UNIVERSITY OF CAMBRIDGE

Cloud Computing

Introduction

Eva Kalyvianaki
ek264@cam.ac.uk

Anil Madhavapeddy
anil@recoil.org

My Background and Contact Details

- Dr **Eva**ngelia Kalyvianaki
- Senior Lecturer since Oct. 2018
- PhD from the CL, Cambridge University (srg, netos group)
- Postdoc from Imperial College London
- Lecturer/Senior Lecturer in City University London
- I like building systems, working in Cloud computing, distributed systems management and autonomic computing

- Office in FN15
- ek264@cam.ac.uk
- http://www.cst.cam.ac.uk/~ek264/

Course Logistics

- Resources:
  - Web page: http://www.cl.cam.ac.uk/teaching/1819/CloudComp/
  - Book:
    - "*Cloud Computing, Theory and Practice*" Dan C. Marinescu, Morgan Kaufmann
  - Research papers (will be given per lecture)

- One coursework project performed in groups of two students each. The project will be assessed via a project report and code testing.
- Deadline for groups by the **20th of October** and send me an email. If not, I will randomly assign you into groups.

Course Contents

1. Introduction to Cloud Computing
2. Virtualization I
3. Virtualization II
4. Data Center Networking
5. MapReduce Batch Processing
6. MapReduce in Heterogeneous Environments
7. Large-Scale Resource Management
8. Resource Management, VM CPU Schedulers
9. Cloud Distributed Storage
10. Real-Time Data Stream Processing

Topic	Lecturer	Date	Room
1. Introduction	E. Kalyvianaki	4/10	LT2
2. Virtualization I	A. Madhavapeddy	9/10	LT1
3. Virtualization II	A. Madhavapeddy	11/10	LT1
4. MapReduce I	E. Kalyvianaki	16/10	LT1
5. Data Center Networking	Dr. Paolo Costa (MSRC)	18/10	LT1
6. MapReduce II	E. Kalyvianaki	23/10	LT1
7. Large-Scale Resource Man.	E. Kalyvianaki	25/10	LT1
8. VM CPU Scheduling	E. Kalyvianaki	30/10	LT1
9. Tutorial I	Dr. Javad Zarrin	1/11	LT1
10. Tutorial II	Dr. Javad Zarrin	6/11	LT2
11. Tutorial III	Dr. Javad Zarrin	8/11	LT1
12. Cloud Storage	E. Kalyvianaki	13/11	LT1
13. Real-time Data Stream Processing	E. Kalyvianaki	15/11	LT1

Lecture Contents

- What is Cloud Computing?
- Early models of Cloud Computing.
- Delivery models and services.
- Ethical issues in Cloud Computing.
- Cloud vulnerabilities.
- Parallel Computing.
- Distributed Systems.

