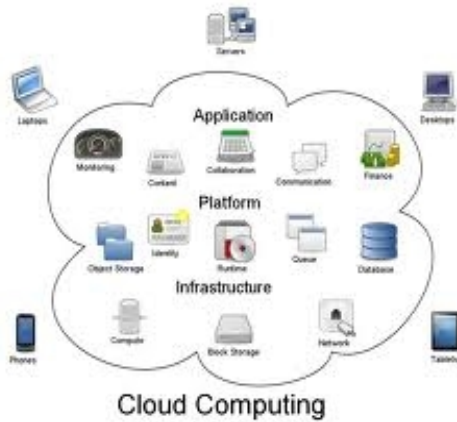




Services for the cloud

Boris Teabe

boris.teabe@inp-toulouse.fr



Some services in the cloud

Services are provided to ease the management of cloud applications

- Development
- VM image creation
- Deployment
- Storage
- Administration

Development services

Various forms

- Programming or integration models (PaaS)
 - Ex: Beantalk from AWS, web and worker role from Azure, Roboconf
- Plug-ins for IDE: deployment and test in the Cloud
 - Ex: Windows Azure via Visual studio, Google App Engine via Eclipse
- REST APIs allowing to perform actions in the cloud
 - Ex: AWS API. Roboconf relies on these APIs

VM image creation services

- Construction and generation of custom VM images
 - Installation in a standard VM and save image
 - Customization of a minimalized OS
 - Ex: Just-enough OS (JeOS)
- Unified deployment in heterogeneous clouds
 - Languages for describing appliances
 - Ex: UshareSoft (OVF format: Open Virtual Format)
 - Conversion between virtualization solutions

Deployment services

- Application installation (binaries)
 - Dynamic installation on standard VMs
 - Language for describing installations
 - Construction of VM images
- Application configuration and launch
 - Centralized or distributed
 - Language for describing configuration and launch

Tools like **Cloudify**, **Roboconf** provide such services

Storage

- Various usages
 - Data used by applications
 - Managed by the Cloud (ex: S3)
 - Managed by the user in VMs
 - VM images or snapshots
- Expected properties
 - Durable
 - Privacy
 - Available anytime
 - Accessible from anywhere

Administration services

- Tools for enabling dynamic administration

- Monitoring
- Reconfiguration
- Fault tolerance
- Scalability
- Security
- Accounting



Tools like **Cloudfify**, **Roboconf** provide such services

Replication and load-balancing

- Replication of a server on several machines
 - Allow to tolerate a machine failure
 - Allow scalability
- Load-balancing of requests between servers
 - Distribution according to the capacity and load of servers
- Different types of protocols (TCP, HTTP, JDBC ...)
- Example: web server or database

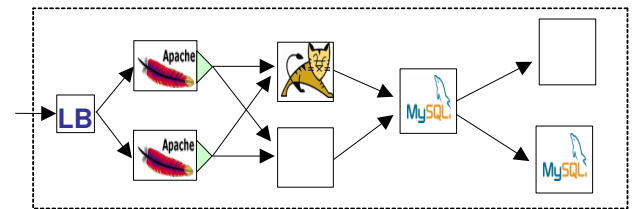


Monitoring

- Observation of runtime conditions
- Detection of particular events
 - Failure of an application or a VM
 - Overuse of a resource
 - Degradation of the Quality of Service
 - Intrusion in an application
- A set of probes in VMs
- A communication infrastructure for gathering metrics
- Used for triggering reconfigurations

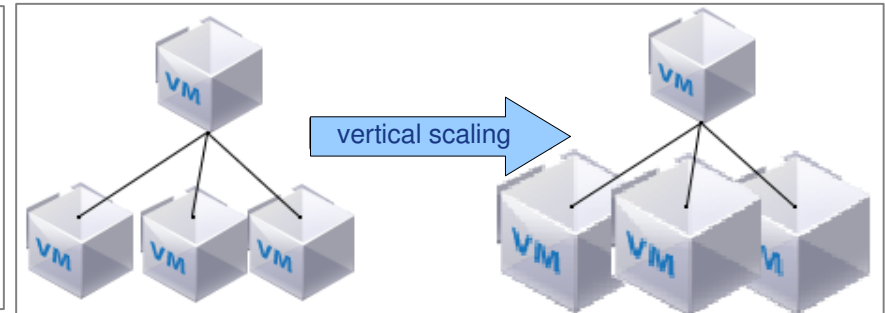
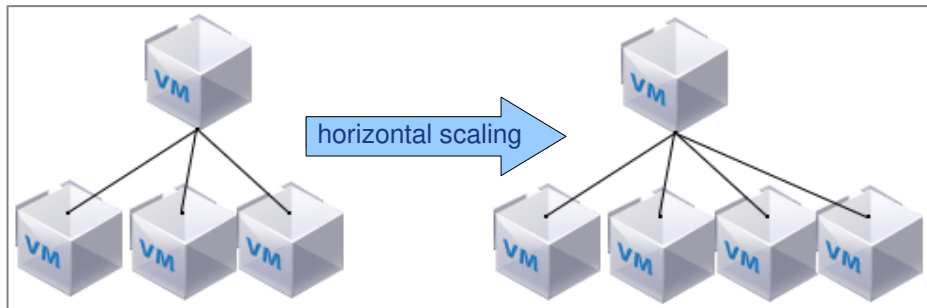
Fault tolerance

