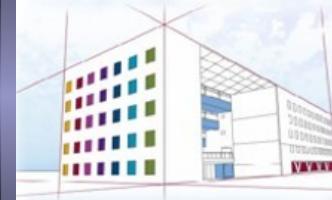




Université
Paul Sabatier
TOULOUSE III



IRIT
Institut de Recherche en Informatique de Toulouse
Red Blue Yellow Cyan Magenta Green

CloudMIP

GreenIT and DFS research in the
Cloud at Toulouse.



2013-06

Plan

- France-Grilles CloudMIP,
- The SEPIA team,
- GreenIT (**addon** : extreme density data center with RECS*),
- Distributed Filesystems,
- Experiments timespan,
- Future work.

*Christmann GmbH <http://www.christmann.info>

GIS France-Grilles

- Grid **production** sites (e.g. EGEE tier1),
- Cloud **researches and experiments** platforms (Toulouse, Lyon ...).



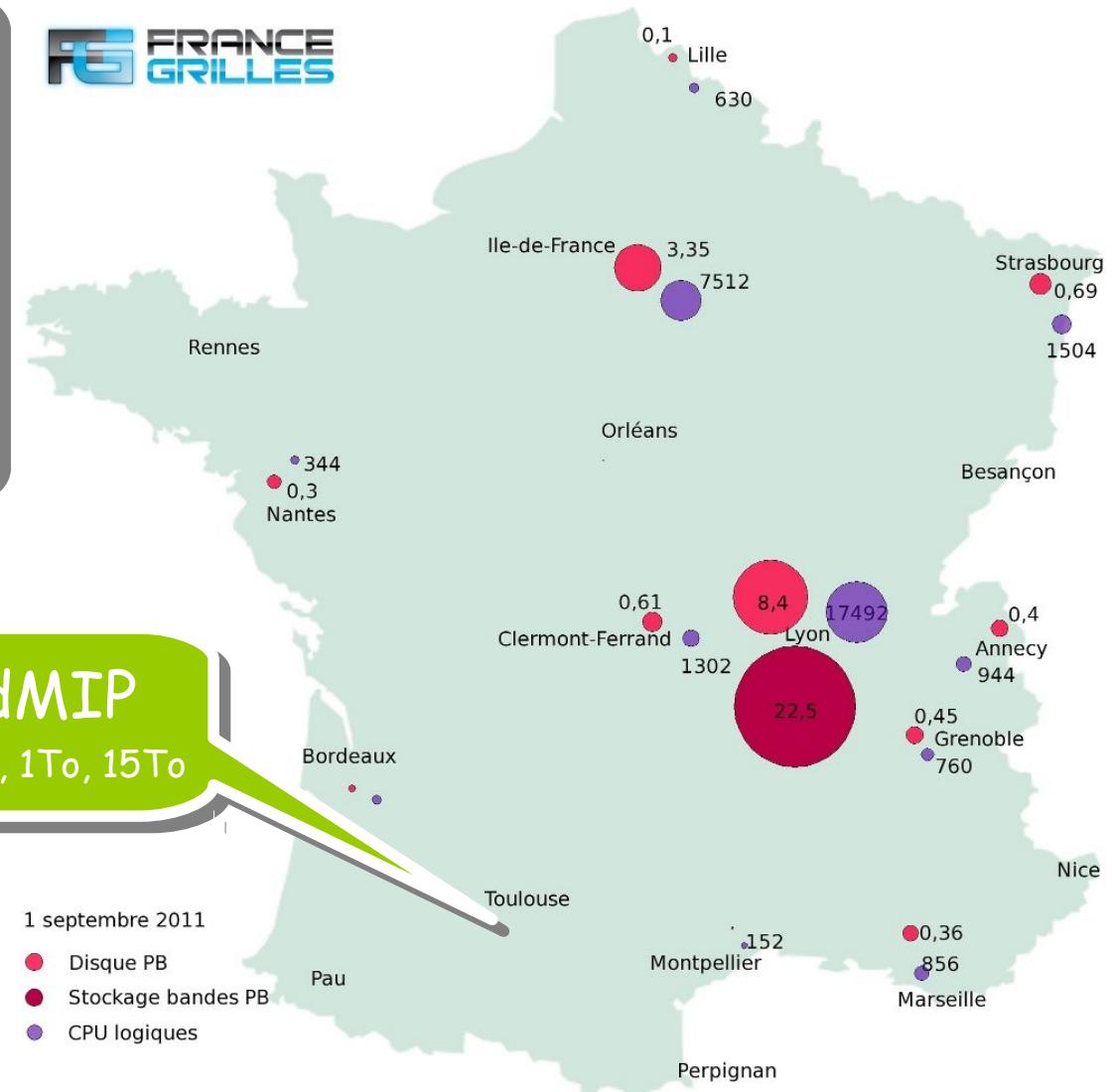
CloudMIP
256 cores*, 1To, 15To

1 septembre 2011

- Disque PB
- Stockage bandes PB
- CPU logiques

*512 cores with hyperthreading activated.

