

# CloudMIP

GreenIT and DFS research in the  
Cloud at Toulouse.

<http://cloudmip.univ-tlse3.fr>

Pr. Jean-Marc Pierson

Dr. François Thiebolt

{pierson, thiebolt}@irit.fr



- The SEPIA team,
- The CloudMIP platform,
- GreenIT: power monitoring @ node-level,
- Future work.

The SEPIA team (IRIT: Pr Jean-Marc Pierson / N7: Pr Daniel Hagimont) mainly focuses on GreenIT, autonomic and distributed systems (cloud filesystems).

10 permanents (4 Pr, 6 MCF, 1 Dr-engineer)

2 engineers, 1 post-doc,  
2 associated researchers,  
16 PhD students.

CoolEmAll (FP7),  
SVC (Grand Emprunt),  
SOP and Control Green (ANR).

Toulouse platforms :

Manager : Pr Jean-Marc Pierson

- Grid5000-Toulouse (560),
- GridMIP (128),
- CloudMIP (256),
- RECS2.0 (72) / 1U.

Pau platform :

- PireCloud (128).

(number of cores)

Data Nov. 2013

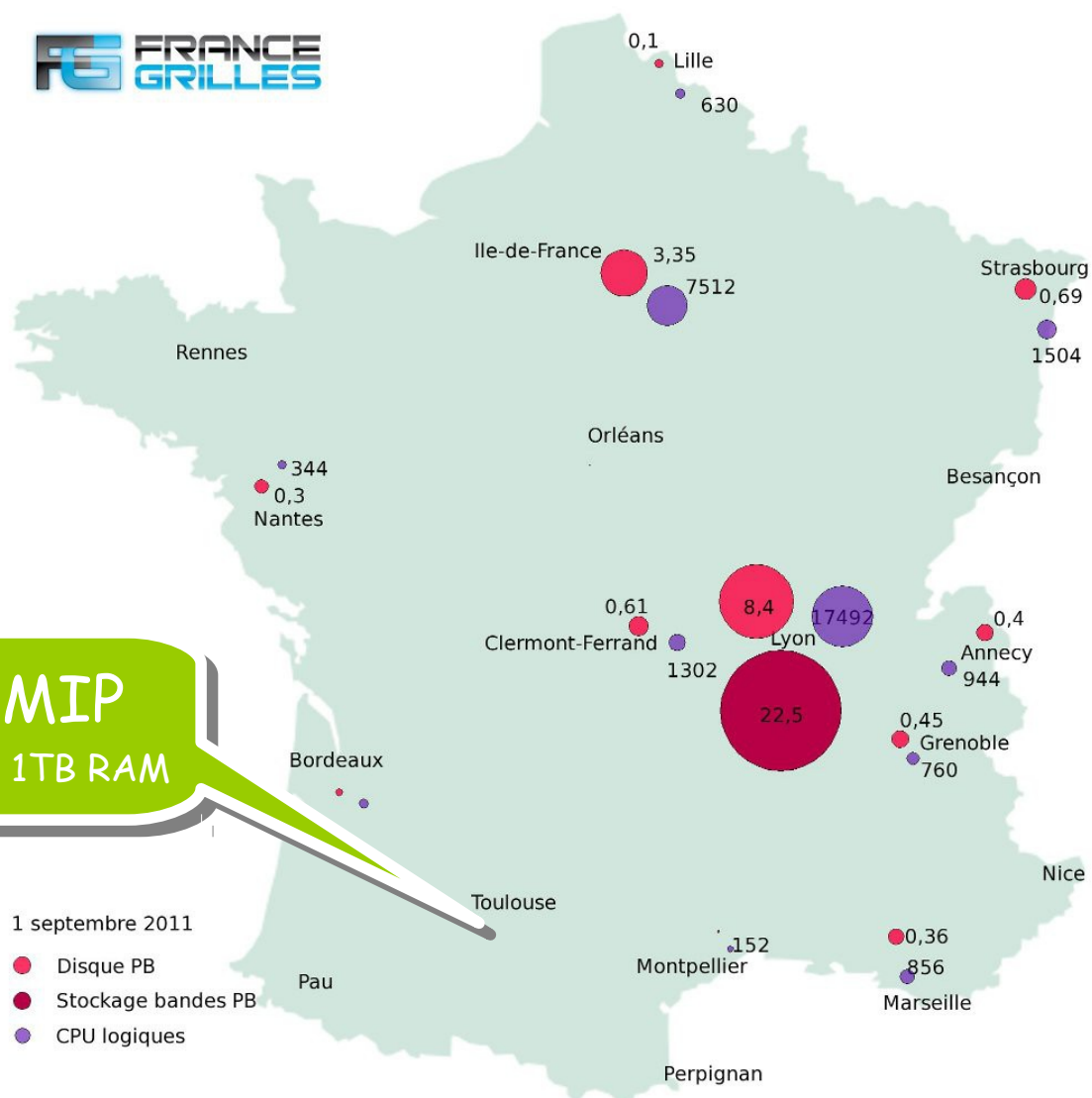
## Cloud initiative @ France-Grilles

● Experimental testbed built upon the following platforms :