- Target application availability
- Two main strategies
 - Anticipation
 - Replication of application tiers
 - Some tiers are still available in case of failure
 - Repair
 - Failure detection
 - Restart of the failed tier (application or VM)
 - Need frequent backups (replicated)
 - Prevent disaster
 - Geo-replication



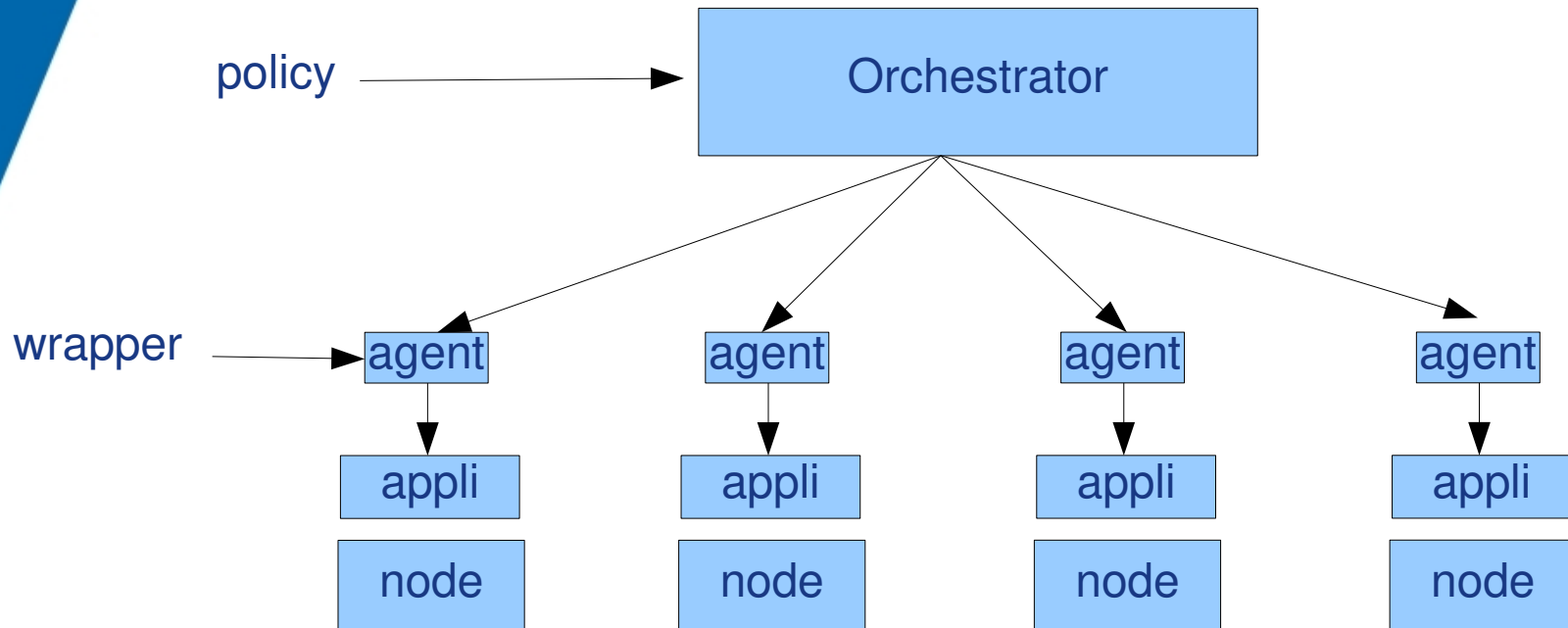
Scalability (or elasticity)

- Detection of overload situations
- Allocation of additional resources
- Two strategies
 - Allocation of additional resources to the VMs which host the application (vertical scaling)
 - More memory or CPU
 - Creation of a new tier (VM and application) (horizontal scaling)
 - Load-balancer based structure