**FRANCE
GRILLES**



The SEPIA team

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

10 permanents (4 Pr, 5 MCF, 1 Dr-engineer)

1 engineer (CoolEmAll project),
21 PhD students.

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

Toulouse platforms :

Director : Pr Jean-Marc Pierson

- Grid5000-Toulouse (560),
- GridMIP (128),
- CloudMIP (256),
- Christmann 18nodes 1U board.

Pau platform :

- PireCloud (128).

The CloudMIP platform

Facts and resources

- Who : Pr Jean-Marc Pierson (manager), Dr François Thiebolt (system), DTSI Network team,
- Where : hosted at the Paul Sabatier University's Data Center (along with Grid5000-Toulouse and GridMIP platforms),
- Taskforce : 1 Dr-enginner (40%), 1 software engineer (30%),
- Fluids consumption annual cost (est.) : between 4k€ and 10k€.

Dell M1000e 16 blades



Hardware, system, middleware ...

- 2 x Dell M1000e chassis each filled with 16 blades ➔ 256 **physical** cores, 1TB RAM,
- System : Scientific Linux 6.3 x86_64,
- VM provisioning system : **OpenNebula** 3.8 (spice protocol support) with **KVM** hypervisor,
- Monitoring : Zabbix (<http://www.zabbix.com>) ---combine Nagios and Ganglia capabilities.

+ power monitoring

+ seconds to launch a several Gb VMs

+ everything is open-source software

The CloudMIP platform

... additional details

- The front-end node : cloudmip.univ-tlse3.fr,
- The monitoring system : [cloudmip.univ-tlse3.fr/zabbix*](http://cloudmip.univ-tlse3.fr/zabbix),
- 32 blades (8 cores @ 2.4Ghz, 32GB ram, 2 x 146GB SAS 15ktpm RAID0),
 - ➡ means upto 256 VMs featuring 1 **physical** (not HT) CPU and 4GB ram)
- OpenNebula 3.8 (KVM) with Qcow2 delta images to speedup deployment,
 - ➡ a hundred of VMs in just seconds :)
- 1s resolution power monitoring of each node (sent to Zabbix monitoring),
- 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, display redirection (spice),
 - ▶ #1000 ports on the front-end node dedicated to routing (**tests**)**,
 - ▶ #60 dedicated public IPs with dynamic routing (**on way**).

