Introduction to cloud computing

Daniel Hagimont

hagimont@enseeiht.fr

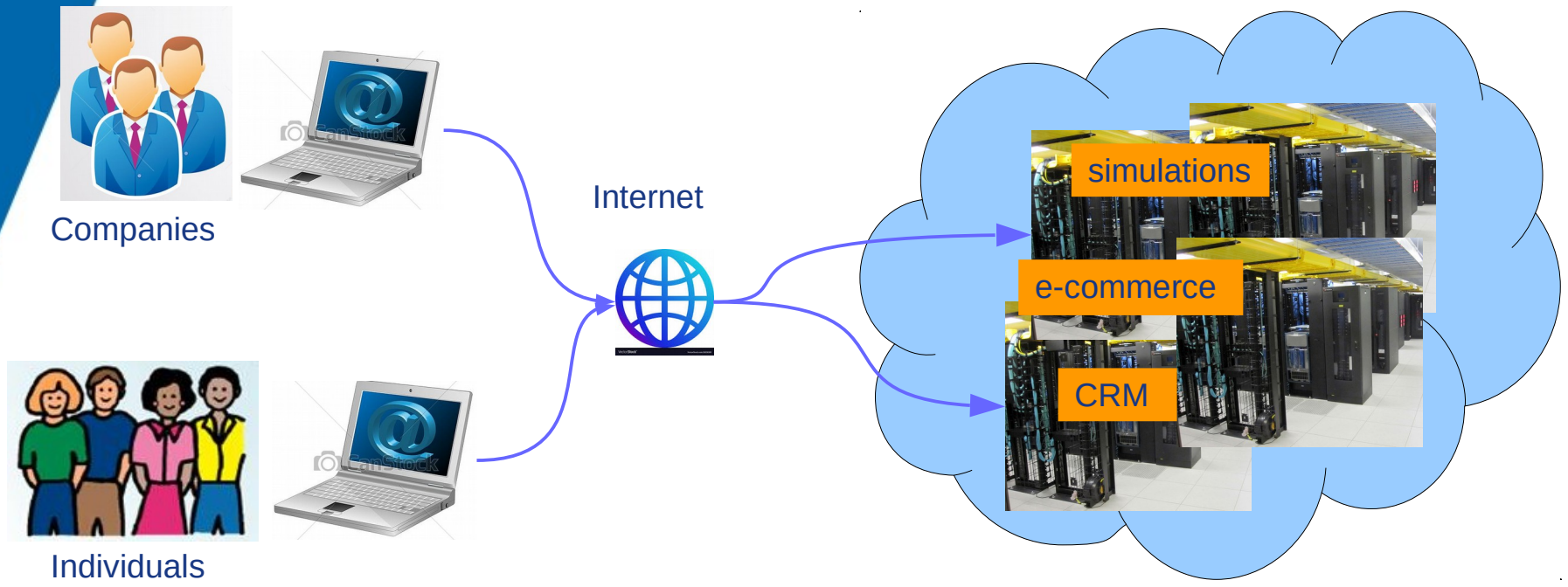
Cloud computing: the idea

- 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

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)

General scheme : externalization



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

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

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

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)

Costs reduction

- For the client
  - Less investments: 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)