Few characteristics

- Architecture

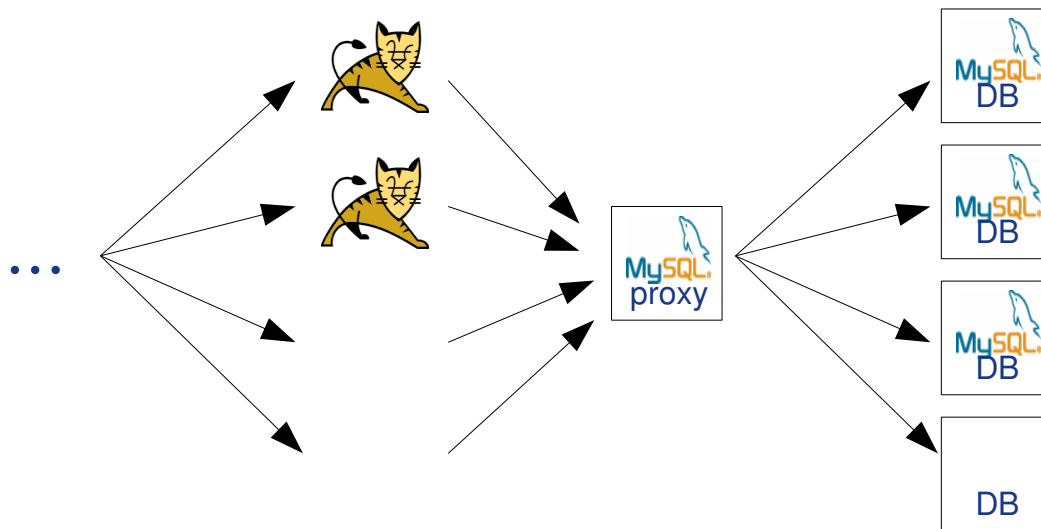


Few characteristics

- Orchestration language
 - Description of wrappers
 - Actions on applications (effectors)
 - Observations on applications (sensors)
 - Description of policies
 - Applications to install/deploy
 - Reconfiguration rules
 - Control loops (sensor → decision → effector)