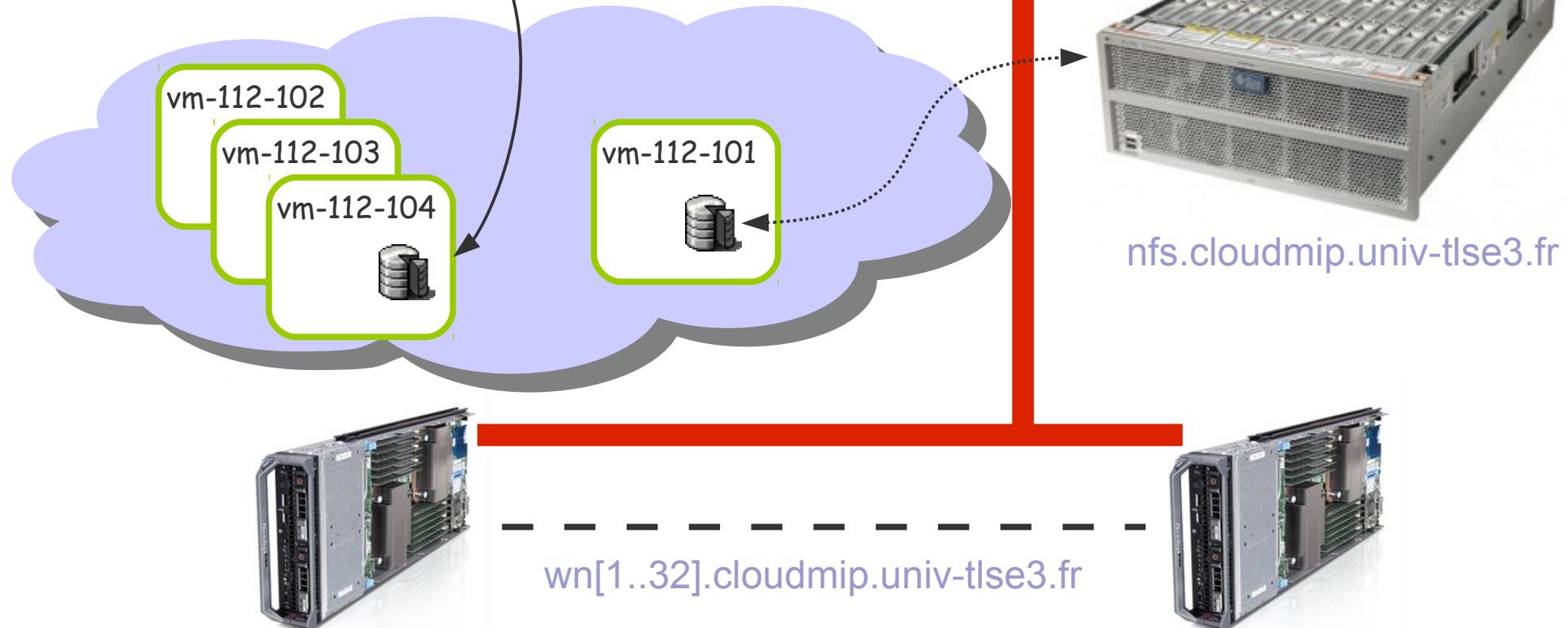
* login : **green**, passwd : **cloudmip** **#> *spicy -h cloudmip.univ-tlse3.fr -p 10000*

