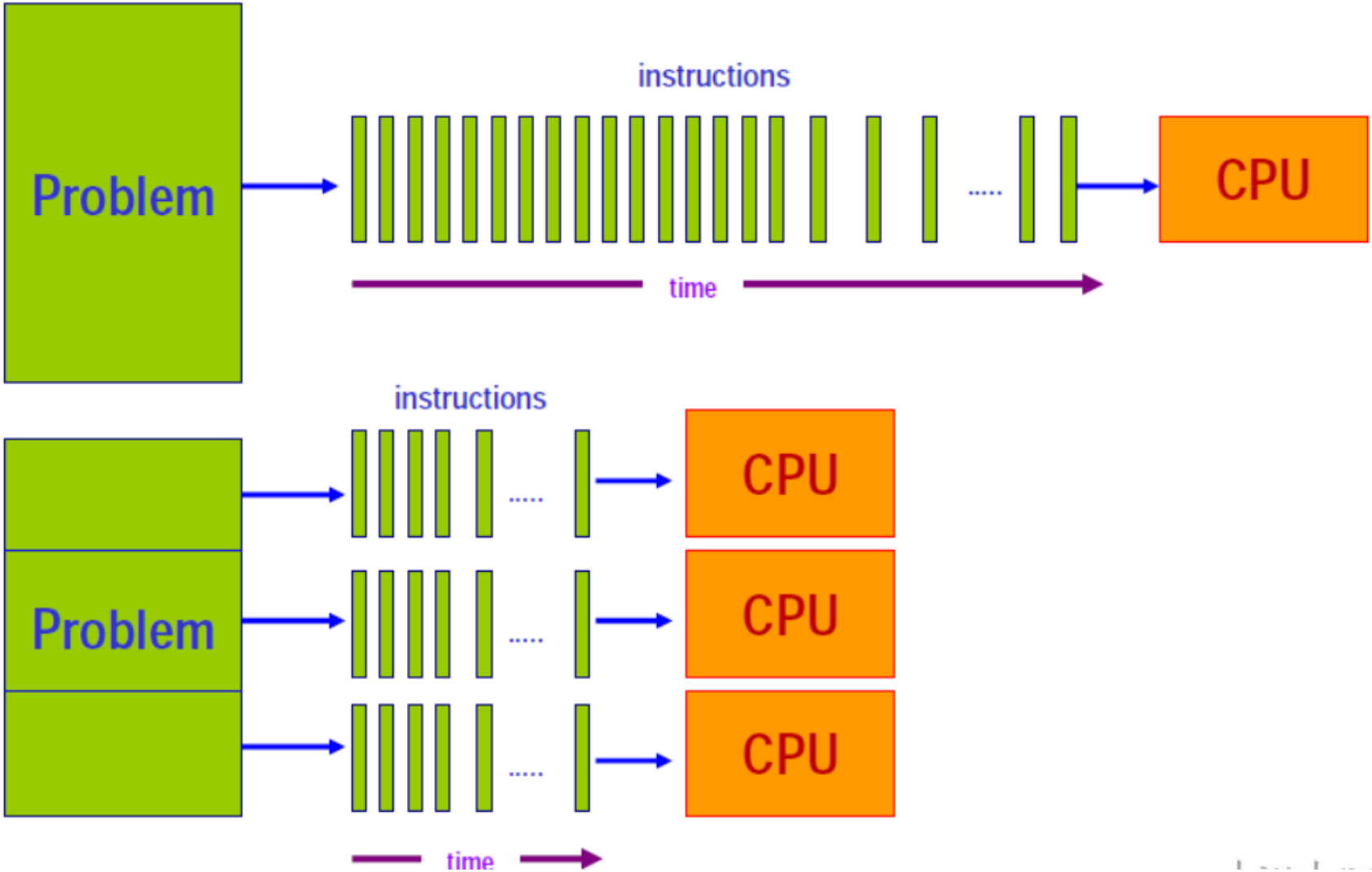
Parallel Computing

- *"**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.**

What is Parallel Computing

- Calculations of large problems are divided into smaller parts and carried out simultaneously/concurrently on different processors.



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:

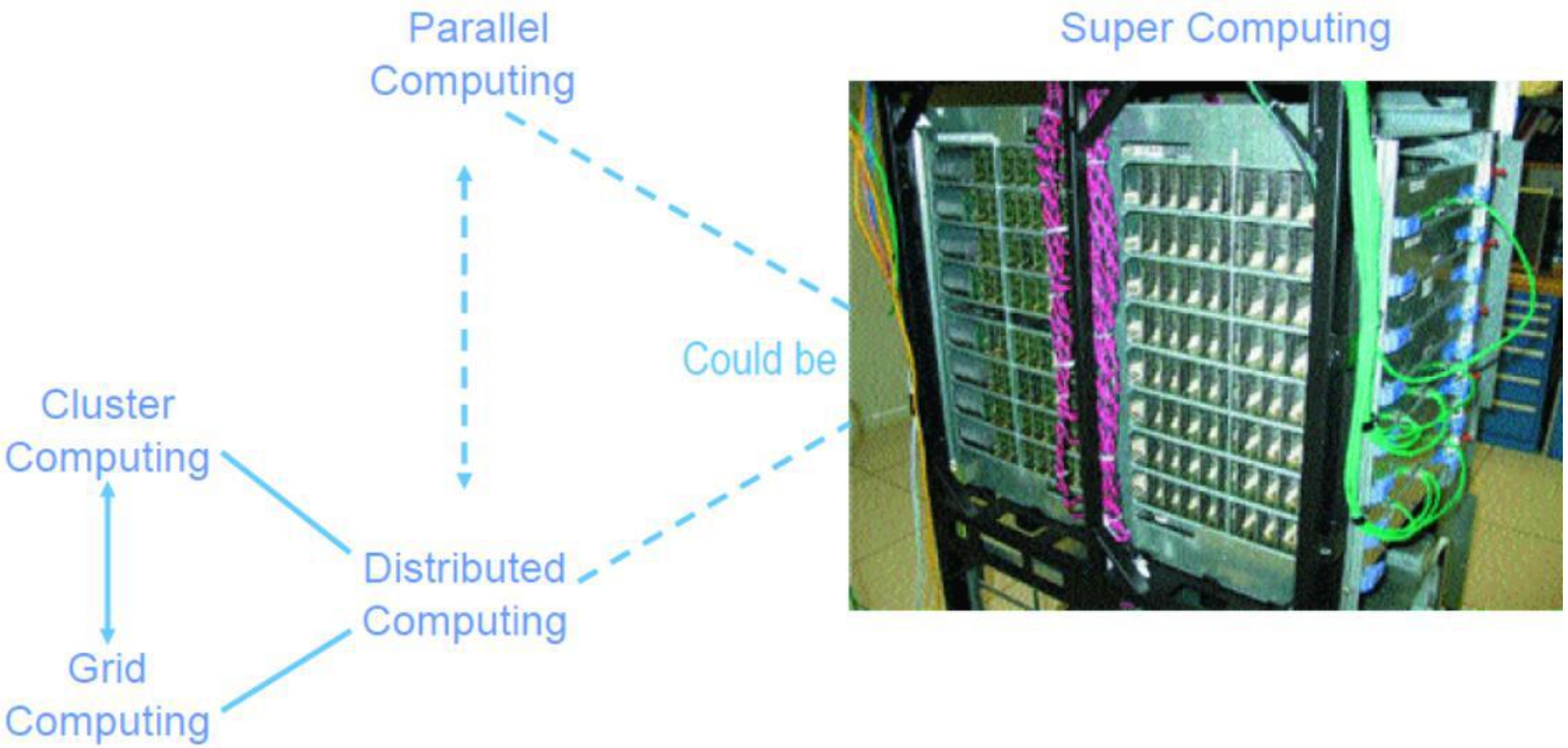
$$S \cong 1 / \alpha$$

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

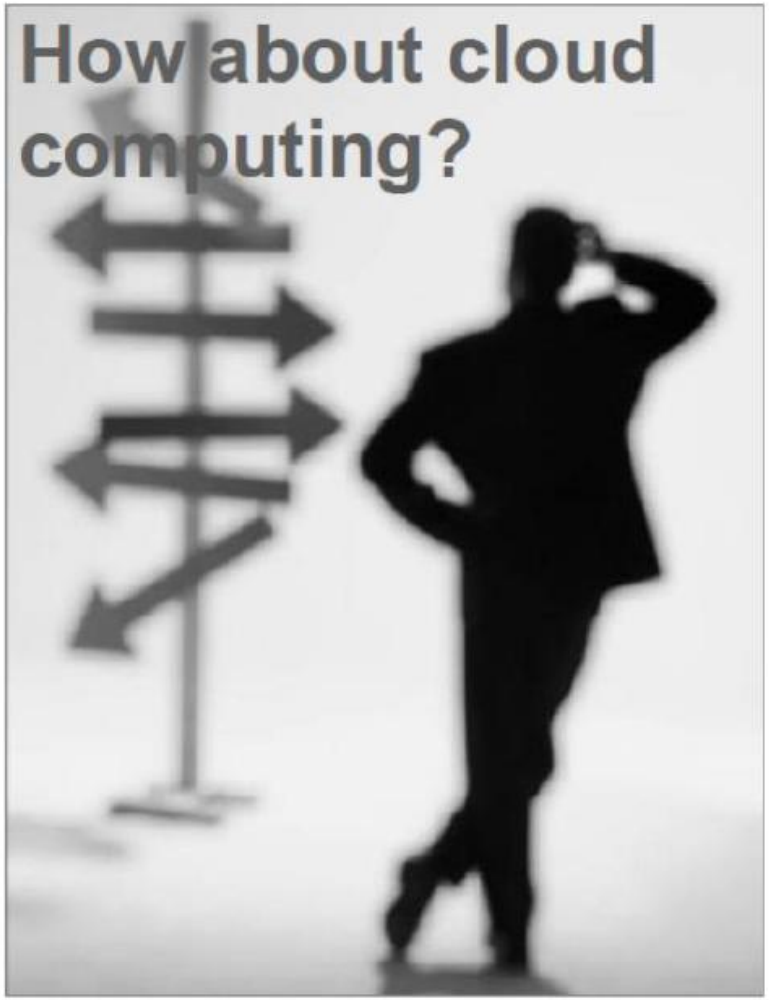
Blue Group

- Super Computing
  - Thousands of processors
  - Used for compute-intensive problems
    - Days instead of Years!!!
  - introduced in the 1960s

Blue Group


Parallel Computing
Super Computing
Could be
Cluster Computing
Distributed Computing
Grid Computing

Cloud Computing



Cloud Computing

1.2 Evolution (On-premises → Cloud)

Think of it this way …

- **Transportation**
  - Which one should you pick?
  - Should you buy/rent?

amazon

1.2 Evolution (On-premises → Cloud)

- Basic reasoning: information and data processing can be done more efficiently on large farms of computing and storage systems accessible via the Internet.

Enabling Technologies

- Virtualization
- Web 2.0
- Distributed Storage
- Distributed Computing
- Utility Computing
- Network Bandwidth & Latency
- Fault-Tolerant Systems

1.3 What is Cloud Computing?

- What do you think?

- ***"Cloud computing*** *is an information technology (IT) paradigm that enables ubiquitous access to shared pools of configurable system resources and higher-level services that can be rapidly provisioned with minimal management effort, often over the Internet. Cloud computing relies on sharing of resources to achieve coherence and economies of scale, similar to a public utility."* 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/*