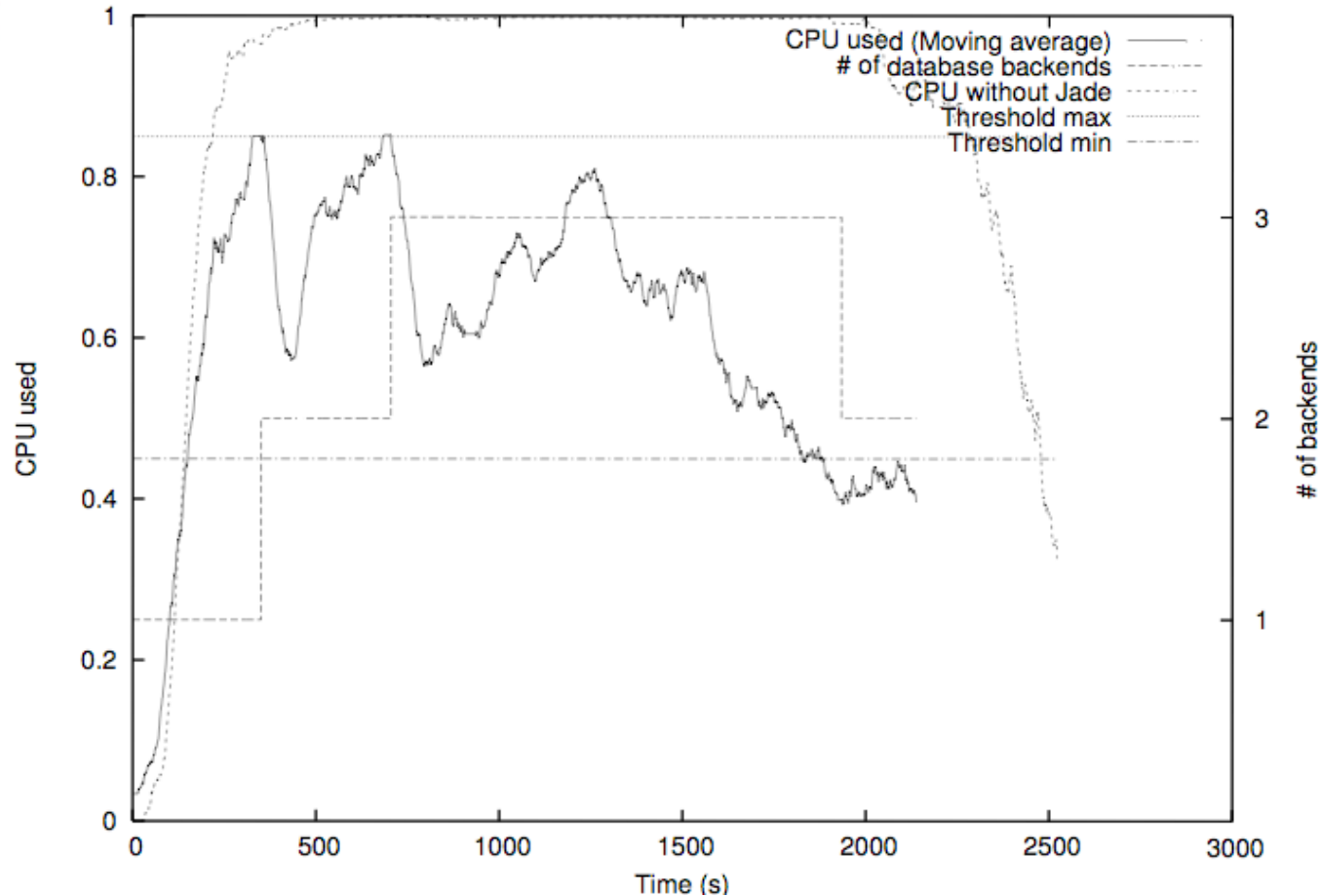
Example: self-optimizing

- Experiment with Jade (research prototype) 15 years ago



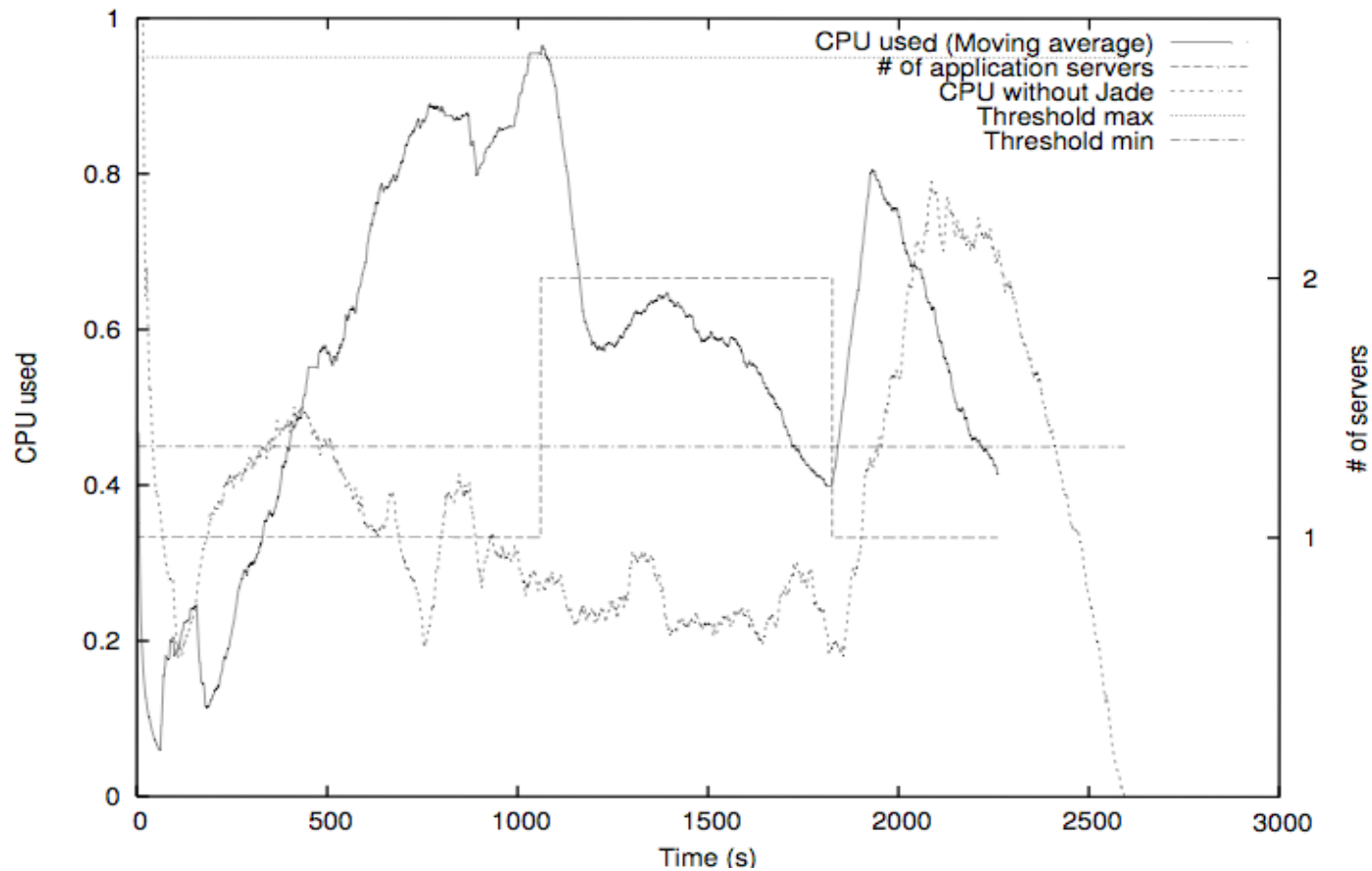
Example: self-optimizing

- Dynamic dimensioning on MySQL



Example: self-optimizing

- Dynamic dimensioning on Tomcat



Products

- Fabric, Cuisine, Chef, Crowbar, Puppet, Pallet, Juju, Ansible, (R)ex, SaltStack, Distelli, CFEngine, Foreman, Commando, Rundeck, Bcfg2, Glu, Pontus Vision, Tstconfig, Dell Cloud Manager, AutoMate, Cloudify, CA Technologies Automation, Cisco Prime Cloud Automation, Qualys, BladeLogic Automation Passport, Cobbler, Sprinkle, Cirba, Mist.io, Scalr, Rdist, HP Cloud Service Automation, ServiceNow, Appcore Cloud Management, Citrix CloudPortal Business Manager, ManageEngine Applications Manager, DivvyCloud, ...
- <https://blog.profitbricks.com/48-best-cloud-tools-for-infrastructure-automation/>