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



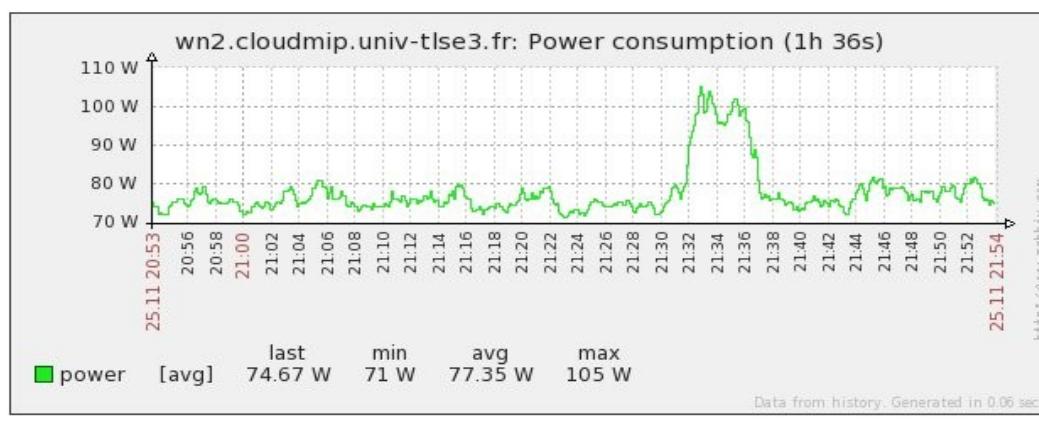
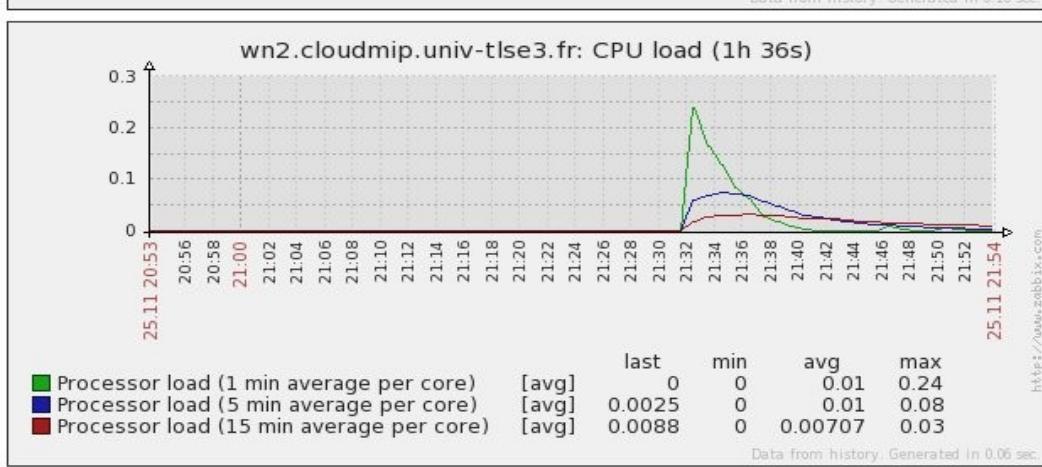
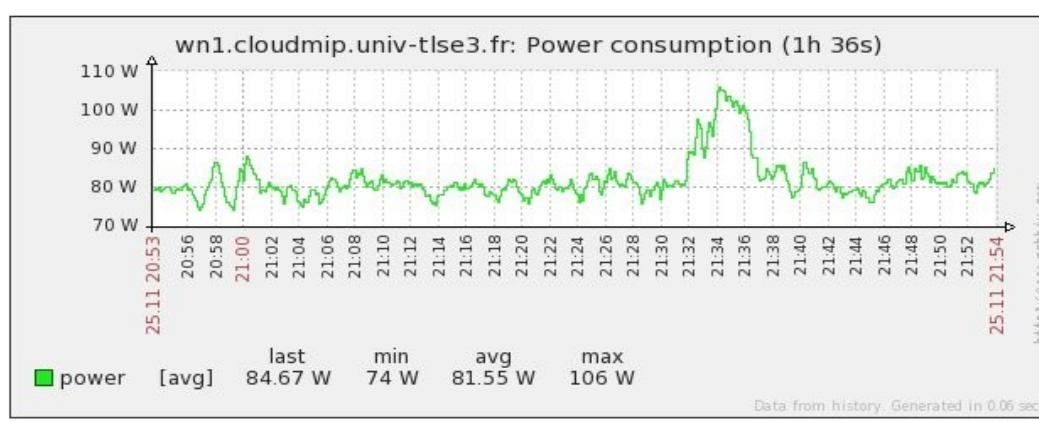
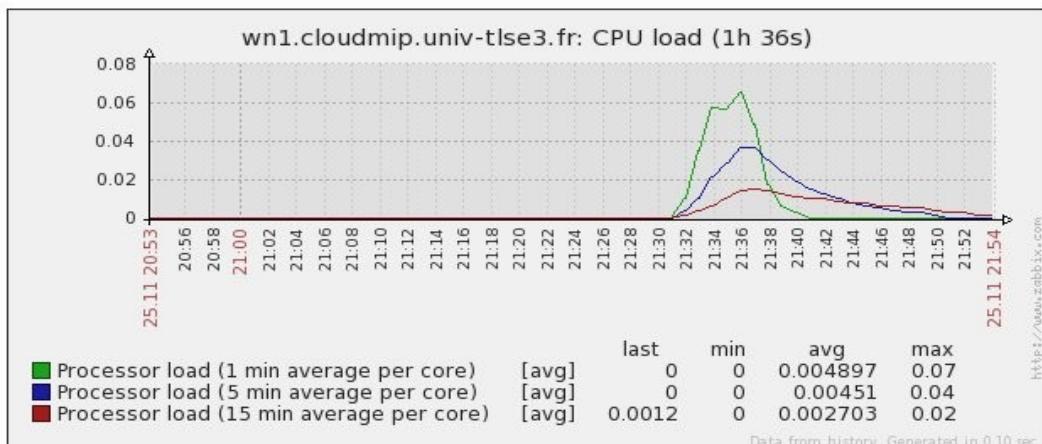
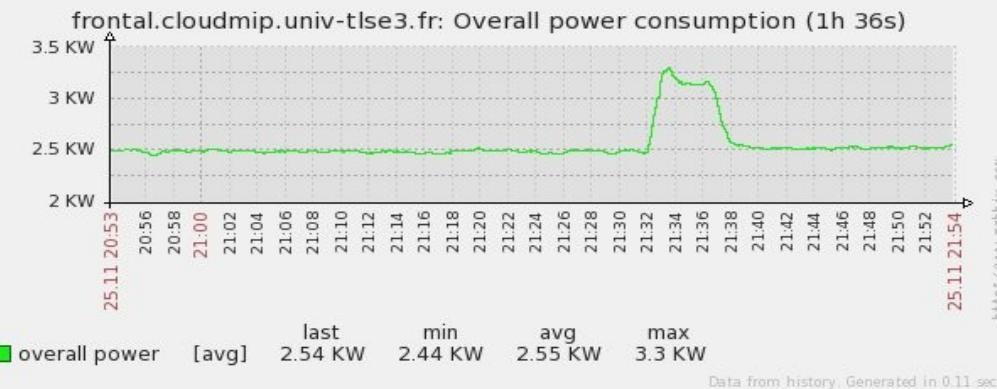
GreenIT : power metering