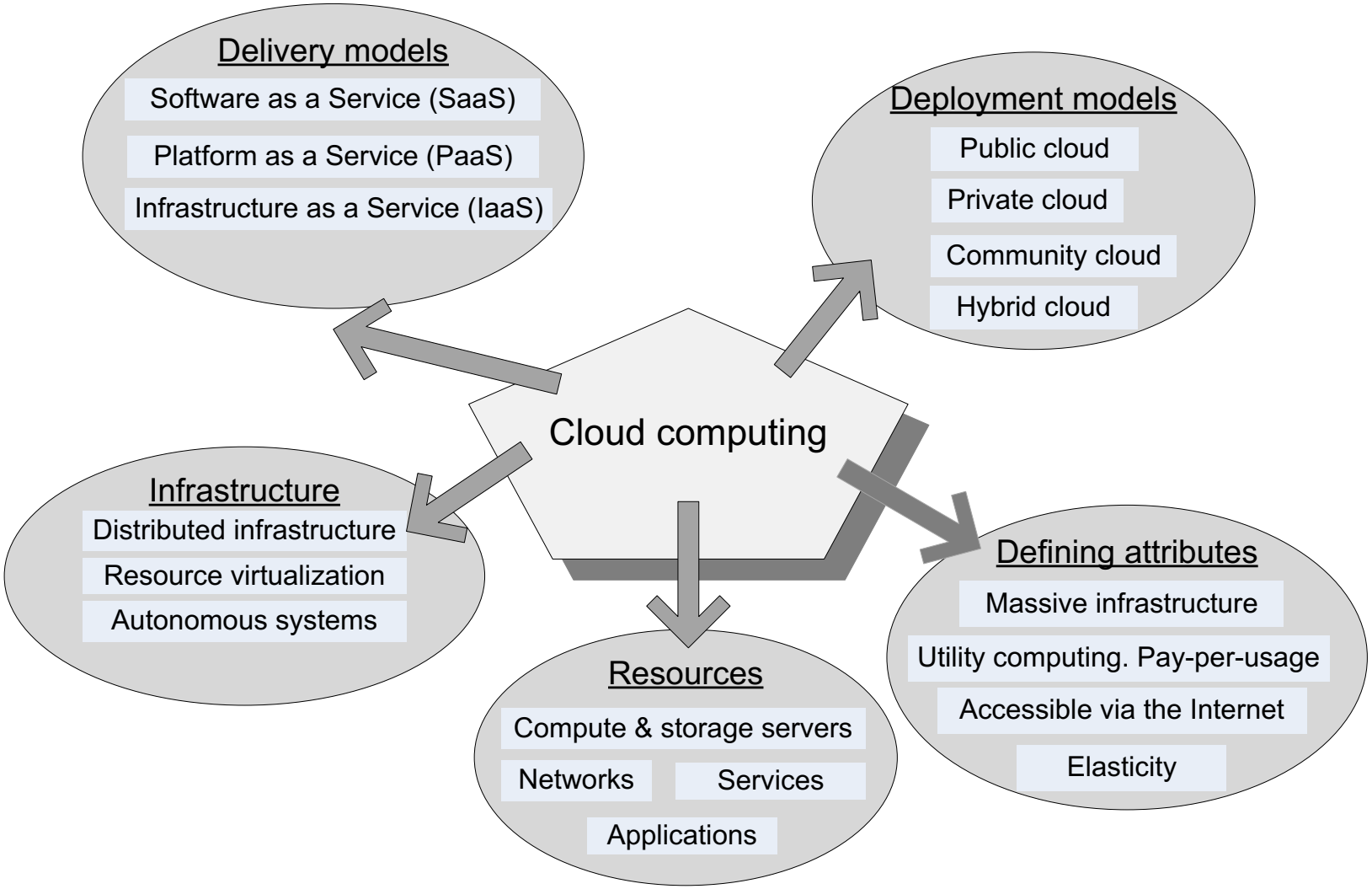
What is Cloud Computing?

- 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/*

Cloud Computing Models, Resources, Attributes



Early Models of Cloud Computing

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

1. **Grid computing** – initiated by the National Labs in the early 1990s; targeted primarily at scientific computing.
   - *"Grid computing is the collection of computer resources from multiple locations to reach a common goal. The grid can be thought of as a distributed system with non-interactive workloads that involve a large number of files."* from Wikipedia
2. **Utility computing** – initiated in 2005-2006 by IT companies and targeted at enterprise computing.
   - *"Utility computing is a service provisioning model in which a service provider makes computing resources and infrastructure management available to the customer as needed, and charges them for specific usage rather than a flat rate."* from Wikipedia

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 past 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

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.

1. **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.

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

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.

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?