What is Cloud Computing

- **Represents both the cloud & the provided services**
- **Why call it cloud?**
  - Wikipedia: "*the term derives from the fact that most technology diagrams depict the Internet or IP availability by using a drawing of a cloud.*"
- **Why call it cloud computing?**
  - Some say because the computing happens out there "*in the clouds*"

What is Cloud Computing

- In our daily routine we use this cloud service without our notice like web-based email service, watching movies through the internet, editing documents, storing pictures etc.. uses cloud computing on the back-end.

- Using such cloud technology, we can design and create new applications, store and recover data, host websites etc.

What is Cloud Computing

Data center : Composed of networked computers and storage that vendors use to organize, process, store and disseminate large amounts of data.

Cloud Computing

- **Who is Who…**

Cloud providers

Cloud Users & Service Providers

Service Users

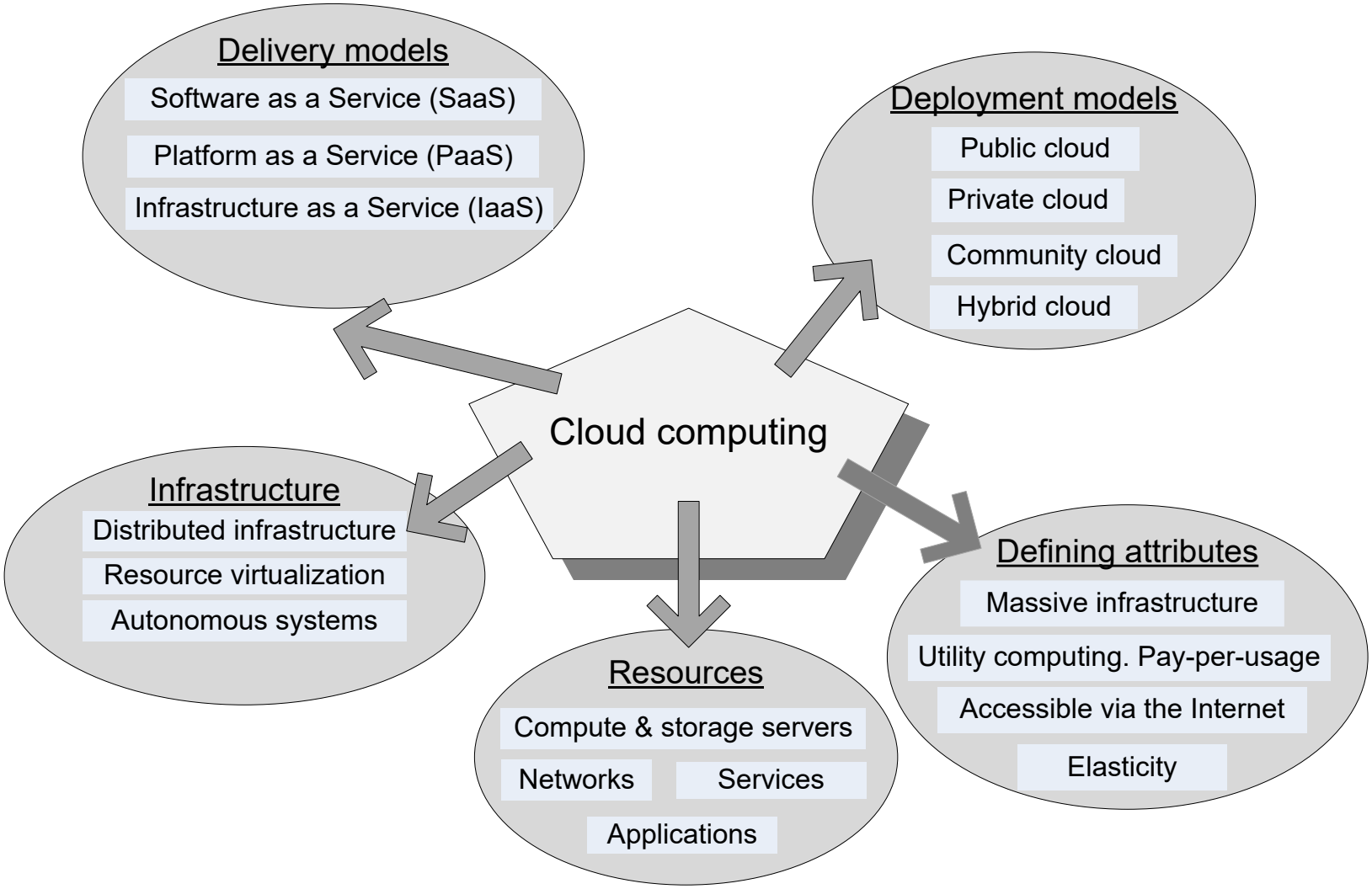
"With Amazon [AWS], on Day One of launch we could scale to the world."

