Introduction to cloud computing

Boris Teabe

boris.teabe@inp-toulouse

Goals

- General introduction
- Definition of main concepts
- Some examples of cloud providers

The goal of this class is to have a general introduction to cloud computing. We are going to define what is cloud computing and why you people should know about it.

Cloud computing: the idea

- Amazon (around 2002)
  - Rent to external users a part of its computing facilities during periods of low use
  - Creation of Amazon Web Services (AWS) (initially for data storage, then for computing)
- My car
  - It could be rented to people who don't have a car
  - It exists !
    - Drivy
    - BlaBlaCar
- Generally
  - An under-used resource (computer, car) can be rented to other users on demand
  - Augment its utilisation rate, therefore its rentability

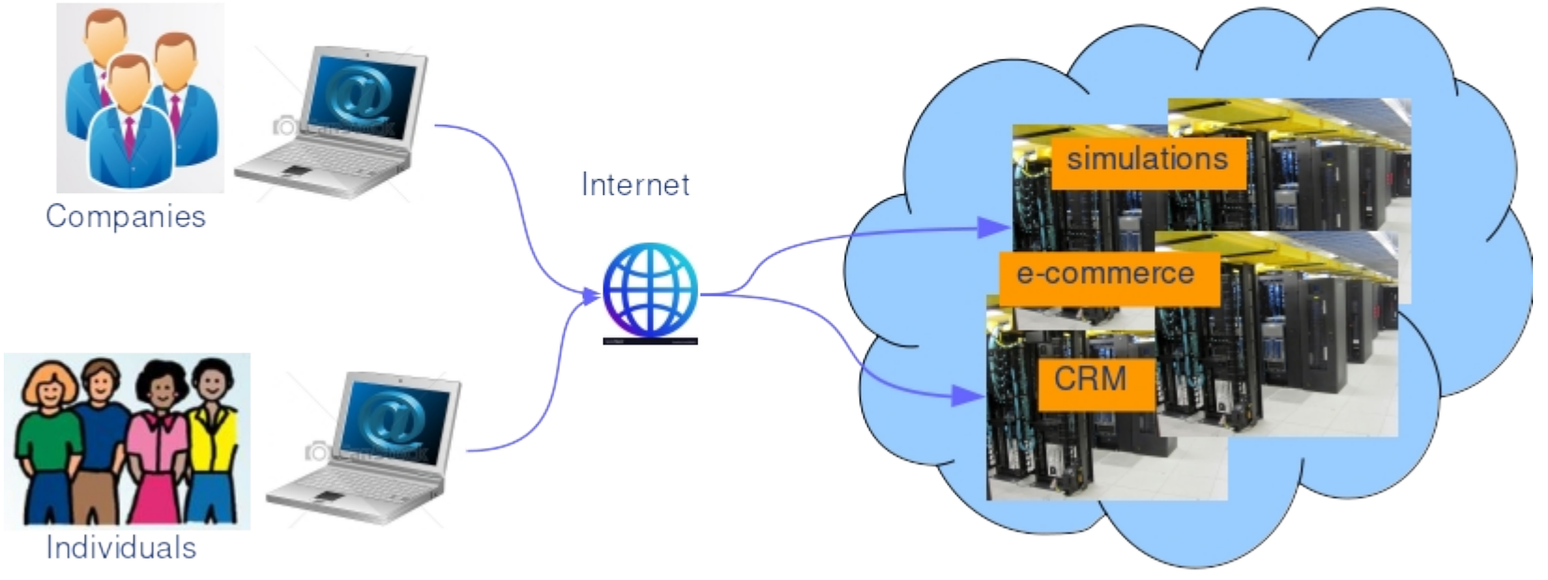
This slide is pretty clear, Amazon was selling goods, but know they are selling infrastructures.

Principles of the cloud

- Mutualize efforts and resources
  - Share an equipment
  - Utilization as needed
- Avantages
  - Reduce costs and wastings
  - Push the limits (alone, this would have been impossible)
- Applied to computer science
  - Shared computers (data centers)
  - Externalize (local => remote)

Share ressources for a better usage, that is easy to understand.

General scheme : externalization



Access based on REST API or a Web portal from the cloud platform

IT Outsourcing. We all pretty much know what this is about. This is the procedure by which a company entrusts all or part of its information system to be managed to an external company, the cloud provider. The company can access his software through a network that can be internet.

Even you can do that. Externalization is for everyone.