Products

- Very popular
 - Chef
 - <https://www.chef.io/>
 - Puppet
 - <https://puppet.com/>
 - Ansible
 - <https://www.ansible.com/>
 - Low level, install/deployment, not dynamic
- Advanced (dynamic)
 - Cloudify
 - <http://getcloudify.org/>
 - Roboconf
 - <http://roboconf.net/>

Examples of public clouds



AWS (IaaS/PaaS/SaaS)

- Initially created to exploit (sell) unused resources in Amazon's infrastructure
- Public cloud platform
 - Initially a IaaS
 - Propose currently a PaaS and a SaaS
- Objectives
 - Provides a web portal where you can buy different types of resources with different sizes

AWS (IaaS/PaaS/SaaS)








Services ▾





Edit ▾

Amazon Web Services





Compute & Networking

-  **Direct Connect**
Dedicated Network Connection to AWS
-  **EC2**
Virtual Servers in the Cloud
-  **Elastic MapReduce**
Managed Hadoop Framework
-  **Route 53**
Scalable Domain Name System
-  **VPC**
Isolated Cloud Resources







Storage & Content Delivery

-  **CloudFront**
Global Content Delivery Network
-  **Glacier**
Archive Storage in the Cloud
-  **S3**
Scalable Storage in the Cloud
-  **Storage Gateway**
Integrates On-Premises IT Environments with Cloud Storage







Database

-  **DynamoDB**
Predictable and Scalable NoSQL Data Store
-  **ElastiCache**
In-Memory Cache
-  **RDS**
Managed Relational Database Service
-  **Redshift** NEW
Managed Petabyte-Scale Data Warehouse Service

Deployment & Management

-  **CloudFormation**
Templated AWS Resource Creation
-  **CloudWatch**
Resource and Application Monitoring
-  **Data Pipeline**
Orchestration for Data-Driven Workflows
-  **Elastic Beanstalk**
AWS Application Container
-  **IAM**
Secure AWS Access Control
-  **OpsWorks** NEW
DevOps Application Management Service

App Services

-  **CloudSearch**
Managed Search Service
-  **Elastic Transcoder** NEW
Easy-to-use Scalable Media Transcoding
-  **SES**
Email Sending Service
-  **SNS**
Push Notification Service
-  **SQS**
Message Queue Service
-  **SWF**
Workflow Service for Coordinating Application Components

AWS (IaaS/PaaS/SaaS)

Amazon EC2 (IaaS)

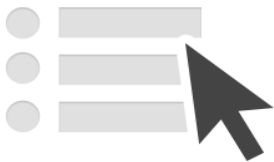
- Amazon's IaaS
- Spread over several continents: America, Europe, Asia
- Virtualized
 - The para-virtualized version of Xen, but switch to KVM
- 750 hours for free when you create an account
- Several types of VMs (14)
- Pay as you go, but also spot VMs
- Marketplace for custom VMs
- Libraries for developers

AWS (IaaS/PaaS/SaaS)

Beanstalk (PaaS)

- Rapid deployment of web applications
- Automatic management
 - Load-balancing
 - Auto-scaling (horizontal scalability)

Get Started in Three Easy Steps



Select a Platform



Upload an Application or Use a Sample



Run it!

Start Now by Selecting Your Platform



Windows Server 2012



node



...and more

Tomcat

Let's Go!

Windows Azure (IaaS/PaaS/SaaS)

- Services
 - IaaS
 - Instantiation of VMs (including Linux)
 - PaaS (core business)
 - Construction web n-tiers applications (.Net, ASP, PHP)
 - Life cycle management (load balancing, scalability, fault tolerance ...)
 - SaaS
 - Microsoft Dynamic CRM (Office, Xbox, etc.)
 - Storage service
 - DB, CDN, etc.
 - Datacenters spread in Europe, USA and Asia

GoogleApp Engine (GAE) (PaaS)

- Google's PaaS for the construction of web applications
 - Application fully managed
 - Load balancing, scalability
- Host applications such as Google sites, calendar, gmail, etc
- Many services for applications' developers
 - URL Fetch – search of web services
 - Images – Treatment of images: dimensioning, rotation, etc.
 - Mail
 - Datastore – small data storage
 - ...
- SDK and eclipse plugin (development and deployment)
- The application does not see the underlying infrastructure



Examples of systems for private clouds



- OpenNebula
 - From the Reservoir european project (Madrid complutense university)
 - Opensource, available for Linux distributions
 - Allow building and managing a IaaS
 - Used by several industrial groups
 - ESA, Telefonica, China Mobile, etc.
 - Help the management of hybrid clouds
 - Support several virtualization technologies
 - Xen, KVM, Vmware
 - Different types of communication interfaces
 - Security enforced by authentication and access rights
 - Manage VMs' fault tolerance

Examples of systems for private clouds



- OpenStack
 - Managed by Rackspace (a public cloud provider)
 - Open source, available for Linux distributions
 - Allow building and managing both public and private cloud platforms (many fonctionnalités)
 - The most widespread
 - Used by: Rackspace, Cloudwatt, CERN, etc.
 - Well suited for large scale infrastructures
 - High overhead for a small platform
 - Support several virtualization technologies
 - Provides a much advanced administration interface
 - Requires high expertise to adapt its components