-Brad Jefferson, Co-Founder & CEO, Animoto

"Animoto has partnered with Amazon to leverage multiple offerings in their Web Services (AWS) platform which, in conjunction with Animoto's own render farm, constitutes the Animoto web infrastructure."

Users use it to produce video pieces from their photos, video clips and music.

1.4 Cloud computing - Characteristics



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 technology to offer **scalable** and **elastic** services.
3. Resources are metered and users are charged accordingly.
4. It is more cost-effective due to **resource-multiplexing.** Lower costs for the cloud service provider are past to the cloud users.

Scalability vs Elasticity

- **Scalability** refers to a system's capacity to handle growing amounts of work or users without compromising performance. It involves the ability to accommodate increased workload by adding resources such as servers, storage, or network bandwidth.

- **Elasticity** refers to the ability of a system to dynamically adjust its resource allocation in response to changing demands (including scalability). Elasticity enables efficient resource utilization, cost optimization, and responsiveness to varying workloads without manual intervention, allowing organizations to effectively manage their IT infrastructure in a dynamic and agile manner.

Cloud computing - Characteristics

Data Storage:

5. 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.

6. The data storage strategy can increase reliability, as well as security, and can lower communication costs.

Management:

7. The maintenance and security are operated by service providers.

8. The service providers can operate more efficiently due to specialization and centralization.

Cloud computing - Characteristics

Why Cloud Computing?

- Large-Scale Data-Intensive Applications
- Flexibility
- Scalability
- Customized to your current needs:
  - Hardware
  - Software
- Effect:
  - Reduce Cost
  - Reduce Maintenance
  - High Utilization
  - High Availability
  - Reduced Carbon Footprint

Cloud computing - Characteristics

Why Cloud Computing?

- **Flexibility**
  - Software: Any software platform
  - Access: access resources from any machine connected to the Internet
  - Deploy infrastructure from anywhere at anytime
    - Software controls infrastructure

Cloud computing - Characteristics

Why Cloud Computing?

- **Scalability**

  Scale [Up (vertical)] vs [out (horizontal)]

  - Instant
  - Control via software
    - Add/cancel/rebuild resources instantly
  - Start small, then scale your resources up/down as you need
  - illusion of infinite resources available on demand

Cloud computing - Characteristics

Why Cloud Computing?

- **Customization**
  - Everything in your wish list
    - Software platforms
    - Storage
    - Network bandwidth
    - Speed

Cloud computing - Characteristics

Why Cloud Computing?

- Cost
  - Pay-as-you-go model
  - Small/medium size companies can tap the infrastructure of corporate giants.
    - Time to service/market
    - No upfront cost

Cloud computing - Characteristics

Why Cloud Computing?

- **Maintenance**
  - Reduce the size of a client's IT department
  - Is the responsibility of the cloud vendor
  - This Includes:
    - Software updates
    - Security patches
    - Monitoring system's health
    - System backup
    - ...etc

Cloud computing - Characteristics

Why Cloud Computing?