► Launching 248 VMs

```
#> onetemplate instantiate SL63 -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

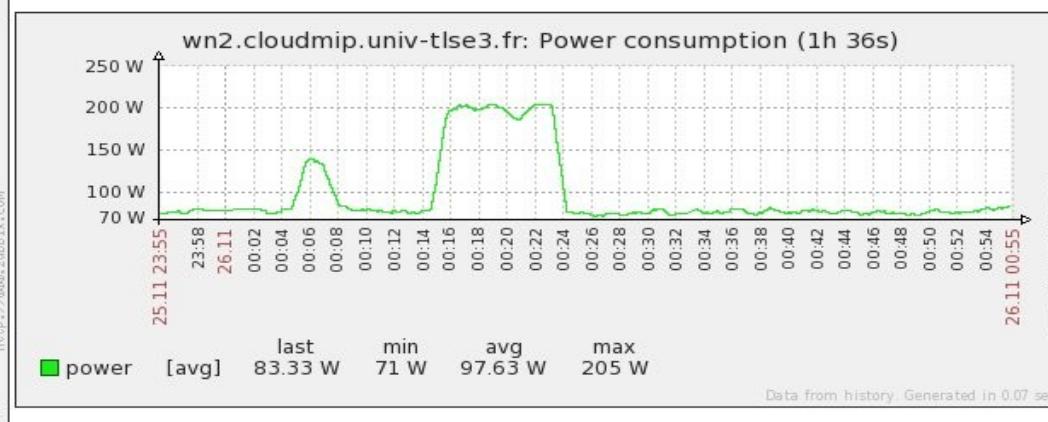
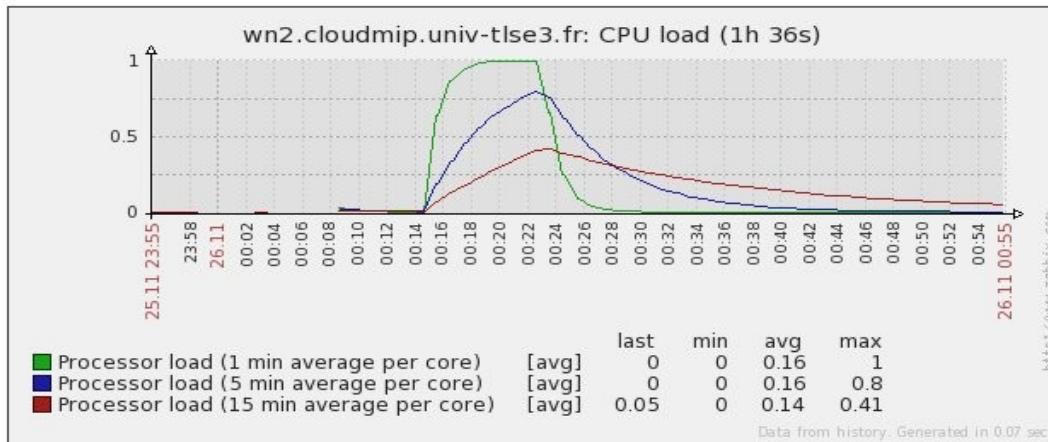
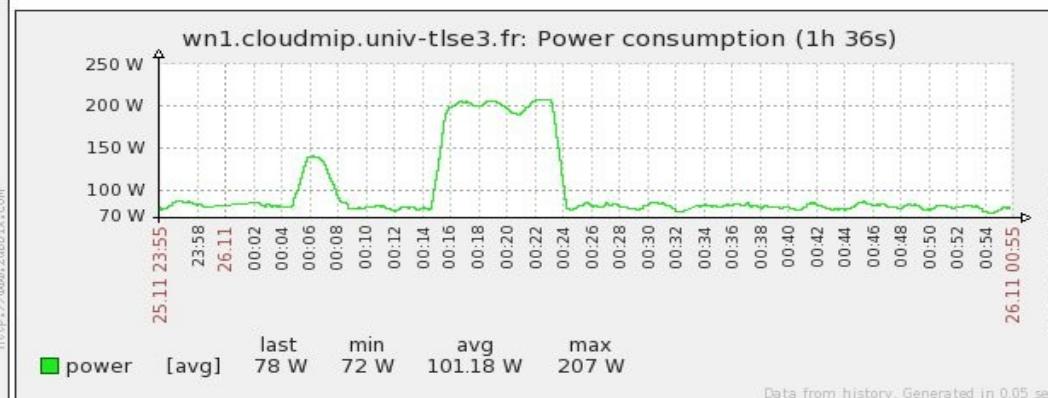
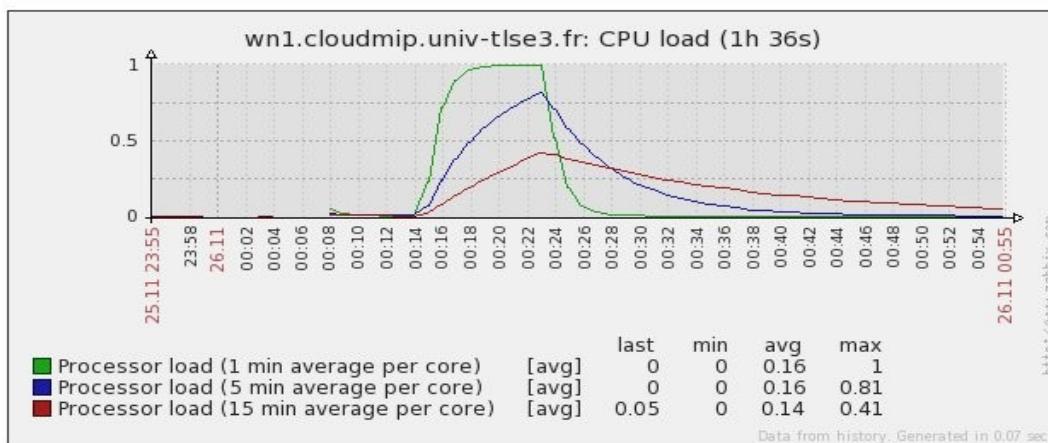
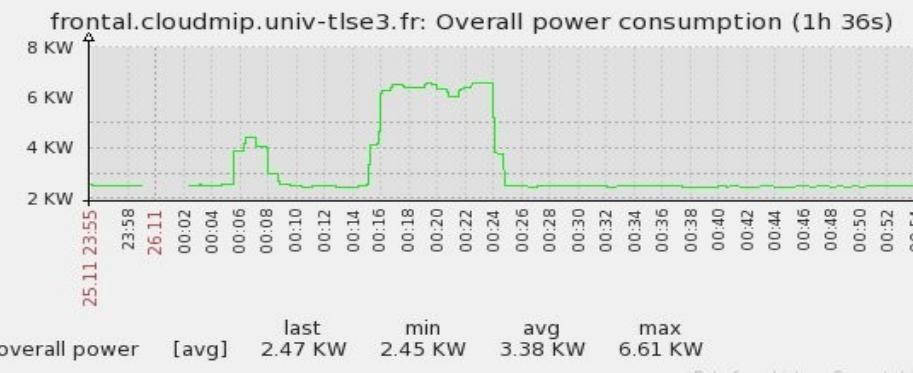