First services

- Amazon EC2 (Elastic Compute Cloud)
  - Sell virtual machines (VMs)
  - The client connects to a VM and uses it as he wishes
  - Comparable to a dedicated server
  - The most widespread service

Historicaly, Pinkham and lead developer Christopher Brown developed the Amazon EC2 service. This was the first service launched by amazon cloud.

Today, EC2 stills exist and is one of the main services used in AWS.

Definition of Cloud

- Just one of many
- Set of resources/applications/services which execute in a distributed environment (hosting center), accessible though web standard protocols, which globally provide a service with the following characteristics:

- Pay as you go (according to quantity and duration)
- Illusion of an infinity of resource (scalability)
- Abstraction of the hardware infrastructure
- Mutualization between many users

NIST Definition of Cloud Computing

The definition of cloud computing provided by NIST has gained significant traction within the IT industry. According to this definition (see [NIST]):

“Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models.”

Difference with the grid

- The Cloud is open to everybody
  - No need to be an expert to use it
- The Grid a devoted to researchers
  - Access is limited to members of a consortium

- The Cloud is more business oriented
  - A given quality of service must be enforced
  - A billing model must be defined
- The Grid is non-profit purpose

- The Cloud abstracts the hardware infrastructure for its users
- The grid users must know about the infrastructure
  - To optimize its usage

Grid computing is a network based computational model that has the ability to process large volumes of data with the help of a group of networked computers that coordinate to solve a problem together.

- The main function of grid computing is job scheduling using all kinds of computing resources where a task is divided into several independent sub-tasks and each machine on a grid is assigned with a task.

- Grid computing is based on a distributed system which means computing resources are distributed among different computing units which are located across different sites, countries, and continents.

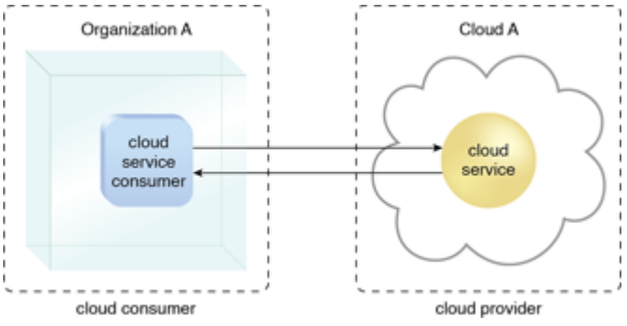
- The grid acts as a distributed system for collaborative sharing of resources. Cloud computing, on the other hand, is a form of computing based on virtualized resources which are located over multiple locations in clusters.

Read more: Difference between Grid Computing and Cloud Computing | Difference Between
http://www.differencebetween.net/technology/difference-between-grid-computing-and-cloud-computing/#ixzz6YrxqqVi5

Roles in the cloud

- **Cloud providers**: provide a hardware infrastructure and a set of services on top
  - Eg: Amazon Web Services (AWS), Microsoft Azure, CloudWatt
- **Cloud Clients**: use resources of cloud platforms
  - Eg: private users, companies
- Cloud resellers: build and sell services, relying on existing cloud plaforms
  - They are both cloud users and cloud providers
  - Eg: RightScale, Scalr
- Cloud developers: produce tools (deployment, self-repair, etc) for the cloud
  - Eg: VMware, research labs & companies (Roboconf)

To illustrate:



You can buy and sell, that makes you a **buy and sellam.**

For those who are interested in a career in cloud computing, you can also become a cloud developer, that means you implement tools for the cloud.

Costs reduction

- For the client
  - Equipments, administration, place, energy, licences ...
- For the provider (data center)
  - Do more with less (profits)
  - Especially energy

Major benefit for the client: pay as you go

- Easy to use and flexible
  - Allocation/de-allocation (fast and on demand) of resources
  - No administrative procedure
  - Accessible from everywhere via internet (24/7)
- Reduced TCO (Total Cost of Ownership)
  - No need of important investments
  - Less staff, no need to manage a local infrastructure
  - Less software licences to pay
- Billing
  - According to usage
  - Monthly/annually
  - Discount VM (spot VM)

Reduced IT costs. Moving to cloud computing may reduce the cost of managing and maintaining your IT systems. Rather than purchasing expensive systems and equipment for your business, you can reduce your costs by using the resources of your cloud computing service provider.

Scalability. Your business can scale up or scale down your operation and storage needs quickly to suit your situation, allowing flexibility as your needs change. Rather than purchasing and installing expensive upgrades yourself, your cloud computer service provider can handle this for you.

Access to automatic updates. Access to automatic updates for your IT requirements may be included in your service fee.