**IRIT**, IPHC (soon), CC-IN2P3, LAL and Montpellier (?).



**CloudMIP**  
256 cores\*, 1TB RAM



1 septembre 2011

- Disque PB
- Stockage bandes PB
- CPU logiques

*\*physical cores. Up to 512 cores with hyper-threading.*

## Facts and resources

- Funded by France-Grilles, installed in 2012,
- Who : Pr Jean-Marc Pierson (manager), Dr François Thiebolt, DTSI Network team,
- Location : Toulouse 3 university's Data Center,
- Taskforce : 1 Dr-engineer (30%), 1 engineer (soon),
- Fluids consumption annual cost (est.) : between 4k€ and 10k€,
- Status: **pre-production**,

Dell M1000e 16 blades



## Hardware, system, middleware ...

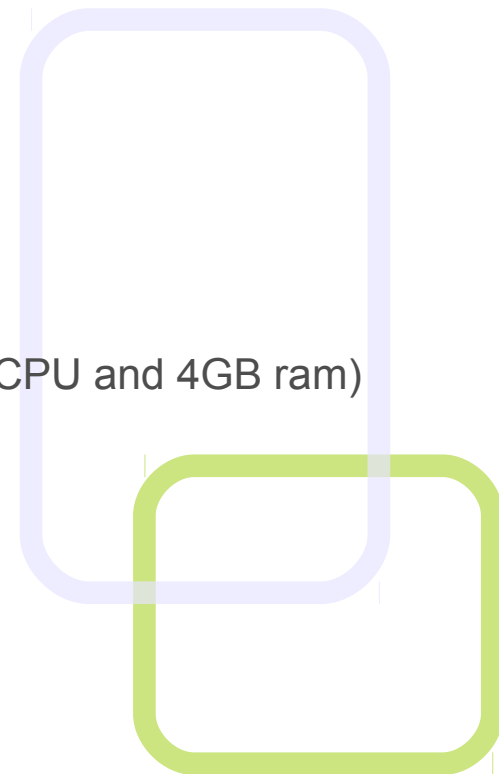
- 2 x Dell M1000e chassis each filled with 16 blades ➡ 256 **physical** cores, 1TB RAM, 15TB disk,
- System : Scientific Linux 6.4 x86\_64,
- VM management : **OpenNebula** 4.2 (spice protocol support) with **KVM** hypervisor,
- OpenNebula GUI **Sunstone**,
- **Zabbix** monitoring.

+ power monitoring      + seconds to launch a hundred of multi-Gb VMs      + open-source software



## ... additional details

- Wiki → <http://cloudmip.univ-tlse3.fr/>
- Monitoring\* → <http://cloudmip.univ-tlse3.fr/zabbix>
- GUI → <http://cloudmip.univ-tlse3.fr:11000>
- 32 blades (8 cores @ 2.4Ghz, 32GB ram, 2 x 146GB SAS 15ktpm RAID0),  
 ➡ means upto 256 Amazon EC2 M1 instances (1 **physical** dedicated CPU and 4GB ram)
- OpenNebula 4.2 (KVM) with Qcow2 delta images to speedup deployment,  
 ➡ a hundred of VMs in just a few seconds :)
- 1s resolution power monitoring of nodes,
- A 24TB, 700MB/s NFS server\*\* shared with Grid5000 and GridMIP,
- Ways for the users to gain access to their VMs from the Internet :
  - ssh, vpn, spice display forwarding, Sunstone GUI (**done**),
  - #1000 ports on the front-end node dedicated to routing (**done**),
  - #60 dedicated public IPs (**done**) with dynamic routing (**on way**).



\* login : **green**, passwd : **cloudmip**

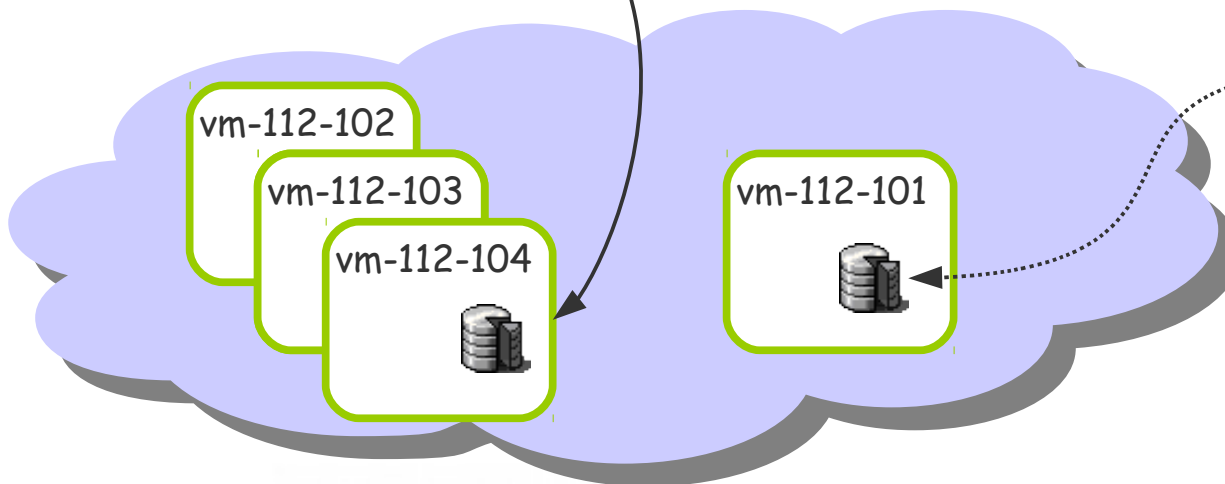
\*\* under heavy rebuilding due to numerous failures of hard drives [nov.13]