GreenIT : power metering

► Launching stress test on all nodes

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

==> Peak power consumption is about 6.6kw



GreenIT : RECS platform

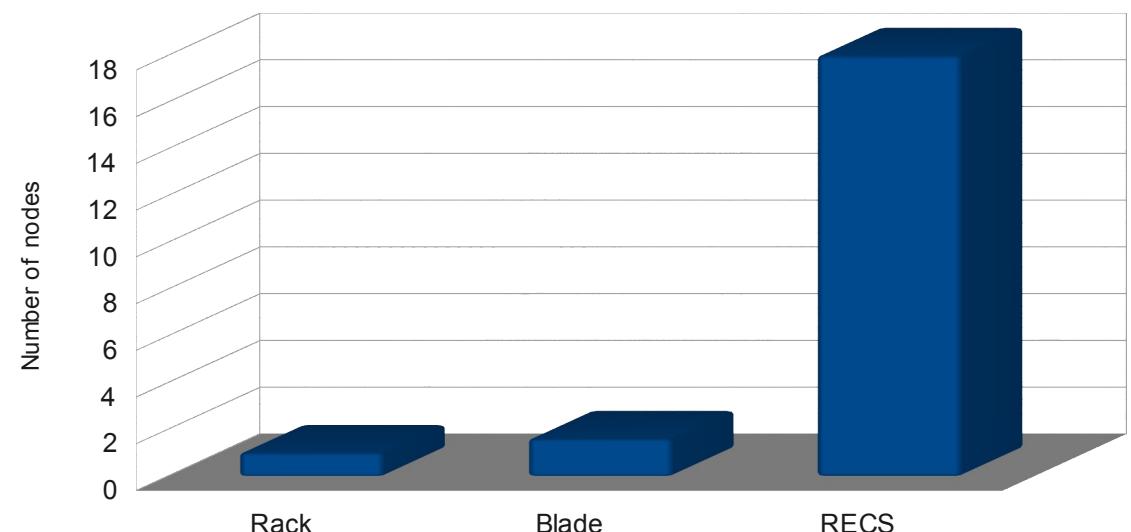
RECS : extreme density servers from Christmann GmbH*

RECS enables up to 18 nodes* (Atom, i7, ARM ...) to fit in a standard 1U rack !