Collaboration efficiency. Collaboration in a cloud environment gives your business the ability to communicate and share more easily outside of the traditional methods.

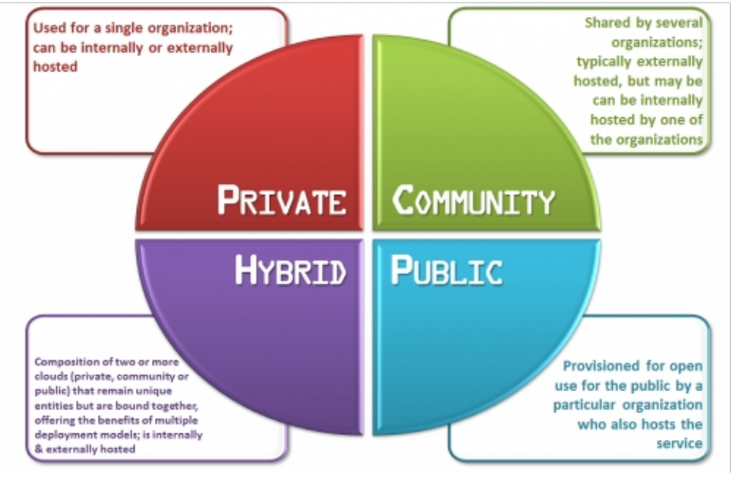
Classification wrt property

- **Community cloud**: the cloud is built to be shared between several organizations

  Eg: UnivCloud
- **Private cloud**: the cloud is built by a company (or an institution) for its internal use
- **Public cloud**: the cloud is built by a company for business purpose and it is open to all users

  Eg: Amazon web services, Microsoft Azure, Eolas, etc.
- **Hybrid cloud**: combination of a private cloud, extended with one or several public clouds