- Utilization
  - Consolidation of a large number of resources
    - CPU cycles
    - Storage
    - Network Bandwidth

Cloud computing - Characteristics

Why Cloud Computing?

- **Availability**
  - Having access to software, platform, infrastructure from anywhere at any time
  - All you need is a device connected to the internet

- **Reliability**

The system's fault tolerance is managed by the cloud providers and users no longer need to worry about it.

Cloud computing - Characteristics

Why Cloud Computing?

- CO2 Footprint
  - Consolidation of servers
  - Higher utilization
  - Reduced power usage

1.5 Cloud 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 standardized technology that enables data and application portability.

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

What is a public cloud?

- A public cloud is a computing model where a cloud service provider makes computing resources (e.g., software applications, development platforms, VMs, bare metal servers, etc.) available to users over the public internet.

- CSPs sell these resources according to subscription-based or pay-per-usage pricing models.

- According to a Gartner report (link resides outside ibm.com), worldwide end-user spending on public cloud spending is forecasted to total $679 billion and is projected to exceed $1 trillion in 2027.

What is a public cloud?

- Public cloud environments are multi-tenant, where users share a pool of virtual resources automatically provisioned for and allocated to individual tenants through a self-service interface.
- In this scenario, multiple tenants' workloads might run CPU instances simultaneously on a shared physical server, yet their data remains separate.
- By having numerous customers share resources, cloud vendors can offer their services to many customers at a lower cost than if these clients maintained their own infrastructure.

Benefits of public cloud

- **Rapid scalability:** Add compute capacity instantly or automatically in response to unexpected surges in traffic.
- **Affordability:** Avoid the investing required to deploy and maintain on-premises IT infrastructure. Pay only for the resources or services you use to aid cost savings.
- **Ease of set up:** Quickly spin up new servers or other resources without purchasing hardware or configuring physical infrastructure.

What is private cloud?

- A private cloud environment is a cloud computing model dedicated to a single organization.
- Unlike a public cloud, a private cloud is a single-tenant environment—resources are accessible to one client only.
- Private clouds can be hosted on-premises in an organization's data center, on rented infrastructure in an offsite data center or hosted off-premises on a cloud service provider's infrastructure.
- A private cloud offers greater control over security and resource customization than a public cloud.
- Many organizations opt for a private cloud setting to protect sensitive data—a business need that is becoming increasingly important.

Private Cloud Benefits

- **Increased resource control:** Gain more control over IT resources with configurations maintained by internal IT team members.
- **Customization:** Customize hardware and software tailored to meet unique business needs, such as policies for compliance.
- **Robust security:** Gain greater visibility and access control by storing sensitive data and applications behind private firewalls that limit the attack surface.

What is Hybrid Cloud ?

- A hybrid cloud environment combines a public cloud, private cloud and on-premises infrastructure to create a single IT infrastructure so companies can get the most out of all computing environments based on their needs. Essentially, it's the best of both worlds.
- For instance, public cloud resources can scale up quickly, automatically and cost-effectively in response to spikes in traffic without affecting private cloud workloads.
- This configuration method, known as "cloud bursting," helps organizations manage sudden surges in computing demand, which can occur in online retail situations like Black Friday sales.

What is hybrid cloud (continue)?

- Another core function of hybrid cloud is to support microservices (or microservices architecture), the cloud-native architectural approach in which a single application comprises many loosely coupled and independently deployable smaller components or services.
- Cloud-native applications are deployed in containers. Orchestration tools such as Kubernetes or Docker Swarm then schedule the automated deployment, management and scaling of those applications across all cloud computing environments.
- Microservices have become crucial for DevOps methodologies. Microservices help teams develop applications once and across all types of clouds.

What is hybrid cloud (continue)?