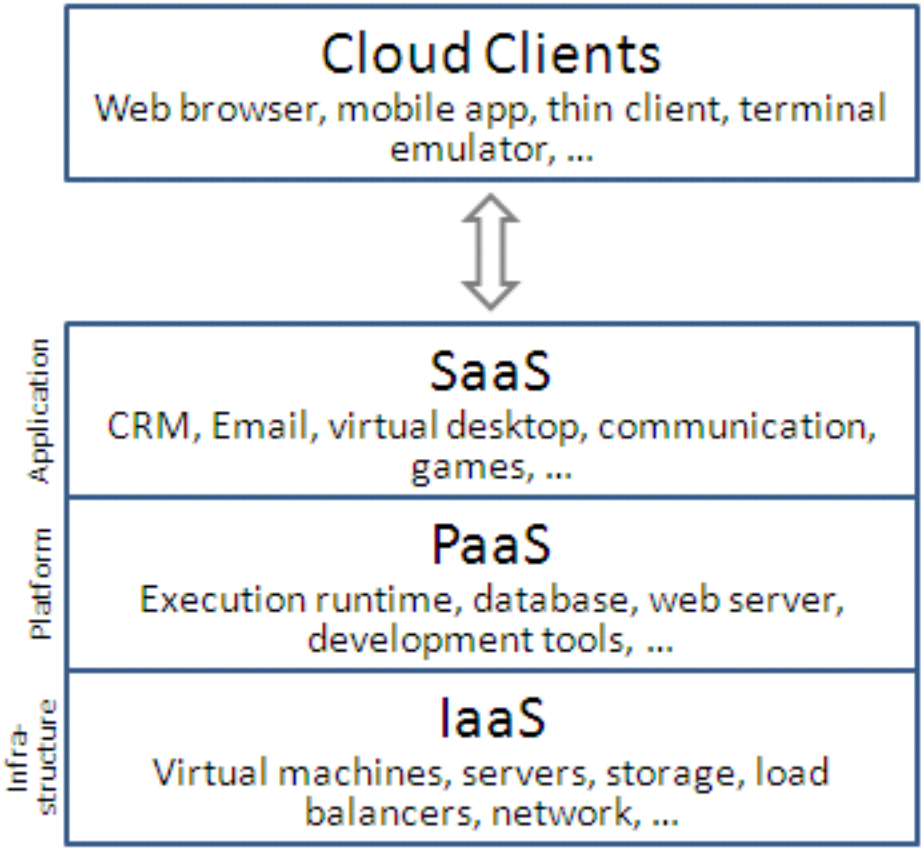
6. Resource management: It is a big challenge to manage different workloads running on large data centers. Are self-organization and self-management the solution?

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

Addressing these challenges is on-going work!

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)



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 real-time applications or for those where data is not allowed to be hosted externally.