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

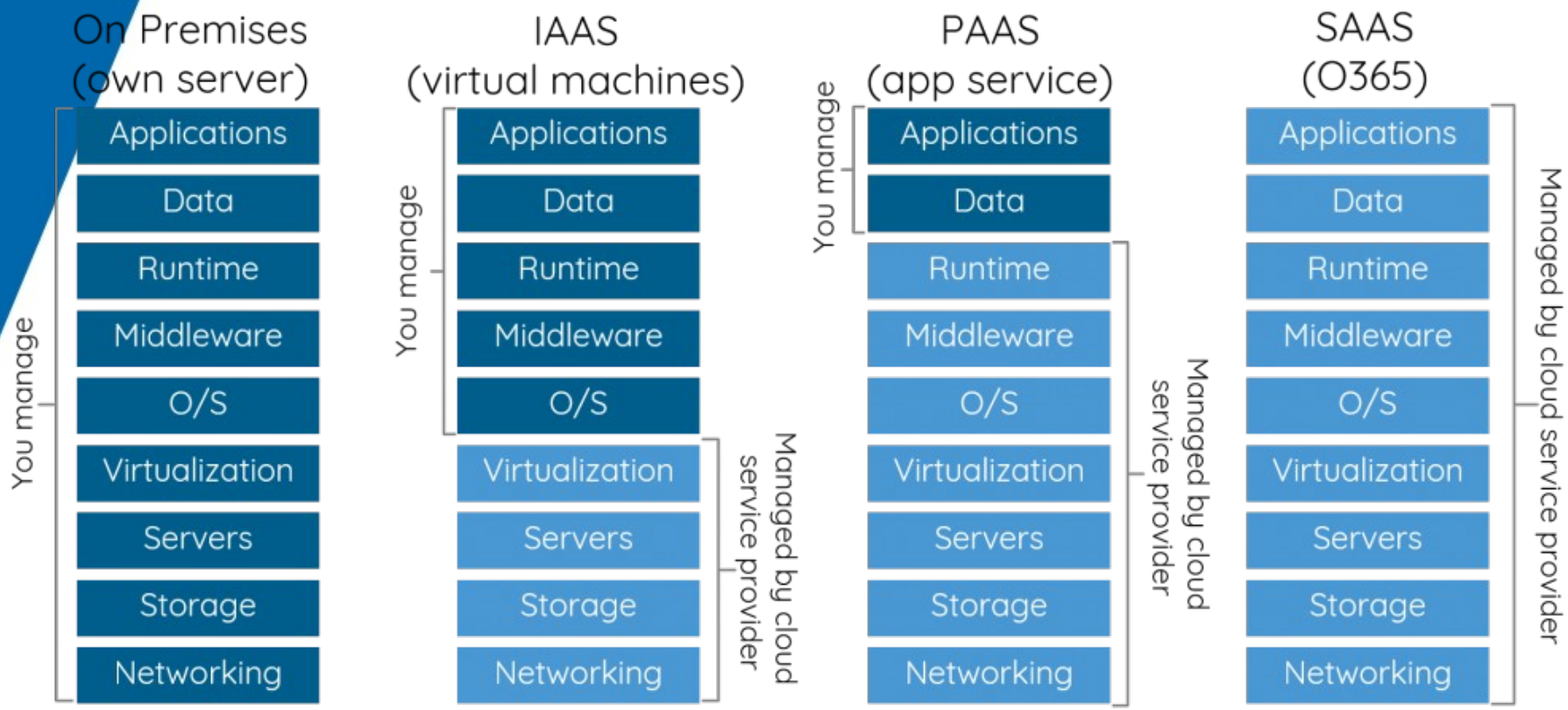
  Eg: Amazon web services, Microsoft Azure, Eolas, etc.
- **Hybrid cloud**: combination of a private cloud, extended with one or several public clouds