Comparison of IaaS management systems

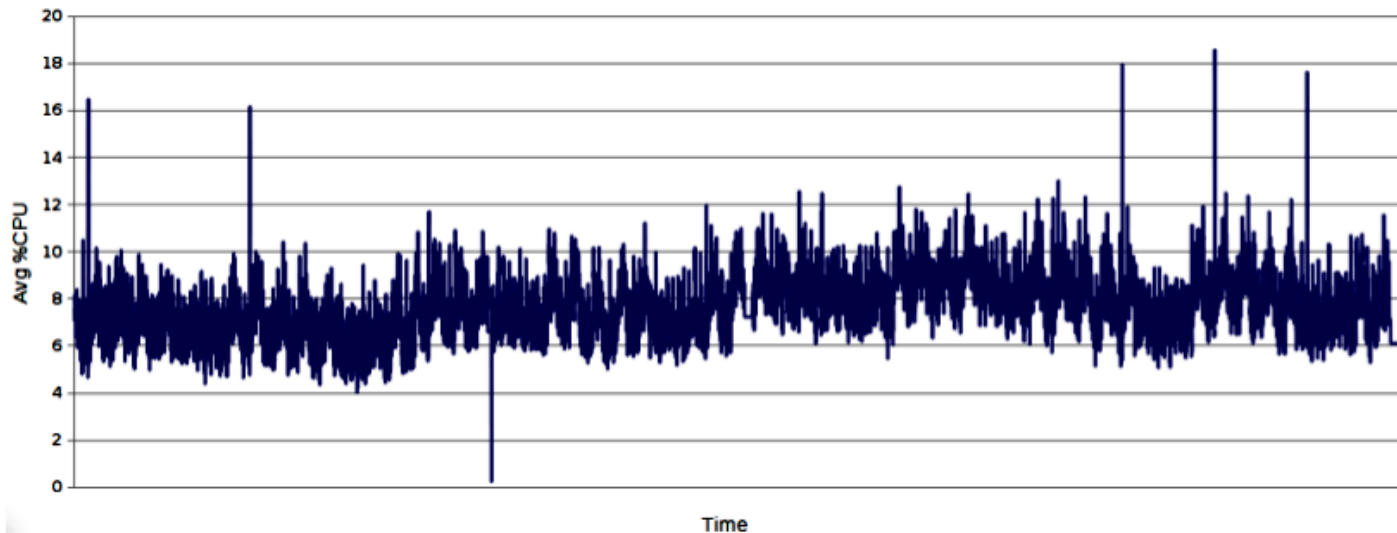
	VMware vCloud Suite	CloudStack	OpenStack	OpenNebula
OpenSource	No	Yes	Yes	Yes
Architecture	Centralized	Centralized	Decentralized	Centralized
Scalability	Yes	Medium	Yes	No
Installation	Require an expertise	Medium	Require an expertise	Easy
Types of hypervisors	A single	Several	Several	Several
Administration interface	Very sophisticated	Medium	Sophisticated	Simple
Interoperability with Amazon	No	Yes	Yes	Yes
Security	Advanced	Advanced	Medium	Simple
VMs management	Much Advanced	Advanced	Much advanced	Medium