# The CloudMIP platform

► Pool of public IP to CloudMIP\*

195.220.53.1  
195.220.53.2  
|||||||  
195.220.53.57

\*subnet 195.220.53.0/26



nfs.cloudmip.univ-tlse3.fr



wn[1..32].cloudmip.univ-tlse3.fr



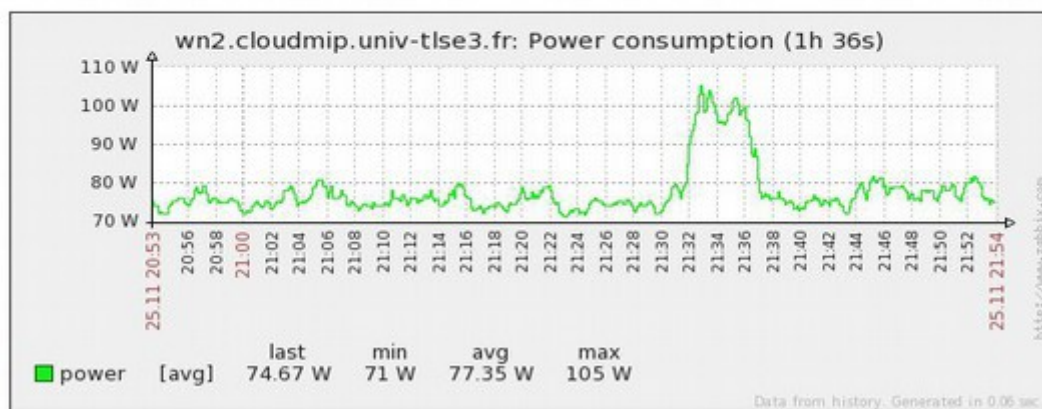
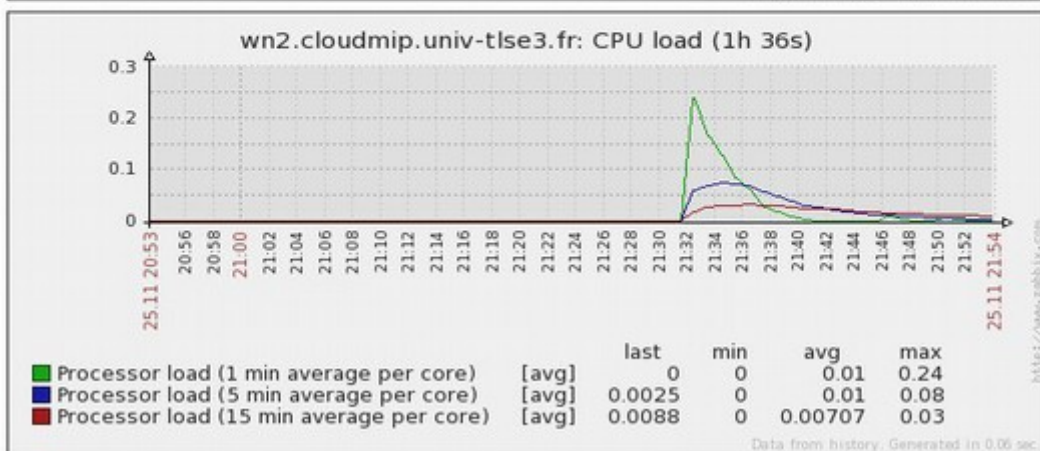
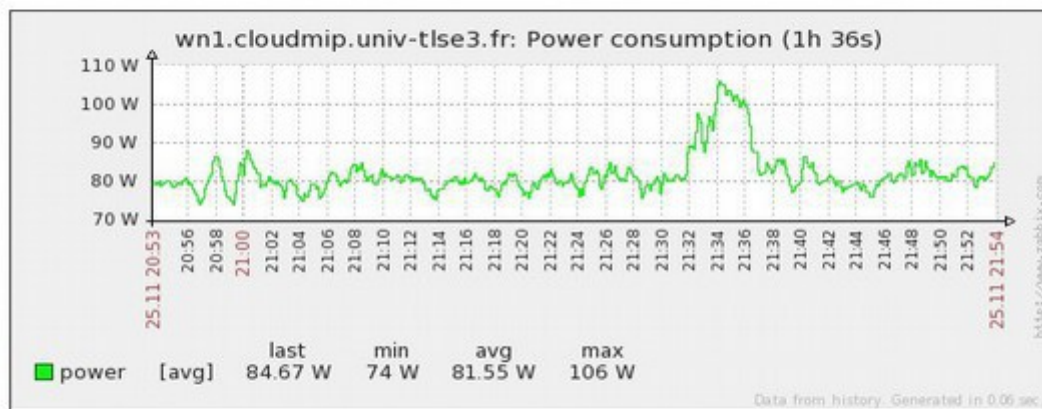
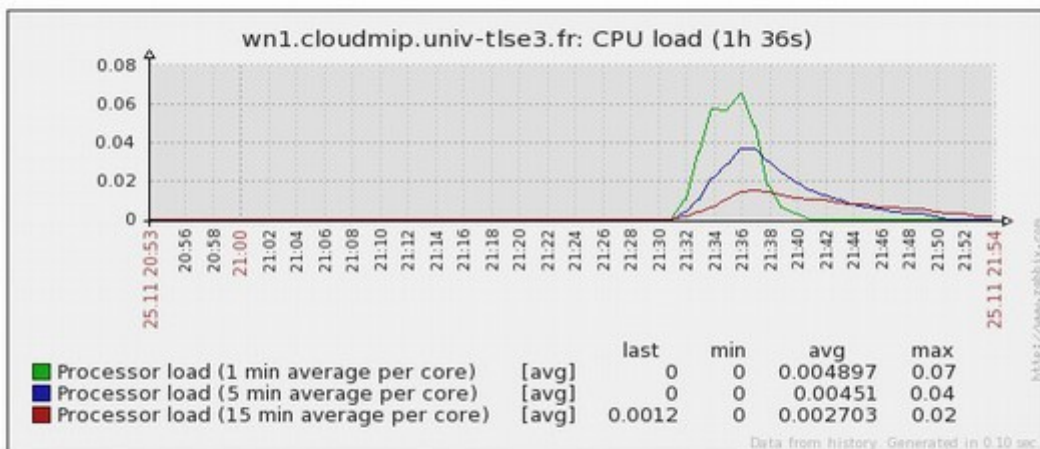
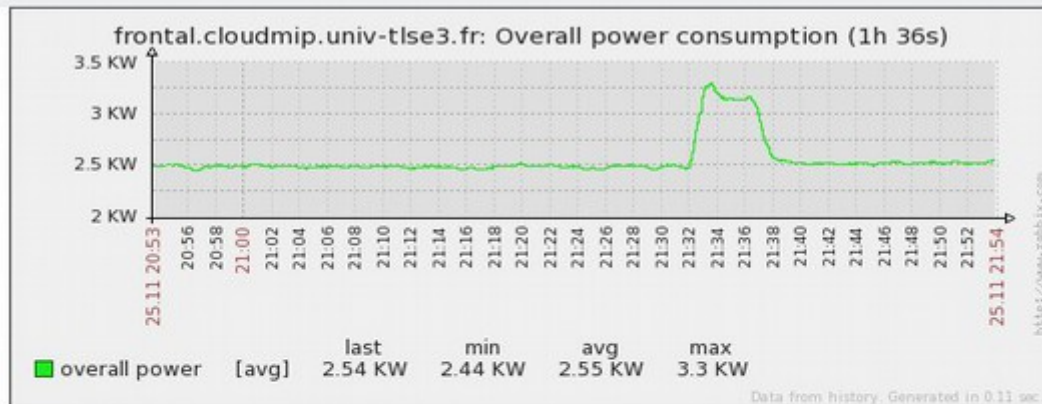
# GreenIT: power metering