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

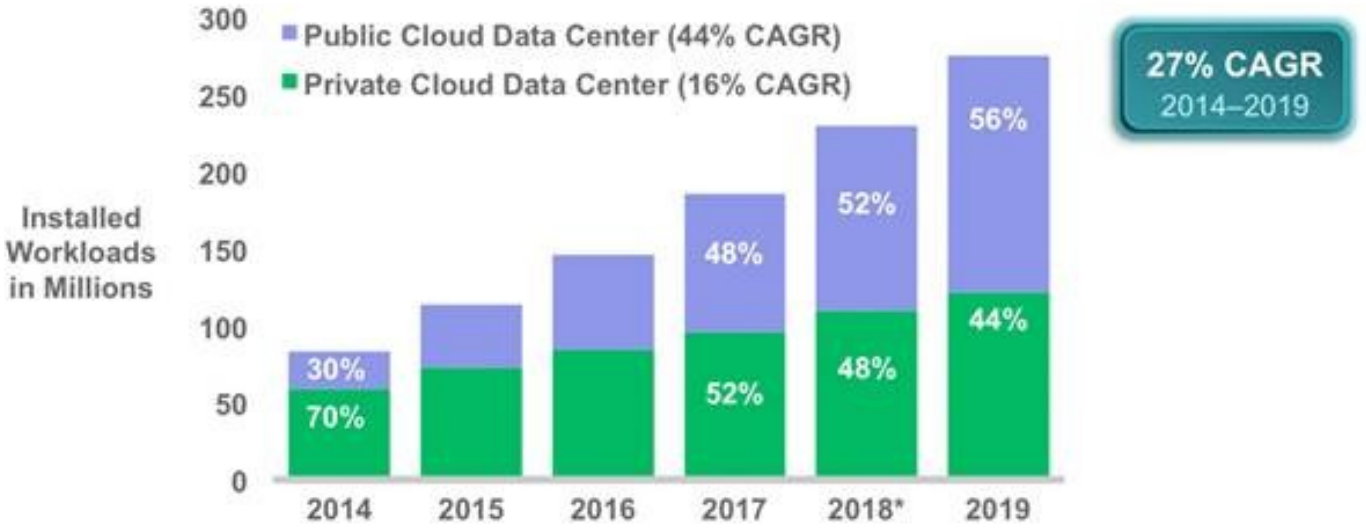
From IaaS to SaaS, this is a story of responsibility transfer from the client to the provider

Classification wrt service



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



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.

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.

Cooling

2 % of electricity in the US, half for 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.

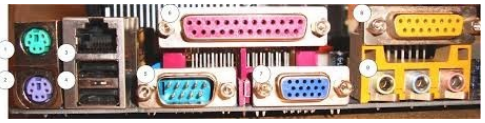
Challenges of cloud

- 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
? = ?

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 ?



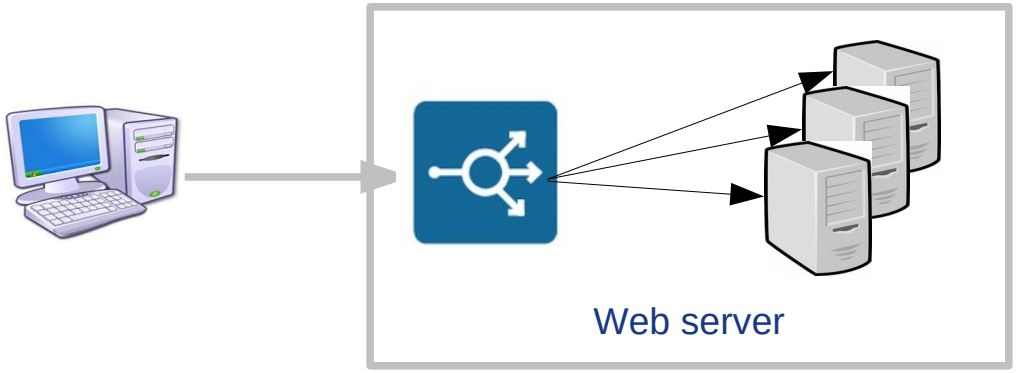
The cloud should be elastic

- 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)

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



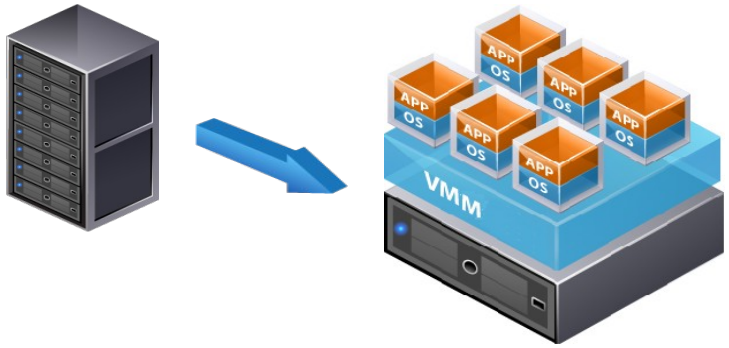
Virtualisation: motivations

- 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.