Nice figure



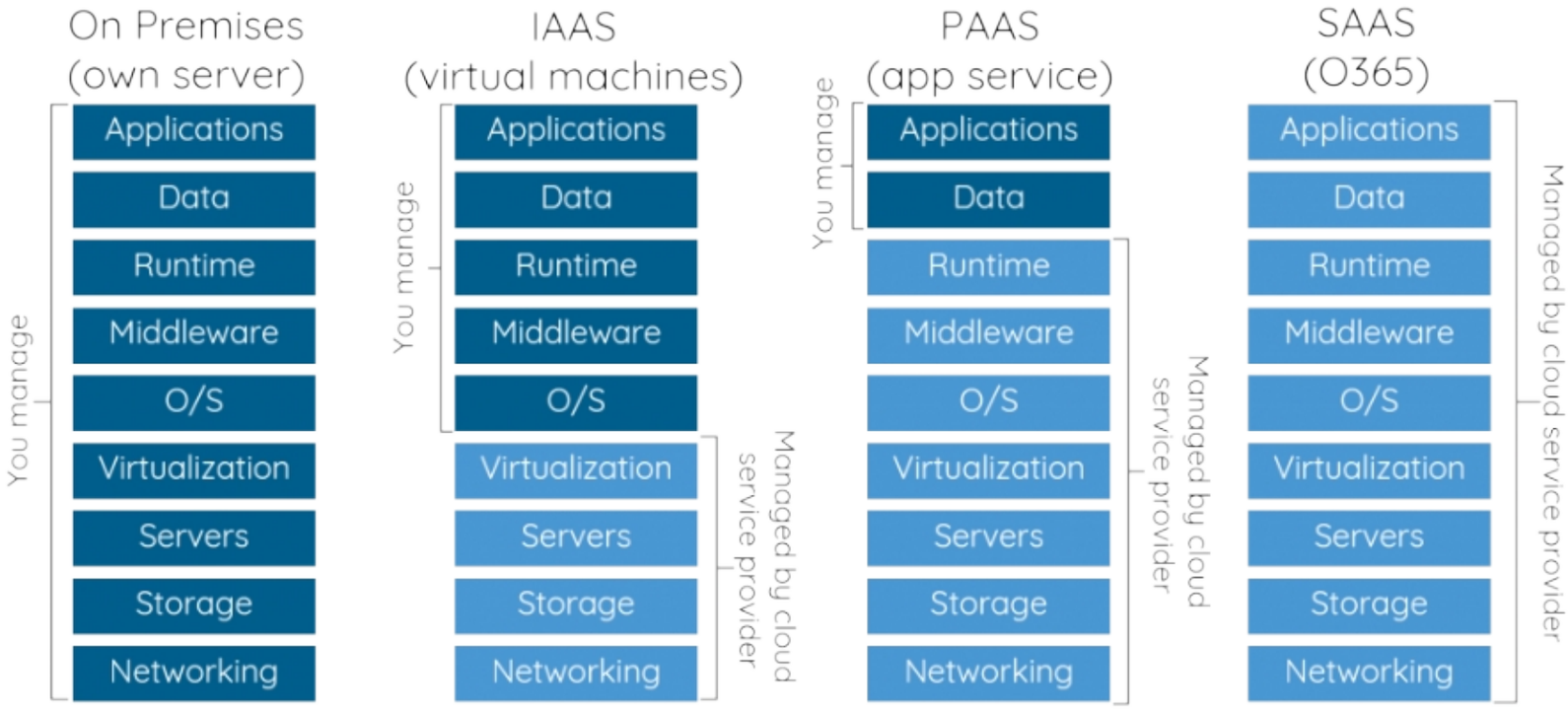
Classification wrt provided service

- **IaaS (Infrastructure as a Service)**: the cloud provides storage and computing facilities. Users can rent machines (or virtual machines)
  - Ex. : Amazon EC2, Windows Azure
- **PaaS (Platform as a Service)**: the cloud provides a platform for the construction and execution of applications in the underlying infrastructure
  - Ex. : Google App Engine, Windows Azure web role
- **SaaS (Service as a Service)**: the cloud directly provides the application that the user needs
  - Ex. : Google docs, SalesForce

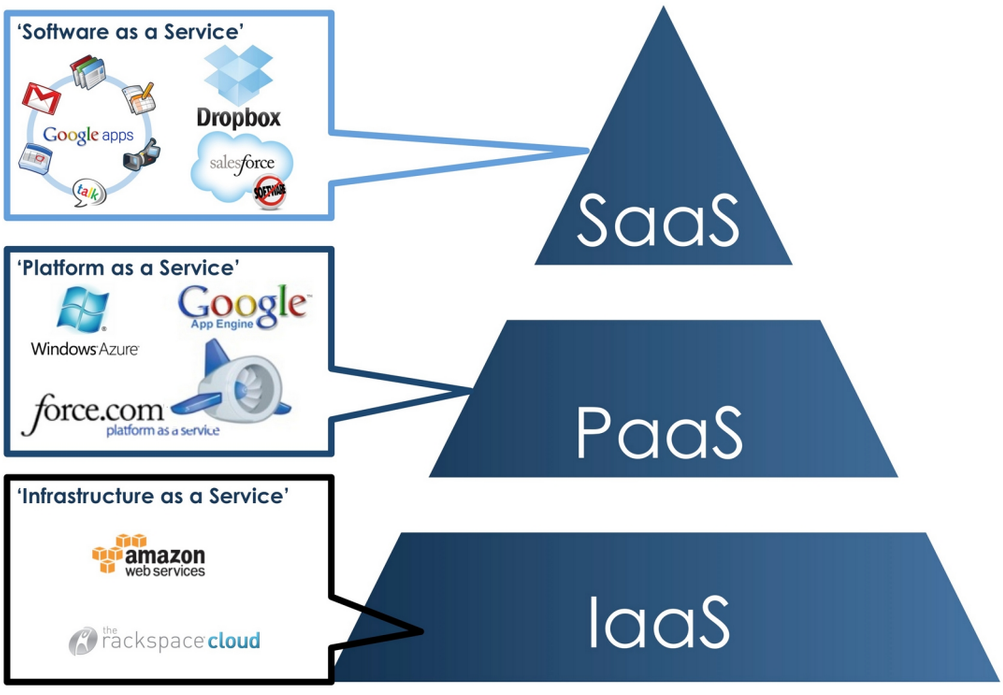
From IaaS to SaaS, this is a story of responsibility transfer from the client to the provider

Classification of cloud systems according to the services they offer. SaaS allows users to run online applications. The vendors own the applications and the users pay a fixed subscription fees. PaaS allows users to create their own cloud applications, providing all the execution and compilation of software as well as operating systems. IaaS allows users to run any applications they want to on cloud hardware of their choice.