One of the main challenges in the cloud

Consolidation for optimized resource management

Server consolidation: motivations

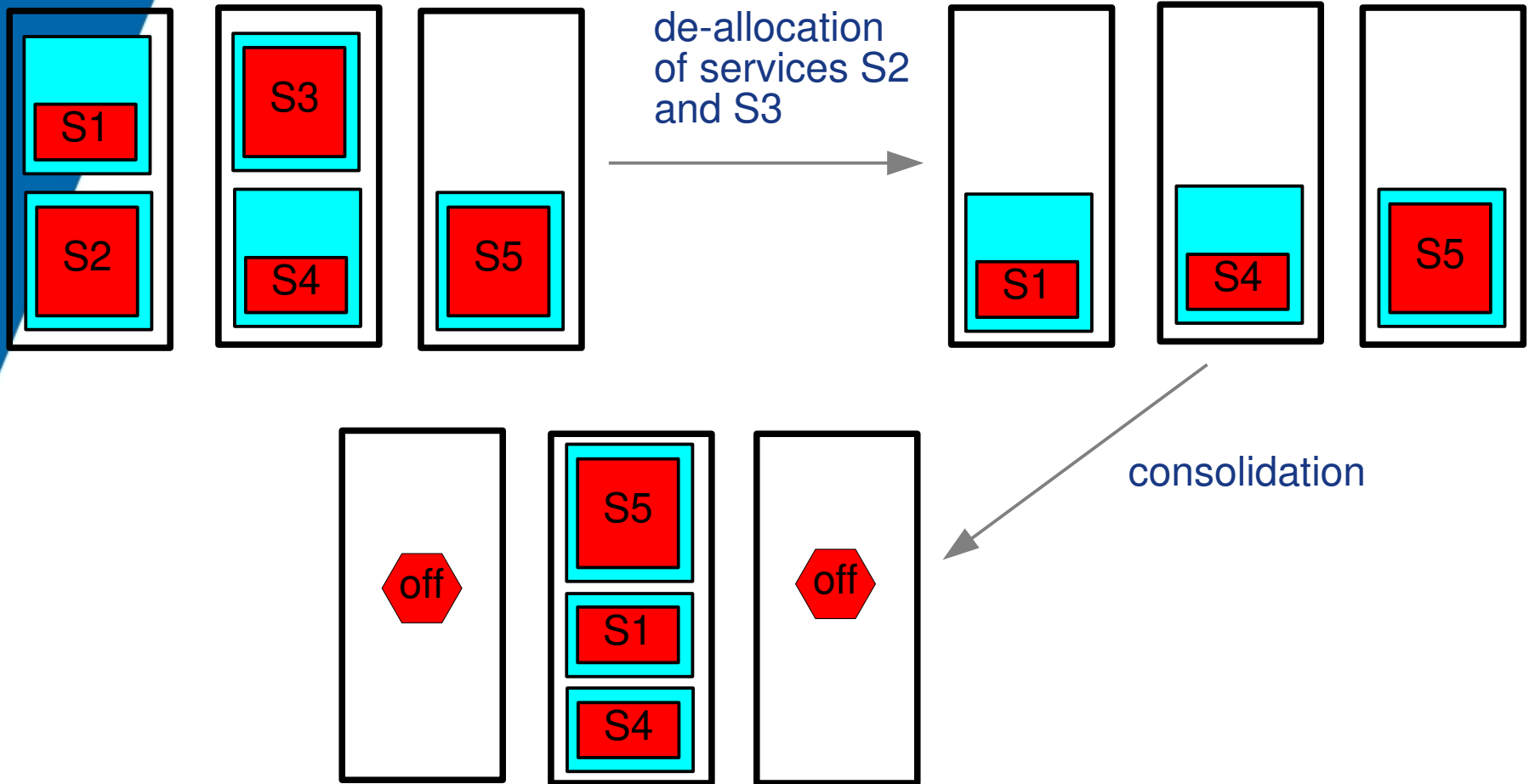
- Resource usage is highly variable
- Average VMs' CPU load in a Eolas cluster
 - Observation over 4 months
 - 805 VMs consolidated on 66 PMs
 - Less than 10 % despites consolidation



Server consolidation: motivations

- Resource usage is highly variable
 - In VMs
 - Unused resources are making holes in physical machines
- The cloud platform is highly dynamic
 - Creation and destruction of VMs
 - Destructions create holes in physical machines

Consolidation



Consolidation

- Role of the consolidator
 - Compute a consolidation plan which minimize the number of used PMs
 - Execute the plan (automatically or after validation by an administrator)
 - Suspend empty PMs
- Live VM migration
 - Displacement of VMs between Pms without service interruption in the VM
 - Has an impact on performance of
 - Migrated VMs
 - VMs on PMs involed in the migration
 - **One of the challenges is to minimize this impact**