## ▶ Launching 248 VMs

```
#> onetemplate instantiate SL64 -m 248
```

=> leads to eight VMs on each of the 31 nodes, thus each VM using one physical CPU.

=> Max. power consumption is 3.3kw



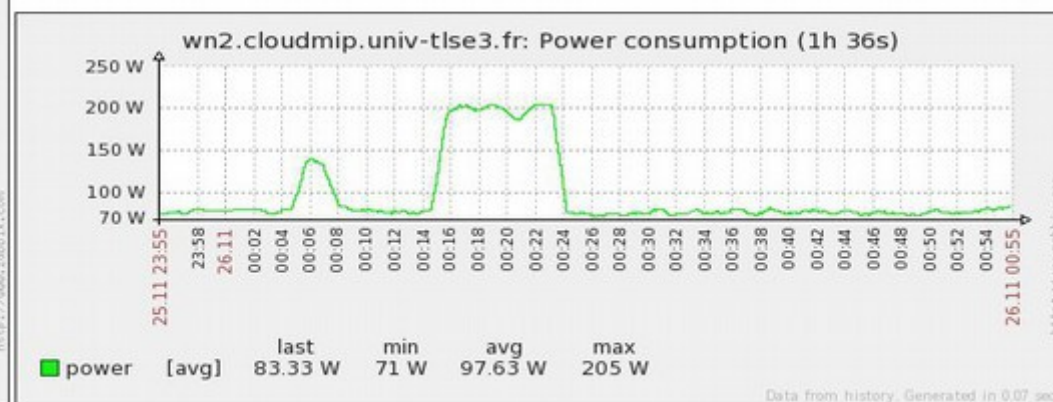
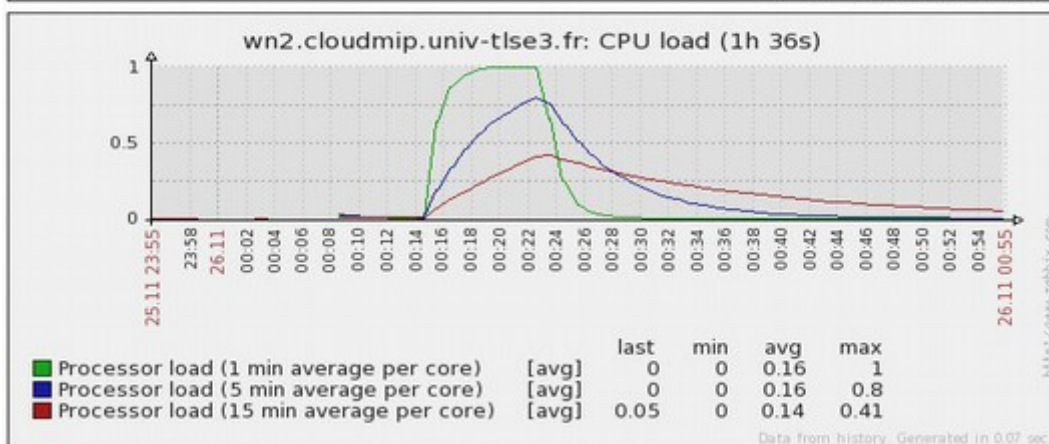
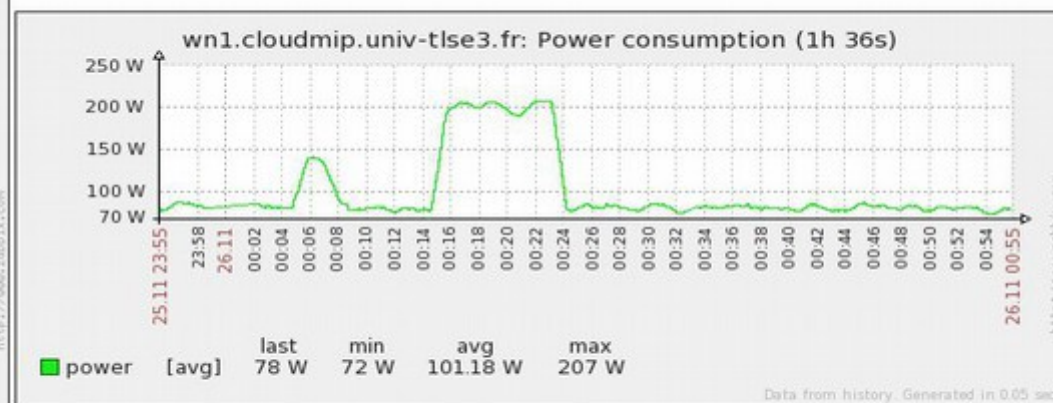
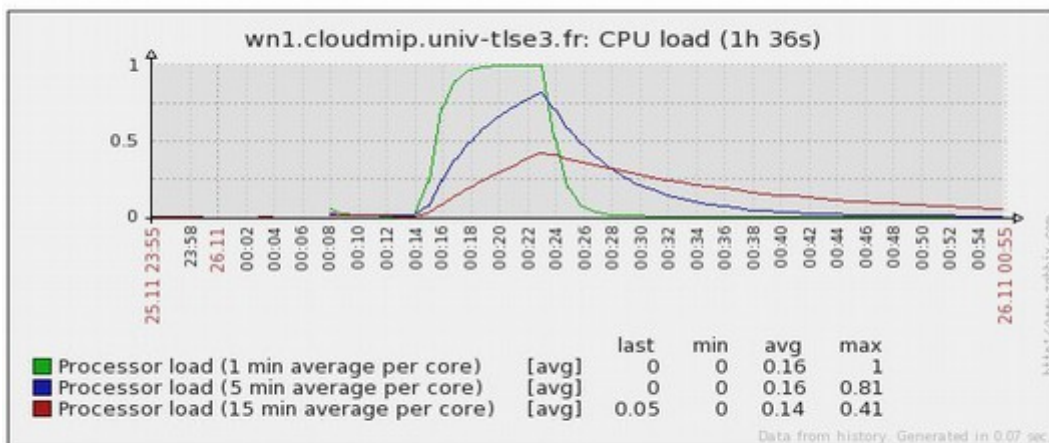
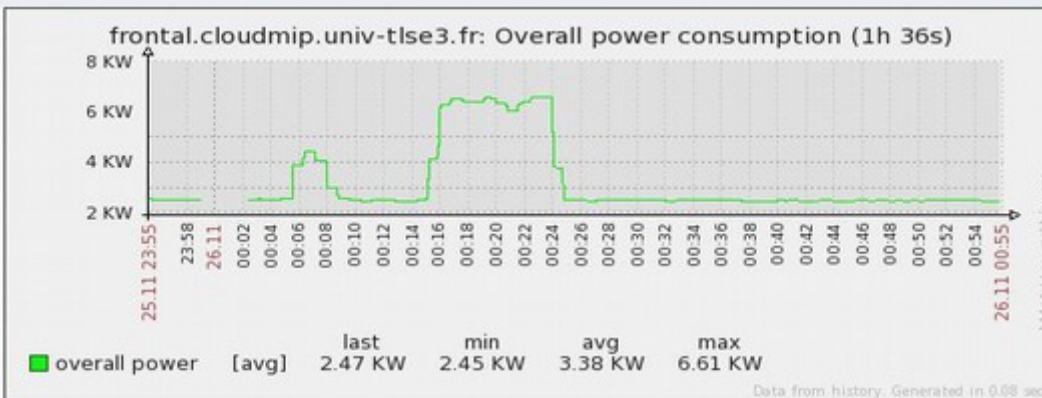