Classification wrt service

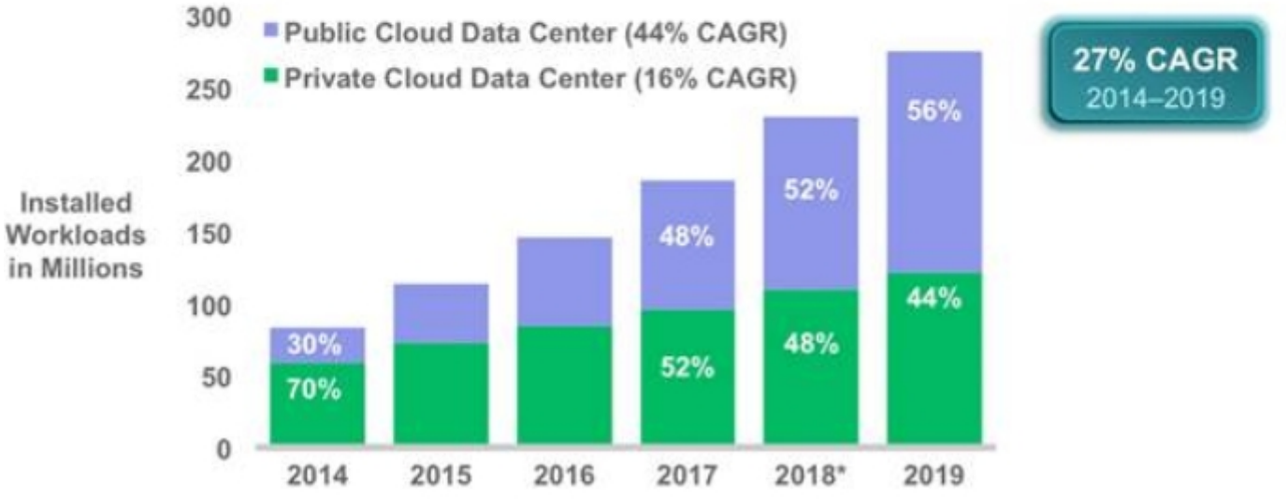


Examples:



Some numbers

- Fast growth of the number of datacenters
  - Salesforce (2009) : 54.000 companies on 1000 servers
  - Amazon EC2 (2012) : 500.000 servers
  - Google (2011) : 900.000 servers
- Energy (2017) :
  - 10% of the electric production for IT
  - 1,8% for datacenters



Some numbers

- Cloud providers (IaaS) in the world

Several datacenters in the world.
You can choose to deploy your workload on the data center close to your location.
They are everywhere except in Africa. The data centers are in Wakanda which is not visible on our maps.

In the heart of clouds

Microsoft Azure Dublin (Ireland)

Opened in 2009, the Microsoft Azure datacenter represents a capacity of 39 000 m^2 of computing rooms. The equivalent of 8 football fields. 80 Microsoft employees work on this site.

Microsoft invests US$230m in Dublin data center expansion. Microsoft said the second expansion at the data center, which first opened in 2009, will increase the facility's computing capacity by 169,000 sq ft.

In the heart of clouds

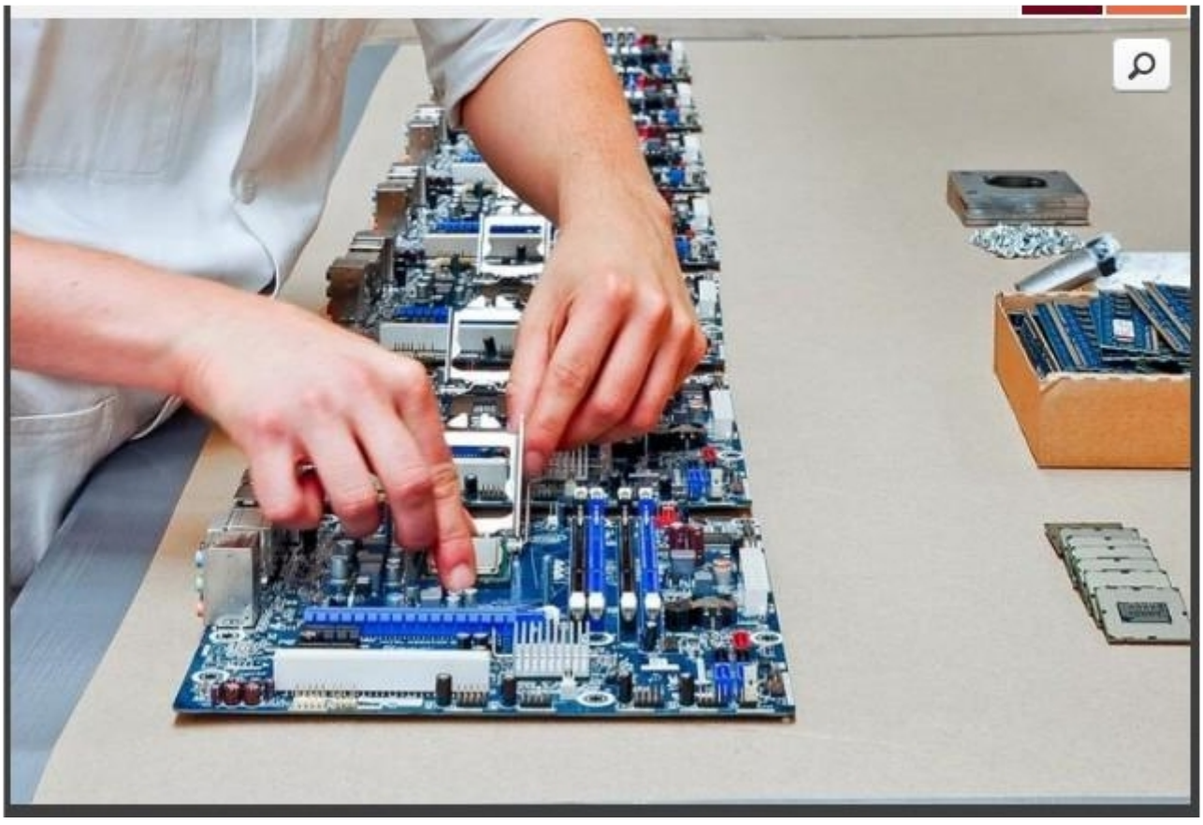
Google IOWA (USA)

