

# Short introduction to cloud computing

Tarja Systä

## What is cloud computing?

- In principle, cloud computing implementations offer seemingly infinite pooled computing resource over the network
  - users can start, stop, and scale its power at will
- Comes close to the idea of *utility computing*:
  - Ideally computing is provided the same way as e.g. water or electricity; available in every home
  - outsourcing all the hardware and getting charged by the use.

## Definitions

- 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.
  - **Cloud service engineering** leverages cloud computing in the context of the Internet in its combined role as a platform for technical, economic, organizational and social networks

## What's new and what's old?

- Cloud computing combines features of
  - cluster computing and
  - grid computing
  - ...with the help of virtualization.

# Grid computing

- Systems are run on multiple interconnected nodes
- The resources are pooled to form a computational grid
  - Note that pooling of resources differs from that of cloud computing due to the on-demand nature of cloud computing
- Mostly utilized in complicated computational or data intensive tasks
- Cloud vs. grid computing
  - Grids and clouds are both intended to be utilised by multiple clients
  - Grids have typically multiple owners, while clouds and clusters are privately owned
  - Pooling of resources in grid computing differs from that of cloud computing due to the on-demand nature of cloud computing;
    - grid computing supports coordinated resource sharing, resource usage is planned in advance and resources are allocated to the users according to their needs.

## Clustered computers

- A cluster is a set of multiple interconnected computers
  - For example, a cluster can run a distributed service where the workload is divided among the computers in the cluster.
- The components of a cluster are commonly, but not always, connected to each other through fast local area networks
- The two most common reasons for clustering are performance and high availability

## Clustered computers

- Cluster vs. grid and cloud computing
  - grid nodes are usually globally distributed, while cluster nodes exist in one physical location and are connected with high speed (local area) network.
  - clusters are privately owned and are built for some specific purpose with homogeneous hardware
  - a cluster is an order of magnitude smaller unit than a grid or a cloud

# Virtualization

- Allows separation of an operating system, and thus the management of resources in a computer, from hardware.
  - This is possible by allowing a host operating system, the system that runs on the hardware, to create a virtual environment that can run any machine code that the hardware supports.
  - In practice, a host operating system runs other operating systems as guests. The guest system is called a virtual machine (VM).
- Allows an easy and secure way to split the resources of a computer among users or software.

## Characteristics

- 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.

## Characteristics

- According to NIST the key characteristics are: (cont'd)
  - 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
    - 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.

## Other claimed characteristics

- Reliability
  - improved if multiple redundant sites are used, which makes well designed cloud computing suitable for business continuity and disaster recovery <sup>[1]</sup> Nonetheless, many major cloud computing services have suffered outages, and IT and business managers can at times do little when they are affected. <sup>[2,3]</sup>
- Security
  - could improve due to centralization of data, increased security-focused resources, etc., but concerns can persist about loss of control over certain sensitive data, and the lack of security for stored kernels. <sup>[4]</sup>
- Maintenance
  - cloud computing applications are easier to maintain, since they do not have to be installed on each user's computer. They are easier to support and to improve since the changes reach the clients instantly.
    - That does not guarantee improved maintenance, though

[1] King, Rachael. "Cloud Computing: Small Companies Take Flight". Businessweek.com. 2008

[2] "Google Apps Admins Jittery About Gmail, Hopeful About Future". Pcworld.com. 2008.

[3] "New Resource, Born of a Cloud Feud". Datacenterknowledge.com. 2009.

[4] "Encrypted Storage and Key Management for the cloud". Cryptoclarity.com. 2009.

## 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**
  - On the other hand, resources, electricity etc. are wasted if the servers are not used extensively enough
  - ...Then again, overloads cause breaks etc.

## Levels of cloud computing

- Infrastructure as a Service (IaaS)
  - the lowest level of abstraction type of cloud computing models
  - offers (leasing of) a computer infrastructure – often a virtual hardware infrastructure - that immediately accessible and ready to use.
- Platform as a Service (PaaS)
  - offers a software stack (that deals e.g. with parallelism and distribution) and a complete abstraction to hardware
  - offers a computing platform and/or solution stack as a service, often consuming IaaS and sustaining SaaS *cloud applications*.
- Software as a Service (SaaS)
  - highest level of cloud computing
  - While IaaS and PaaS are aimed for a software developer, SaaS is often aimed directly to the end user
  - The services range from machine services (that can use SOA technologies to deliver services) to desktop applications.

## Deployment models

- 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 (from the Internet) access is possible
- Hybrid cloud
  - consists of multiple internal and/or external providers
  - a combination of public and private clouds; it utilizes public cloud for non-critical information handling and private cloud for critical information handling