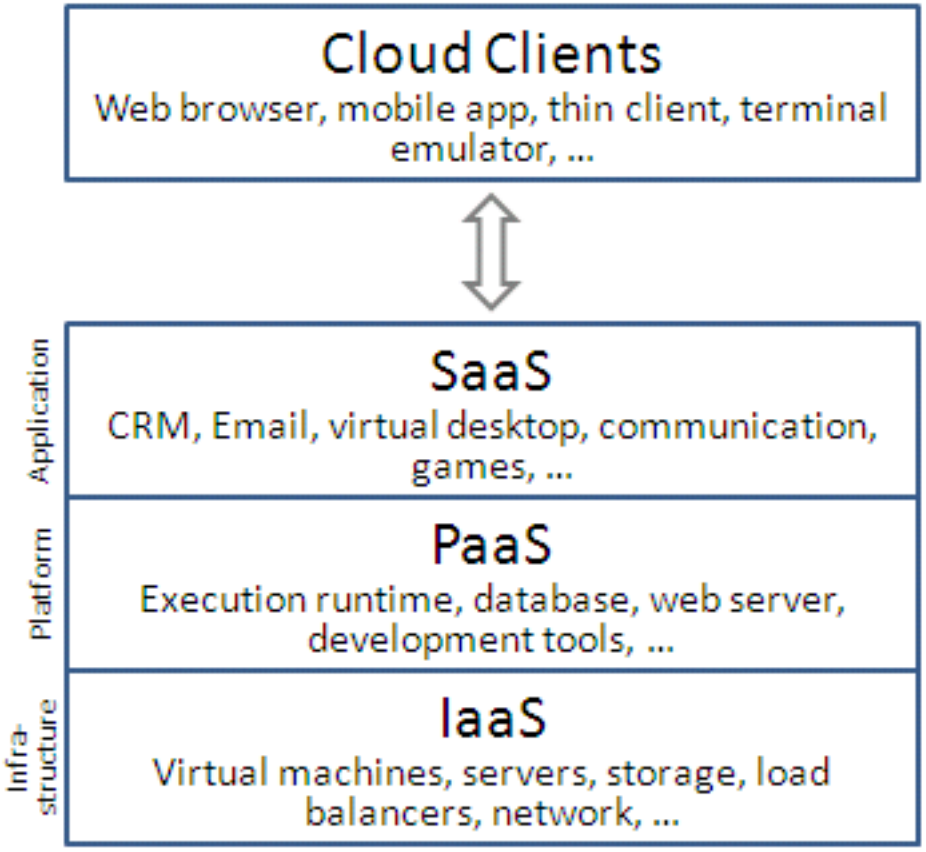
- Uber, for example, depends on a microservices architecture to build and release its ride-hailing and food-delivery services quickly.
- Hybrid cloud architecture also provides resource flexibility to help manage data gathered from multiple edge and Internet of Things (IoT) settings with flexible compute, network and cloud storage resources.
- On a manufacturing factory floor, for example, a hybrid cloud helps to provide an end-to-end solution for gathering insights, analyzing data and delivering predictive maintenance solutions with low latency and no downtime.

Benefits of hybrid cloud:

- **Flexibility:** Allocate workloads to fit best with business needs. Utilize public cloud resources for short-term projects like development and testing. Protect confidential or sensitive data on private cloud infrastructure.
- **Improved application development:** Expand adoption of agile and DevOps methodologies, enabling faster application development and time to market.
- **Digital transformation:** Leverage vast amounts of compute to process big data and harness the latest technologies like generative AI and machine learning (ML).

1.6 Cloud Delivery (Services) Models

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



source Wikipedia

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: server hosting, storage, computing hardware, operating systems, 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.
- Not particularly useful when:
  - The application must be portable.
  - Proprietary programming languages are used.
  - The hardware and software must be customised to improve the performance of the application.
- Examples: Google App Engine, Windows Azure

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.
- Not suitable for applications where data is not allowed to be hosted externally.