The Google datacenter of Council Bluffs, IOWA, covers more than 10 000 m^2. It is used to host Google's search engine and YouTube.

In 2007, google announced plans to construct a data center complex in Council Bluffs, Iowa. Then in 2012 and 2015, they announced expansions of the data center facilities. Over time, they have invested over $2.5 billion

In the heart of clouds

OVH, Google, Facebook build their own servers

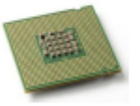


When the staff at OVH's newest data center need to install a new server, they don't need to go far to find one. The huge web hosting firm builds its own custom servers on-site at its facility , creating perhaps the shortest supply line in the industry.

Challenges of cloud

- Security & Trust (mentionned by 70 % of queried people in an IDC study)
  - Where are stored my data ?
  - Is privacy ensured ?
  - What about the laws in these countries ?
- Service guarantee (SLA = Service Level Agreement)
  - The SLA should be enforced
  - The client should trust this guarantee
- Impact on energy: in 2017, 1.8% of worldwide electricity consumption
  - For powering servers
  - For cooling
  - Generate more CO2 than air transport

vCPU
? = ?



Cloud computing presents many unique security issues and challenges. In the cloud, data is stored with a third-party provider and accessed over the internet. This means visibility and control over that data is limited. It also raises the question of how it can be properly secured. It is imperative everyone understands their respective role and the security issues inherent in cloud computing.

Ensure a SLA is challenging because, In circumstances where multi-tenancy is in use, SLA relating to service separation and high obtainability may be challenging to for the cloud service provider.

Energy challenge also. Data centers need a lot of energy.

Cooling

Google IOWA (USA)

In the datacenter of The Dalles, pipes are binging and evacuating water. This is the cooling system of the datacenter: blue pipes for cold water and red pipes for hot water.

2 % of electricity in the US, half for cooling

Cooling

Google Hamina (Finland)

For cooling its datacenter, Google installed water tanks that can store up to 900 000 liters. In some place like here in Finland, sea water is used.

Water cooling. Any cooling technology that uses liquid to evacuate heat from the air. Increasingly, data center liquid cooling refers to specifically direct cooling solutions that expose server components (such as processors) to liquid to cool them more efficiently.

Cooling

Microsoft

For cooling its servers, Microsoft relies on air cooling. Air is captured bay containers right atop racks, on the roof of the datacenter, in order to reduce the distance between capture and injection.

Challenges of cloud

- Precise monitoring for billing
  - The cloud is multi-tenant (shared resources)
  - Billing should be reliable
- Standardization
  - Each provider develop its own APIs
  - How to interoperability and portability ?
- Customization (eg. hardware)
  - Very large companies develop specific hardware for private clouds
  - How can it be used in public (multi-tenant) clouds ?

Cloud monitoring is the process of reviewing and managing the operational workflow and processes within a cloud infrastructure or asset. It's generally implemented through automated monitoring software that gives central access and control over the cloud infrastructure.