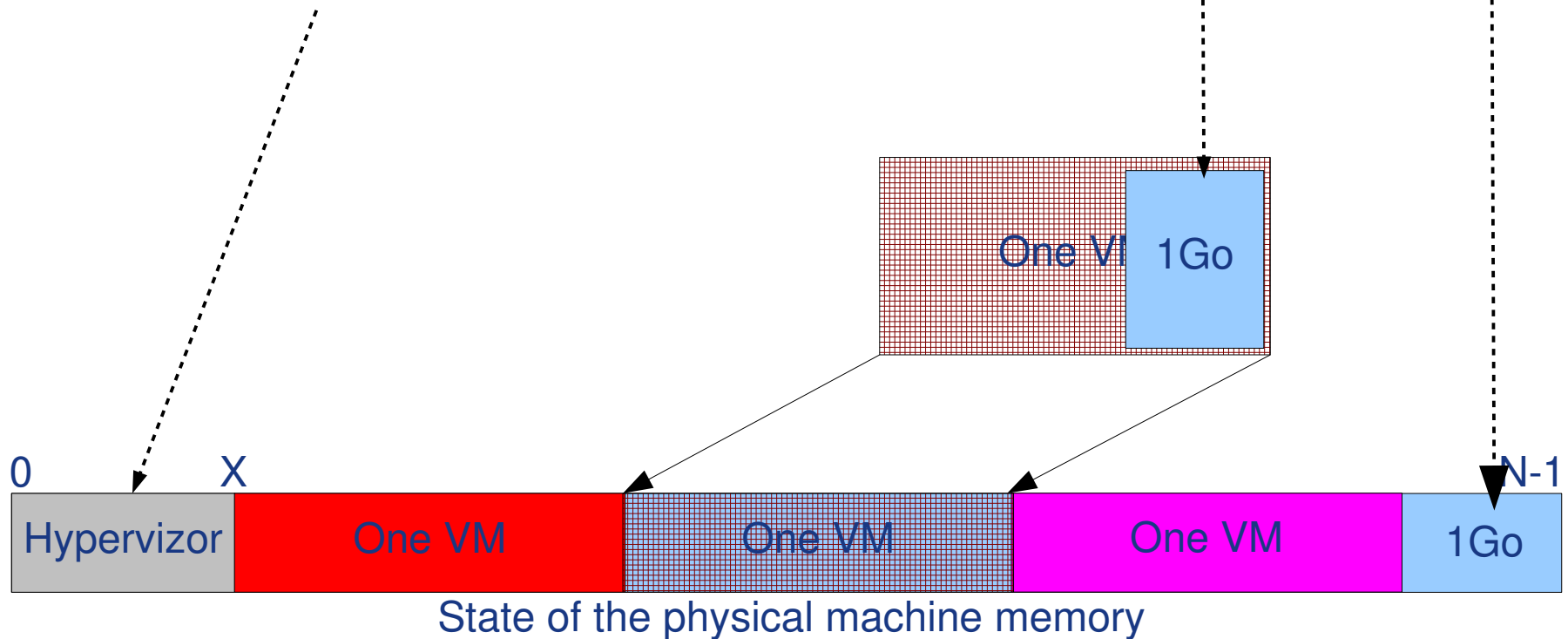
Consolidation

- Consolidation is a NP-Complete problem
- Hints
 - When to consolidate ?
 - Prediction
 - A learning phase to define a prediction model
 - Refinement of the model at runtime
 - Planification
 - Consolidation periods are known in advance
 - On the fly
 - Consolidation relies on runtime monitoring
 - How to consolidate ?
 - Heuristics. Ex: DRS/DPM from VMware
 - Constraints solver. Ex: Entropy

Optimization: ballooning

Problem

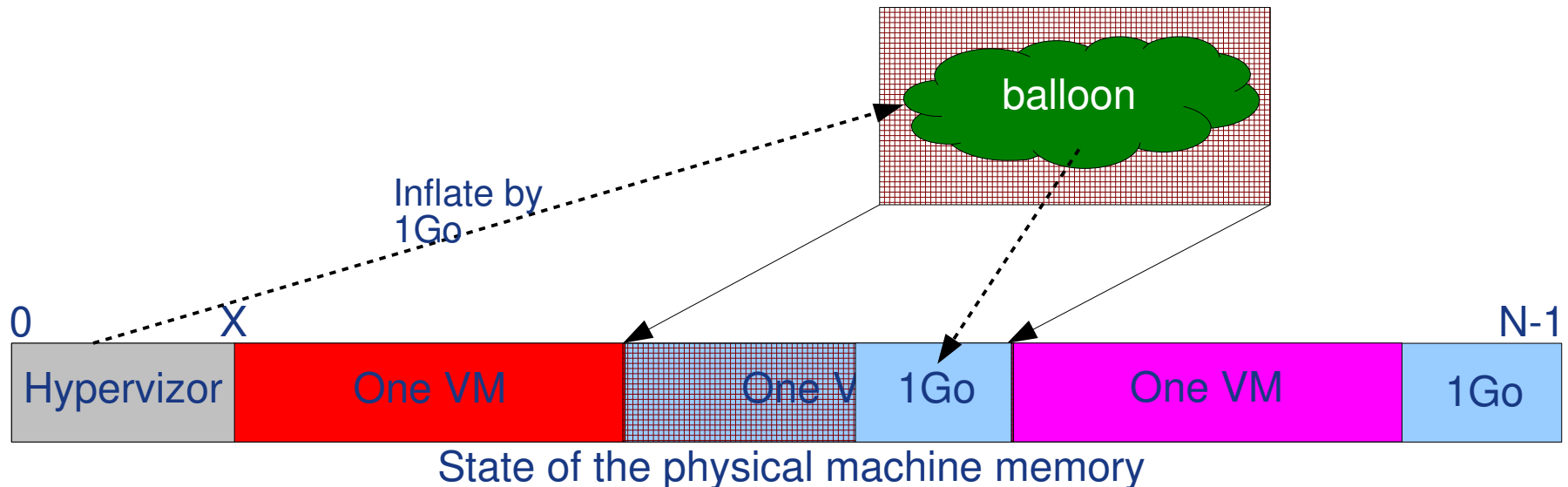
- A VM with 2Gb must be started. Cannot fit in the free space
- Another VM has free (or weakly used) memory
- How to reclaim and use this memory (without extensive VM modifications)



Optimization: ballooning

Solution

- A balloon driver is installed in each VM
- The hypervisor can inflate the balloon in order to reclaim memory
- It may force some pages to swap



Jobs in the cloud domain

- Build a cloud platform such as Amazon
 - Should win to the lottery :)
- Implement tools for cloud users
 - Deployment tools
 - Administration tools
- Be a reseller
 - PaaS or SaaS specialized for a specific domain
- Install (for clients) custom private cloud solutions
- Instructor for client companies
- Cloud application administrator for a company or for a cloud provider
- Work for virtualization systems providers
 - VMware, Citrix (Xen), Microsoft (hyper-V), etc.
- Researcher
 - Issues about energy consumption
 - Issues about QoS management
 - Security issues
 - Etc.