- Examples: Gmail, Salesforce

The Three delivery models of Cloud Computing

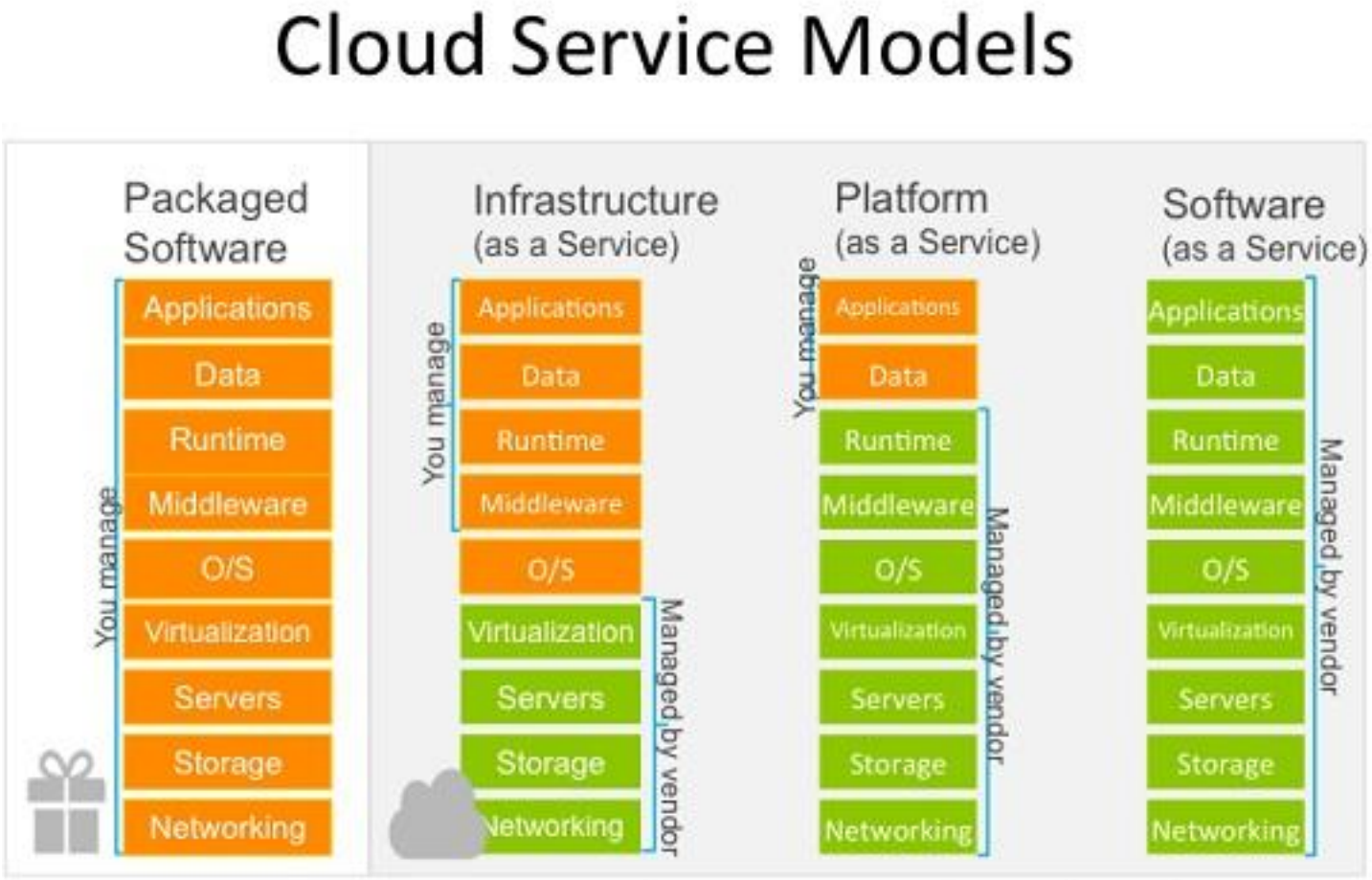
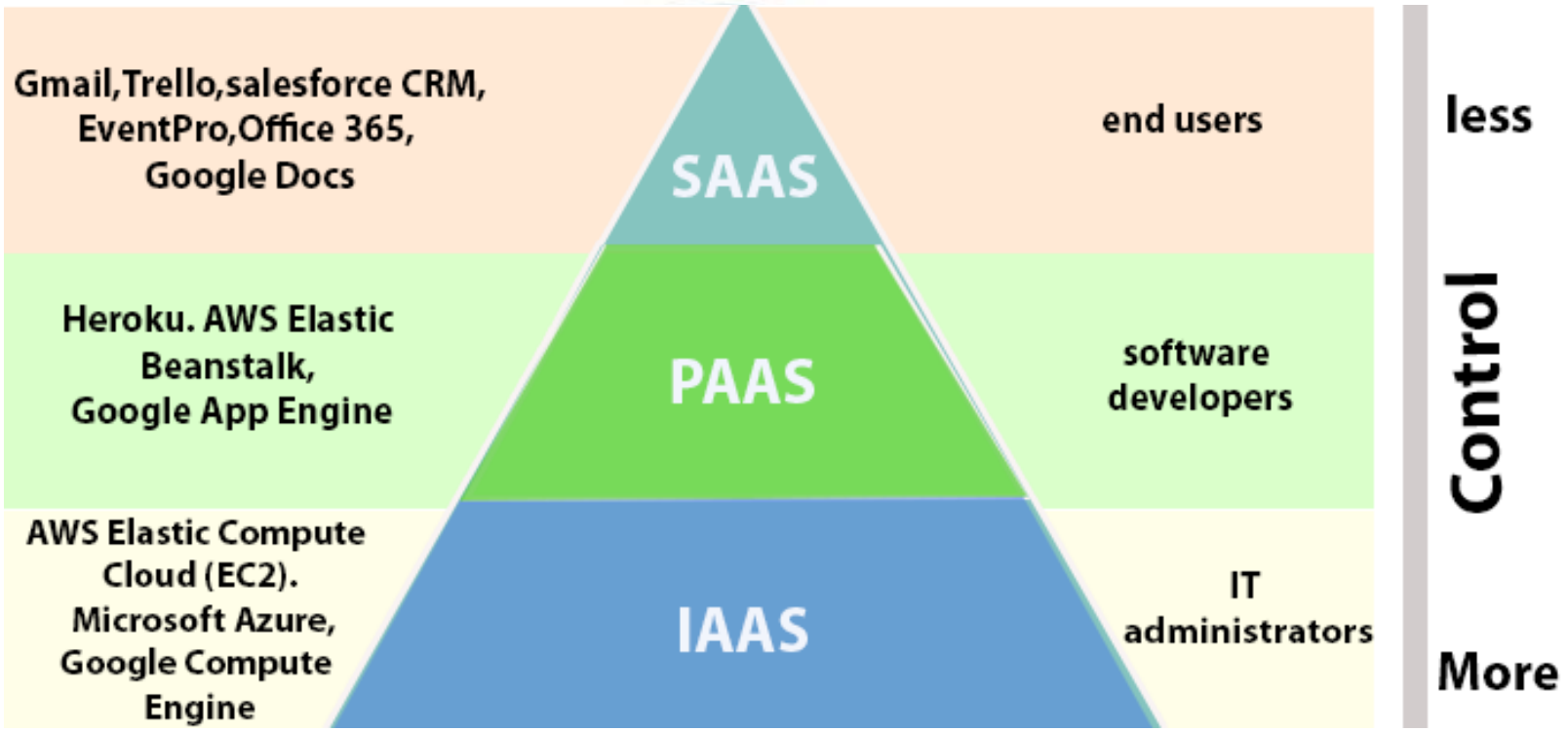