The cloud should be elastic

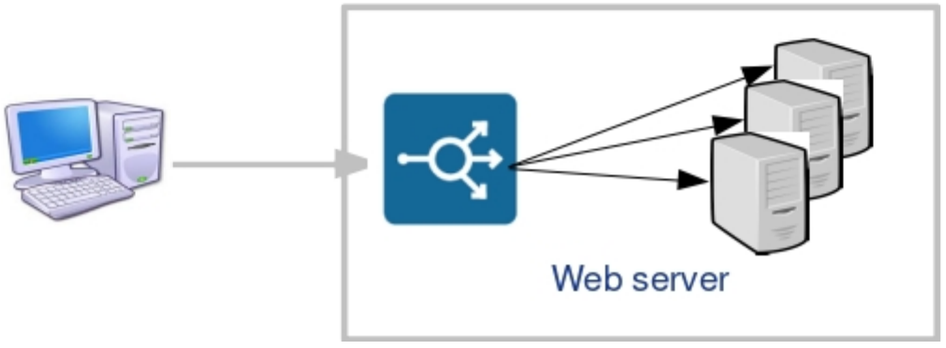
- Fluctuation of the load
  - Applications in the cloud are more or less used over time
  - Examples
    - An e-commerce application
    - Computation of payslips for a company
- For the client : elastic applications
  - Avoid over-booking
  - Allocate resources (machines) on demand
    - Allocation/de-allocation
  - Reduce costs
- For the provider: elastic cloud
  - Optimize his infrastructure
  - Adapt his spendings according to sold resources (especially energy)

In cloud computing, elasticity is defined as "the degree to which a system is able to adapt to workload changes by provisioning and de-provisioning resources in an autonomic manner, such that at each point in time the available resources match the current demand as closely as possible".

Elastic applications

- An elastic application
  - We augment its capacity by adding machines
  - We reduce its capacity by removing machines
  - The application should adapt according to the load (like a lung)

- Example of elastic application
  - Replicated web server
  - Load balancing
  - Adaptation of the number of replica
    - Manage automatically by a tool



Elastic applications can allocate and deallocate resources on demand for specific application components. For example, during its lifetime, a data storage tier of an elastic application might add and remove data storage points due to cost and performance requirements.

Virtualisation: motivations

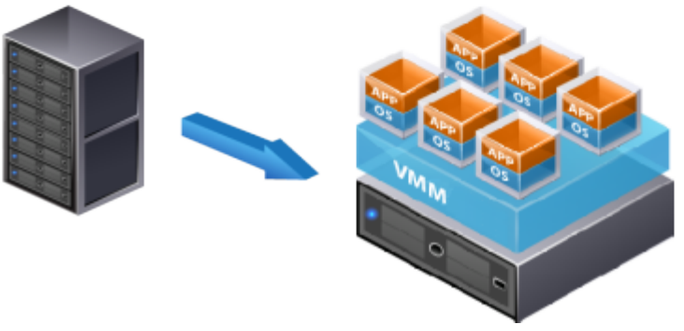
- Historically
  - A cloud was selling physical machines
  - A machine was used by one user only
- Wasting
  - Many under-loaded applications (only one machine)
  - An application uses in average 10-15 % of its resources (source VMware)

Me: Do you track server and CPU utilization?
Wall Street IT Guru: Yes
Me: So it's a metric you report on with other infrastructure KPIs?
Wall Street IT Guru: No way, we don't put it in reports. If people knew how low it really is, we'd all get fired.

Virtualization is the process of running a virtual instance of a computer system in a layer abstracted from the actual hardware. Most commonly, it refers to running multiple operating systems on a computer system simultaneously. To the applications running on top of the virtualized machine, it can appear as if they are on their own dedicated machine, where the operating system, libraries, and other programs are unique to the guest virtualized system and unconnected to the host operating system which sits below it.

Virtualisation: principles

- Virtual machines (VM)
  - Simulate several machines (virtual) on one machine (physical)
  - Users are allocated virtual machines
- Challenges: isolation
  - Security: a VM is protected against potential attacks from other VMs
  - Performance: one VM's performance is not affected by other VMs
  - Failure: one VM's failure should not affect other VMs