Virtualisation: principles

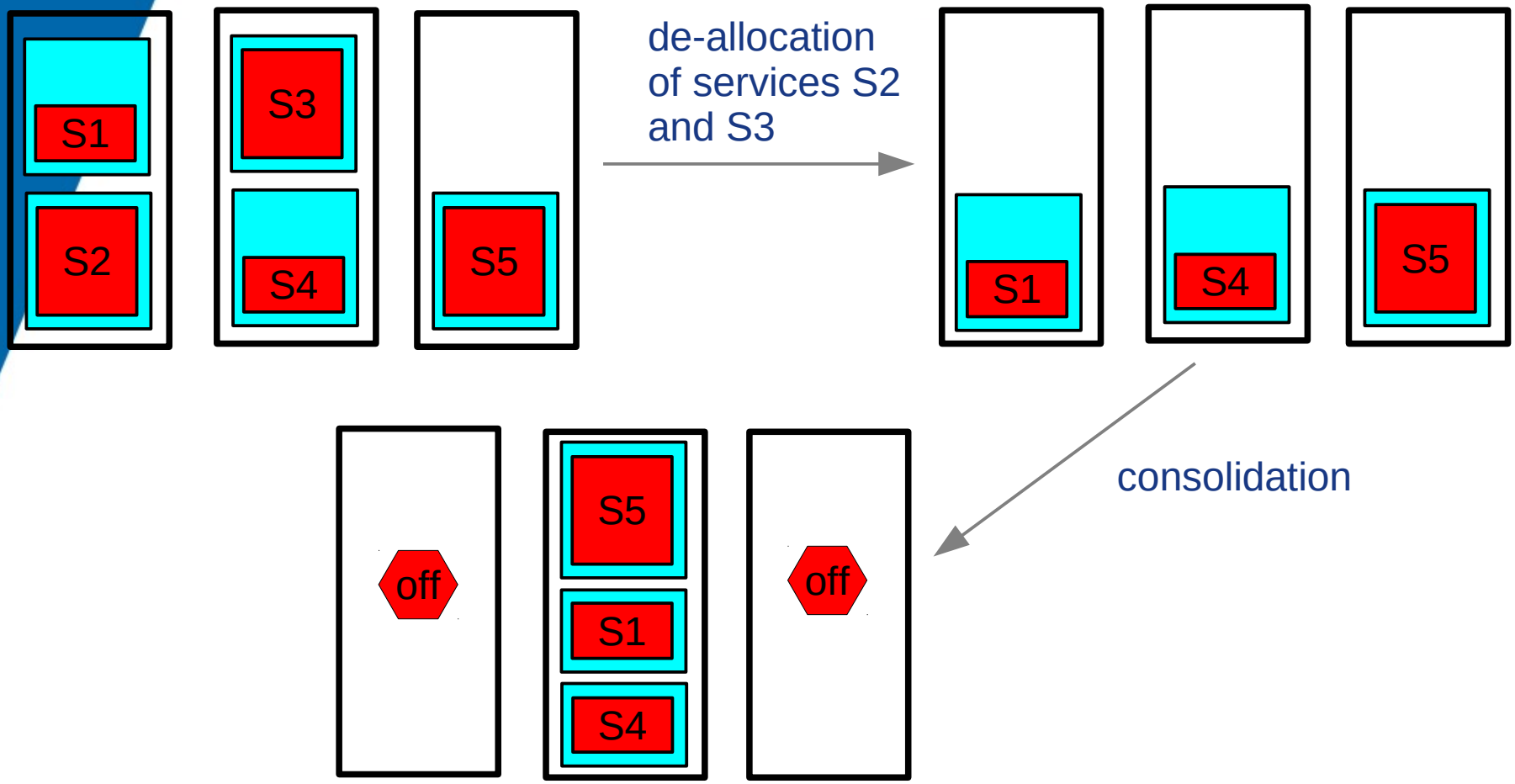
- Virtual machines (VM)
  - Simulate several machines (virtual) on one machine (physical)
  - Users are allocated virtual machines
- Challenges: isolation
  - Security: a VM is protected against potential attacks from other VMs
  - Performance: one VM’s performance is not affected by other VMs
  - Failure: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



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



Is the cloud panacea ?

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