- Examples: Gmail, Salesforce

The Three delivery models of Cloud Computing

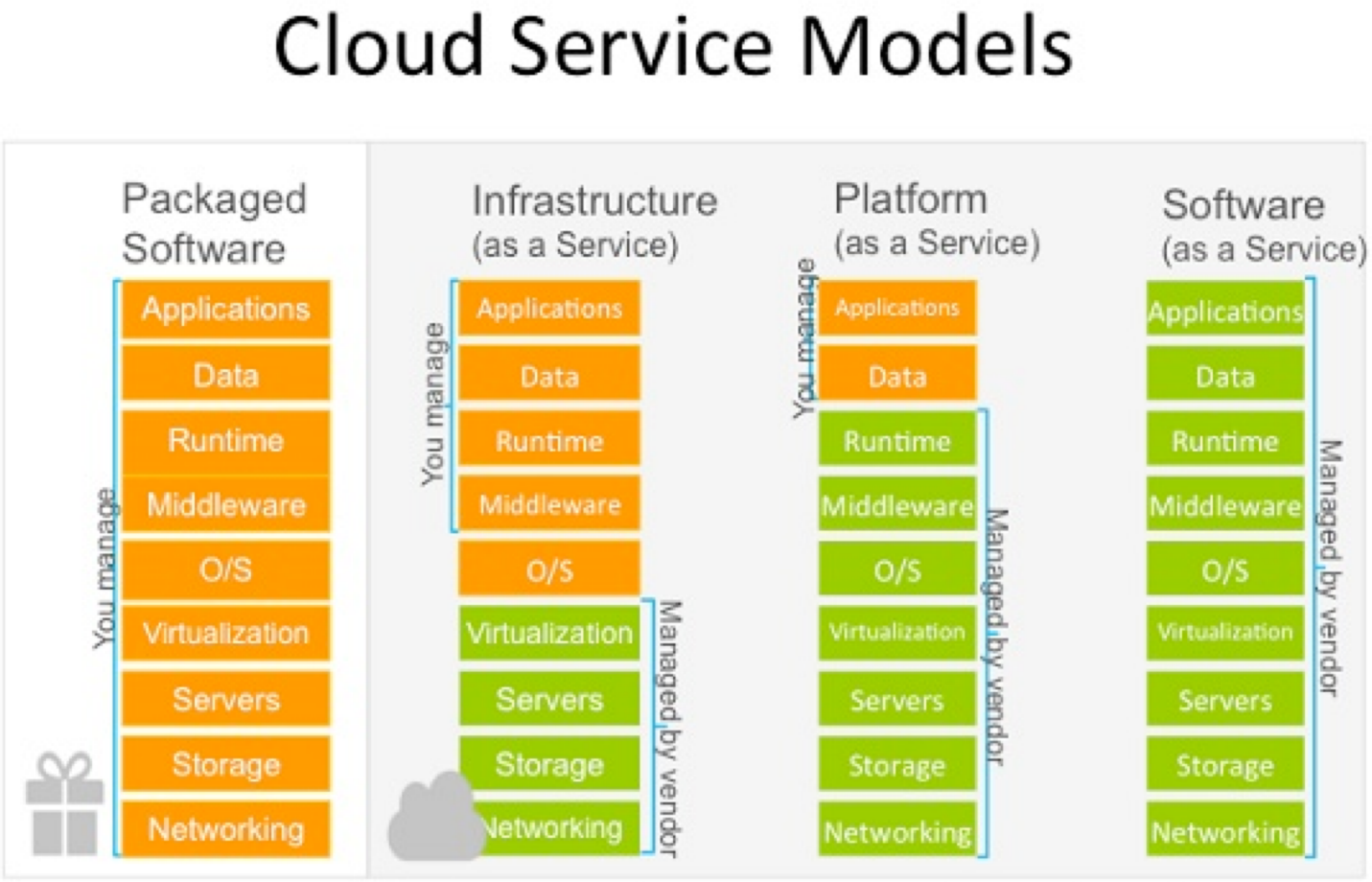


Figure 1. **Source: Microsoft Azure**

Cloud activities

- Service management and provisioning including:
  - Virtualization.
  - Service provisioning.
  - Call center.
  - Operations management.
  - Systems management.
  - QoS management.
  - Billing and accounting, asset management.
  - SLA management.
  - Technical support and backups.

Cloud activities (cont'd)

- Security management including:
  - ID and authentication.
  - Certification and accreditation.
  - Intrusion prevention.
  - Intrusion detection.
  - Virus protection.
  - Cryptography.
  - Physical security, incident response.
  - Access control, audit and trails, and firewalls.

Cloud activities (cont'd)

- Customer services such as:
  - Customer assistance and on-line help.
  - Subscriptions.
  - Business intelligence.
  - Reporting.
  - Customer preferences.
  - Personalization.
- Integration services including:
  - Data management.
  - Development.

Ethical issues

- Paradigm shift with implications on computing ethics:
  - The control is relinquished to third party services.
  - Data is stored on multiple sites administered by several organizations.
  - Multiple services interoperate across the network.
- Implications:
  - Unauthorised access.
  - Data corruption.
  - Infrastructure failure, and service unavailability.

De-perimeterisation

- Systems can span the boundaries of multiple organisations and cross the security borders.
- The complex structure of Cloud services can make it difficult to determine who is responsible in case something undesirable happens.
- Identity fraud and theft are made possible by the unauthorised access to personal data in circulation and by new forms of dissemination through social networks and they could also pose a danger to Cloud Computing.

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.

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 users perceive the system as a single, integrated computing facility.
  2. The components are autonomous.
  3. Scheduling and other resource management and security policies are implemented by each system.
  4. There are multiple points of control and multiple points of failure.
  5. The resources may not be accessible at all times.
  6. Can be scaled by adding additional resources.
  7. Can be designed to maintain availability even at low levels of hardware/software/network reliability.

Summary

- What is Cloud Computing?
- Early models of Cloud Computing.
- Delivery models and services.
- Ethical issues in Cloud Computing.
- Cloud vulnerabilities.
- Parallel Computing and Distributed Systems (brief)