# GreenIT: power metering

## ▶ Launching stress test on all nodes

```
#> pdsh -w wn[1..32] -- openssl speed -multi 8
```

=> Peak power consumption is about 6.6kw



- The SEPIA team,
- The CloudMIP platform,
- GreenIT: power monitoring @ node-level,
- Future work:
  - Apache DeltaCloud CIMI + CloudInit,
  - GreenIT: power monitoring @ VM-level,
  - Offloading Hyperion's embarrassingly parallel code to CloudMIP.

## ► Keys to the Cloud interoperability:

- VM lifecycle management (instantiation, migration, shutdown ...),
- VM contextualisation (network setup, applications to start along with parameters ...).

### VM management@OpenNebula:

- XML-RPC / OpenNebula API,
- OCCI <http://occi-wg.org> or rOCCI (Ruby OCCI),
- Amazon EC2 (AWS).

Apache **DeltaCloud** with **CIMI** interface and OpenNebula driver.

### VM contextualisation@OpenNebula:

- Based on a CDROM image built at boot time and mounted as a cdrom within the VM filesystem.

The upcoming **OpenNebula 4.4** release has support for **Cloud Init**.

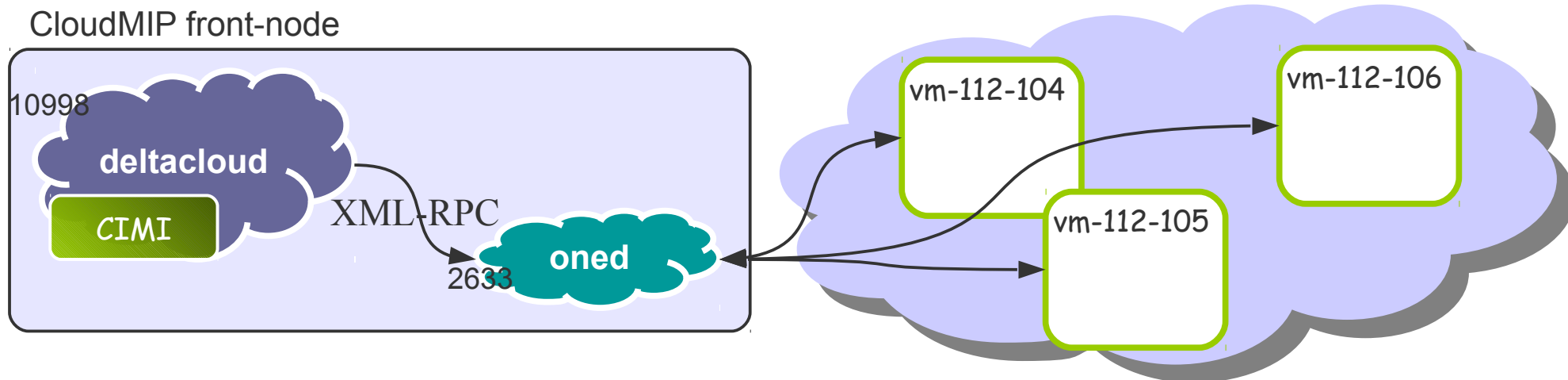
## ► Apache DeltaCloud + OpenNebula 4.4 (next release)

Apache **DeltaCloud** daemon abstracts Clouds API through its own API.

[**CIMI**] New DMTF standard for Cloud Infrastructure Management Interface is available via DeltaCloud!

Deltacloud setup @ CloudMIP is a work in progress.

CloudMIP migration to the upcoming **OpenNebula 4.4** to benefit from the Cloud Init VM contextualisation.