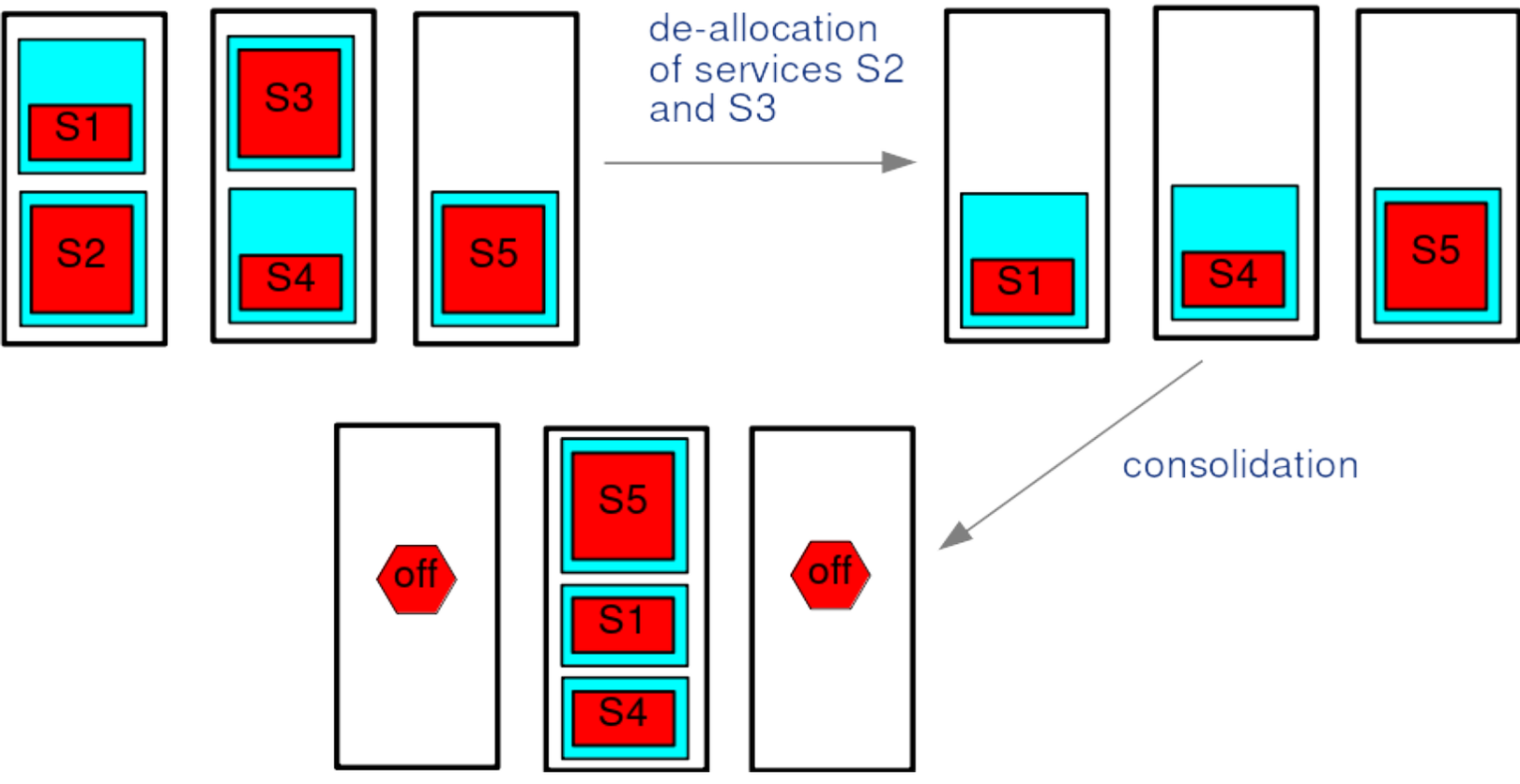


A virtual machine is the emulated equivalent of a computer system that runs on top of another system. Virtual machines may have access to any number of resources: computing power, through hardware-assisted but limited access to the host machine's CPU and memory; one or more physical or virtual disk devices for storage; a virtual or real network inferface; as well as any devices such as video cards, USB devices, or other hardware that are shared with the virtual machine.

An elastic cloud

- Better physical machine occupation
  - With several VMs, the capacity of a physical machine can be fully used
  - In some way, the provider sells fractions of machines
- Consolidation of physical machines
  - VMs can be moved between physical machines (migrated)
  - We can pack VMs on physical machines (according to resources really used)
  - Unused physical machines can be switched-off or suspended
  - Energy savings
- Ideally
  - Physicall machines are used at 100 % of their capacity
  - Other machine are off
  - A provider can sell more resources than available in the datacenter (overbooking as in airline companies)

Consolidation



Server consolidation is the process of migrating VMs from multiple computers to a singular computer.

As illustrated in the above figure, after the migrations two servers can be turned off.

Is the cloud panacea ?

- Opposition
  - Centralization
    - The cloud
    - Passive terminals
  - Decentralization
    - Personal computers
    - Opensource software
- Richard Stallman
  - Cloud computing = careless computing