Figure 1. **Source: Microsoft Azure**


Gmail,Trello,salesforce CRM, EventPro,Office 365, Google Docs
SAAS
end users
Heroku. AWS Elastic Beanstalk, Google App Engine
PAAS
software developers
AWS Elastic Compute Cloud (EC2). Microsoft Azure, Google Compute Engine
IAAS
IT administrators
less
Control
More

More (XaaS): Everything as a Service EaaS

- **Desktop: DaaS**
  - Use your desktop virtually from anywhere
- **Communication: CaaS**
- **Virtualization: VaaS**
- **Hardware: HaaS**
- **...etc**

1.7 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.

Why cloud computing is (could) be successful when other paradigms have failed?

- It is in a better position to exploit recent advances in software, networking, storage, and processor technologies promoted by the same companies who provide Cloud services.
- Economical reasons: It is used for enterprise computing; its adoption by industrial organizations, financial institutions, government, and so on has a huge impact on the economy.
- Infrastructures Management reasons:
  - A single Cloud consists of a mostly homogeneous (now more heterogeneous) set of hardware and software resources.
  - The resources are in a single administrative domain (AD). Security, resource management, fault-tolerance, and quality of service are less challenging than in a heterogeneous environment with resources in multiple ADs.

Challenges for cloud computing

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

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. (Vendor lock-in)(Migration)

4. Data transfer bottleneck; many applications are data-intensive. (Network-dependent)

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.

Addressing these challenges is on-going work!

Privacy issues

- Cloud service providers have already collected petabytes of sensitive personal information stored in data centers around the world. The acceptance of Cloud Computing therefore will be determined by privacy issues addressed by these companies and the countries where the data centers are located.

- Privacy is affected by cultural differences; some cultures favour privacy, others emphasise community. This leads to an ambivalent attitude towards privacy in the Internet which is a global system.

Cloud Vulnerabilities

- Clouds are affected by malicious attacks and failures of the infrastructure, e.g., power failures.

- Such events can affect the Internet domain name servers and prevent access to a Cloud or can directly affect the Clouds:
  - in 2004 an attack at Akamai caused a domain name outage and a major blackout that affected Google, Yahoo, and other sites.
  - in 2009, Google was the target of a denial of service attack which took down Google News and Gmail for several days;
  - in 2012 lightning caused a prolonged down time at Amazon.

SLA basics

- A Service Level Agreement (SLA) is a contract between a cloud service provider and a client that outlines the expected level of service, performance metrics, and responsibilities of both parties.
- SLAs are crucial in cloud computing as they ensure that the service provider meets the agreed-upon standards and provides a clear framework for accountability and performance measurement.

Key Components of SLAs

- Service Availability: This specifies the uptime or availability of the service. For example, an SLA might guarantee 99.95% uptime for a cloud service.
- Performance Metrics: These include response time, latency, and throughput. For instance, an SLA might specify a maximum response time of 2 seconds.
- Reliability: This covers the reliability of service components and their ability to function without failure.
- Accountability: This defines the responsibilities of each party, including who is accountable for specific tasks and actions.
- Warranties and Penalties: These outline the guarantees provided by the service provider and the penalties for failing to meet the SLA terms.

SLA Lifecycle

- Discover Service Provider: Identify a provider that meets the organization's needs
- Define SLA: Establish the service level requirements, metrics, and targets
- Establish Agreement: Formalize the agreement, including penalties for non-compliance
- Monitor SLA Violation: Regularly monitor performance to ensure compliance
- Terminate SLA: End the agreement if the provider fails to meet the objectives
- Enforce Penalties: Impose penalties for SLA violations

Future of computing

1. The intelligent cloud is ubiquitous computing, enabled by the public cloud and artificial intelligence (AI) technology, for every type of intelligent application and system you can envision.
2. The intelligent edge is a continually expanding set of connected systems and devices that gather and analyze data—close to your users, the data or both. Users get real-time insights and experiences, delivered by highly responsive and contextually aware apps.
3. Combine the virtually limitless computing power of the cloud with intelligent and perceptive devices at the edge of your network to create a framework for building immersive and impactful business solutions.

Future of Cloud