► Power consumption of processes\* : **ECTOP**

Tool to estimate the power consumed on each running process of the machine

Light weight\*\*, several sensors (PerfCounters, CPU%, Memory, CPU temperature, ...) and wattmeters (ACPI, G5K PDUs, CloudMIP, RECS, ...).

Two estimators implemented

Inverse model (PE\_IC): calibration with power meter

$$P^{PID} = \frac{P^{Node} \times CPU_{time}^{PID}}{CPU_{time}^{Node}}$$

Linear model (PE\_MMC2):

$$P^{PID} = \frac{(P_{max}^{Node} - P_{min}^{Node}) \times CPU_{time}^{PID}}{CPU_{time}^{Node}} + \frac{P_{min}^{Node}}{procs}$$

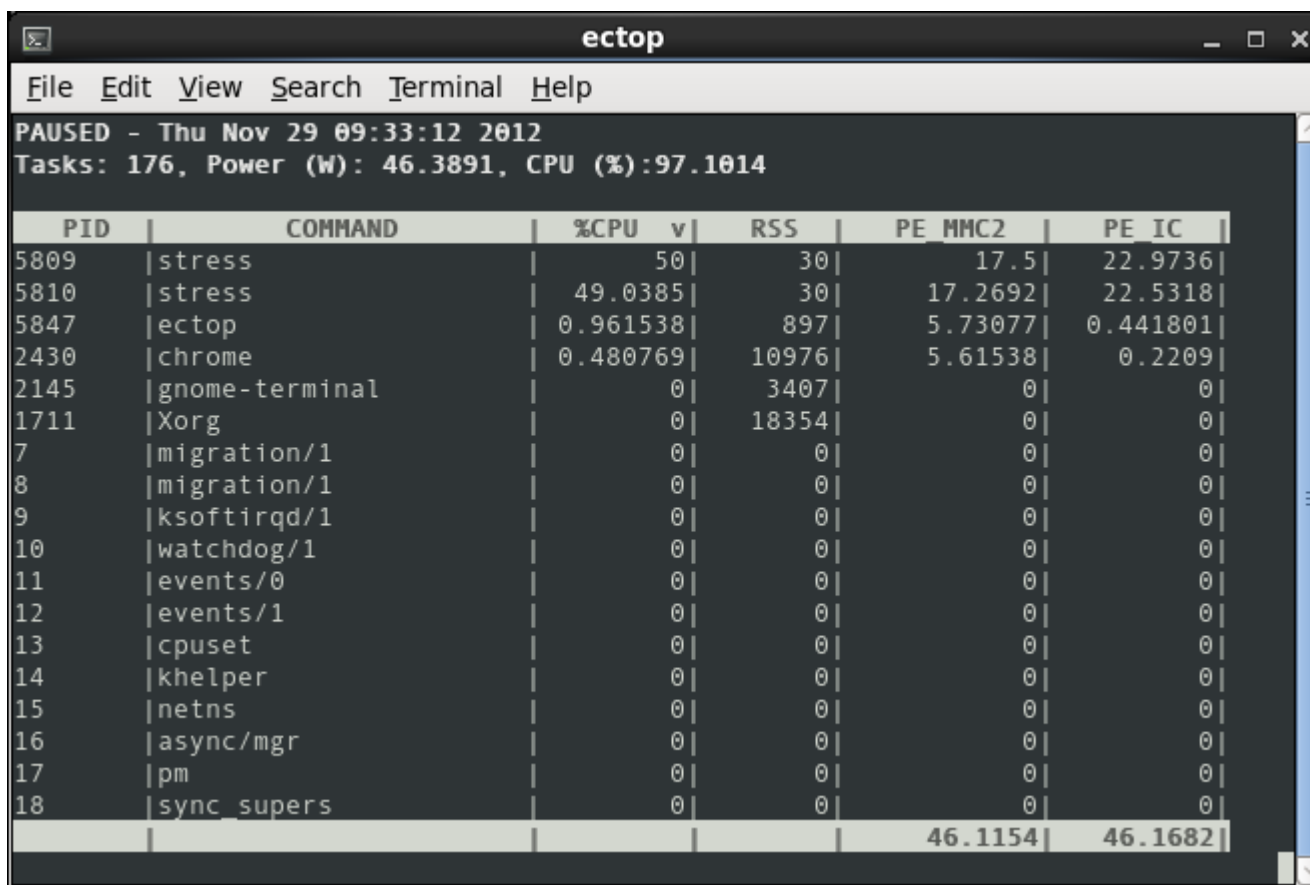
\*\* as low as Memory: 3Kb, CPU: 0.3%

**KVM hypervisor ==> each VM is a process thus leading to a possible evaluation of the power consumption of each VM!**

\*ongoing researches from Leandro Fontoura Cupertino email:fontoura@irit.fr

► Power consumption of processes (cont.)

ECTop (GPL license, available at <http://coolemall.eu>,  
alternate at <http://www.irit.fr/~Georges.Da-Costa/code.html>)



The screenshot shows the 'ectop' application window. The title bar is 'ectop'. The menu bar includes 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. The status bar at the top shows 'PAUSED - Thu Nov 29 09:33:12 2012' and 'Tasks: 176, Power (W): 46.3891, CPU (%):97.1014'. The main display area shows a table of process power consumption data.

PID	COMMAND	%CPU	v	RSS	PE	MMC2	PE	IC
5809	stress	50		30		17.5		22.9736
5810	stress	49.0385		30		17.2692		22.5318
5847	ectop	0.961538		897		5.73077		0.441801
2430	chrome	0.480769		10976		5.61538		0.2209
2145	gnome-terminal	0		3407		0		0
1711	Xorg	0		18354		0		0
7	migration/1	0		0		0		0
8	migration/1	0		0		0		0
9	ksoftirqd/1	0		0		0		0
10	watchdog/1	0		0		0		0
11	events/0	0		0		0		0
12	events/1	0		0		0		0
13	cpuset	0		0		0		0
14	khelper	0		0		0		0
15	netns	0		0		0		0
16	async/mgr	0		0		0		0
17	pm	0		0		0		0
18	sync_supers	0		0		0		0
						46.1154		46.1682

... and next to come is **Valgreen**, a high sampling rate (250ms), hardware independent statistical profiler!

## Near future ...

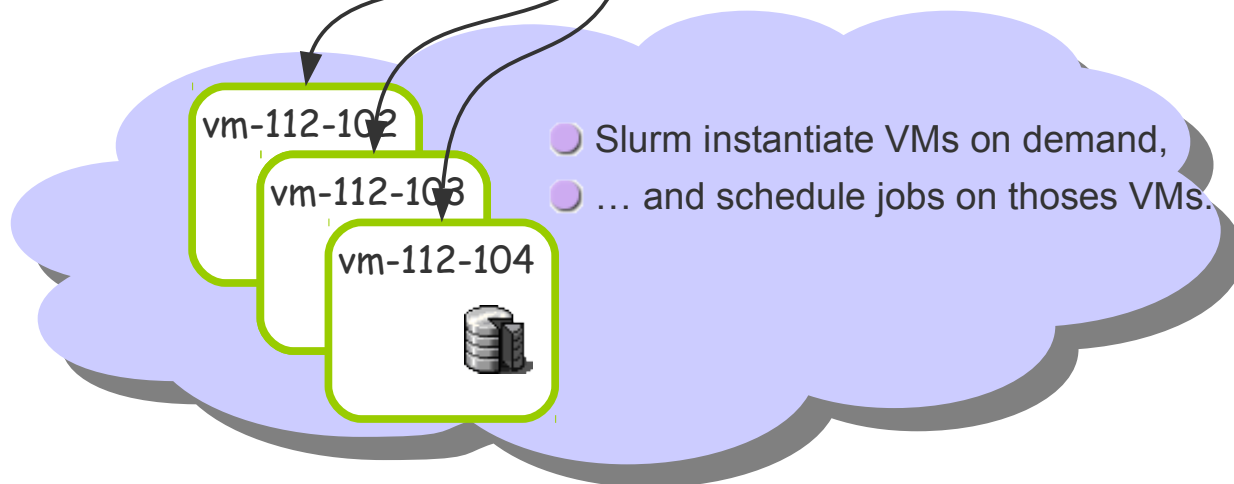
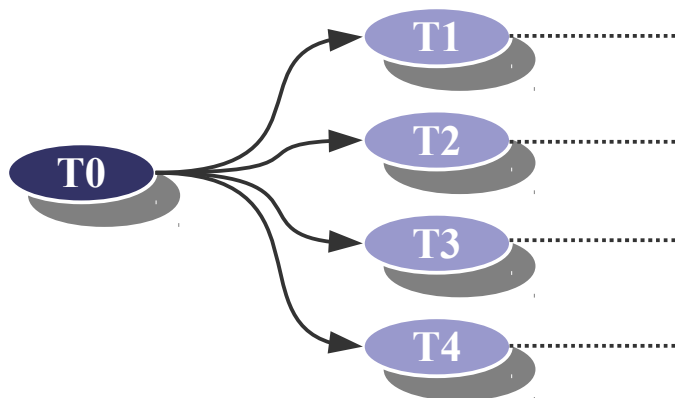
---

- Ectop : the per-process power consumption accounting (Leandro's work),
- Ectop calibration and integration in CloudMIP,
- Availability of a per-VM power metering through Zabbix,
- Integration of the per-VM power metering in Sunstone (OpenNebula GUI)  
→ enables a per-user view of VMs along with their power consumption,
- VM power metering data availability in an Inter-Cloud configuration ?

... a bit further ...

▶ Offloading embarrassingly parallel workload from the Hyperion supercomputer to CloudMIP.

● CALMIP supercomputer Hyperion





# Questions ?



Want to collaborate with France-Grilles ? <http://www.france-grilles.fr/>  
*To benefit from our expertise, to access powerful platforms and to share yours in innovative academic / business workflows!*