*COM-Express boards



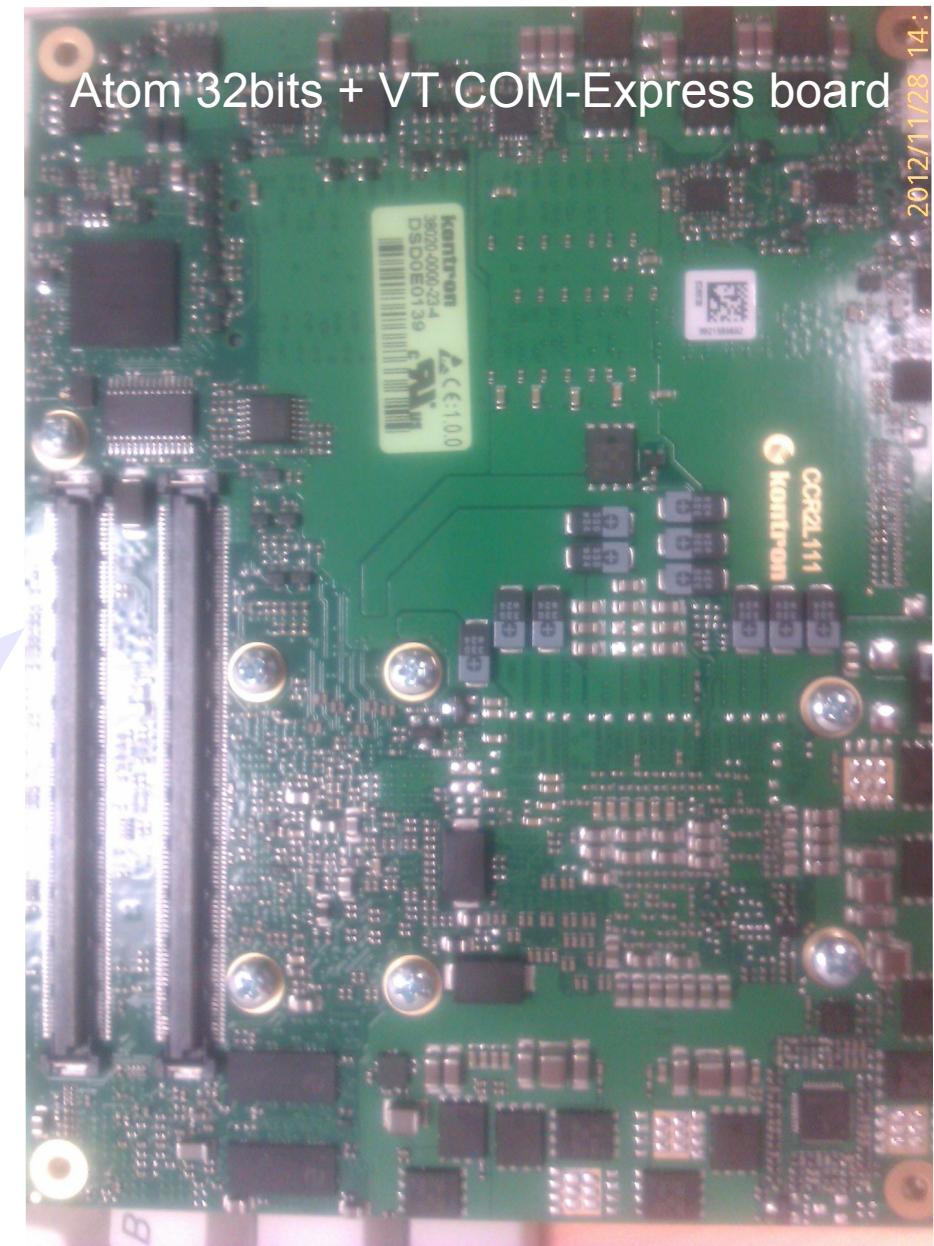
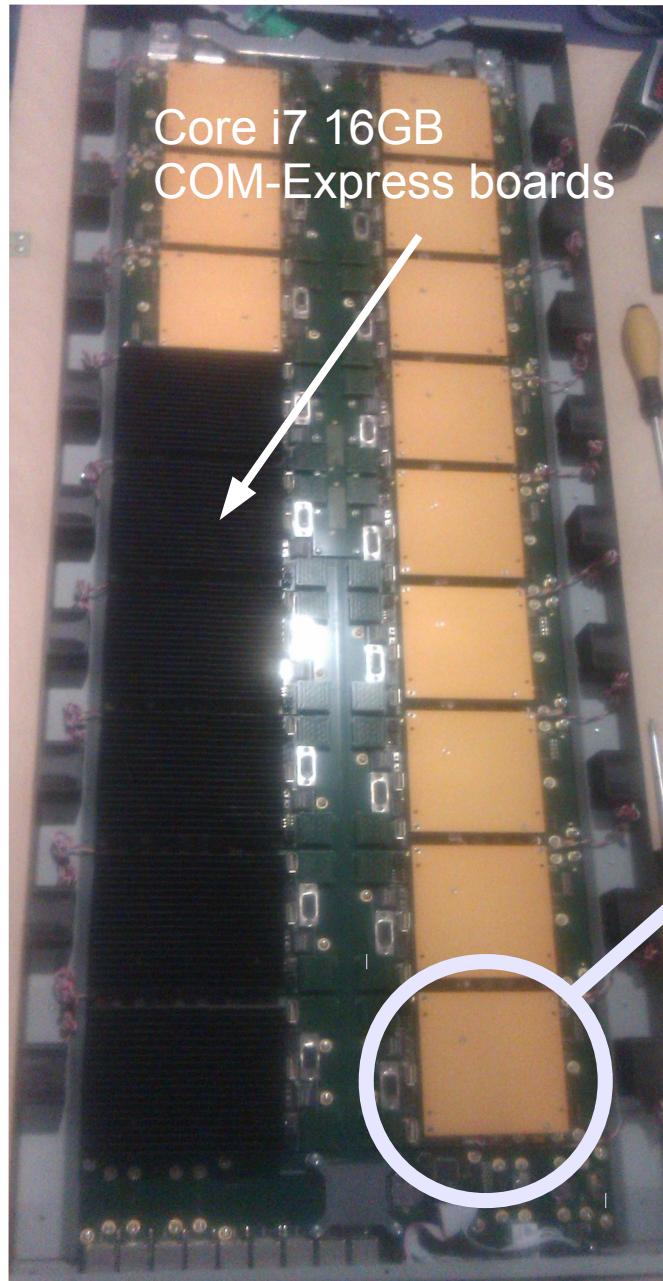
Server density : node(s) per 1U



*<http://www.christmann.info/>

Member of the CoolEmAll project (FP7)

GreenIT : RECS platform



2012/11/28 14:

GreenIT : RECS platform

... external 12v power supply connection detail



GreenIT : RECS platform

RECS Interface - Mozilla Firefox (on priam.irit.fr)

File Edit View History Bookmarks Tools Help

RECS Interface +

◀ 192.168.0.250/interface.html ★ Google 🔍 🏠

Connected ... 192.168.0.250 Christmann RECS | Box Server v2.0 (Sirius)

Installed Boards: 18

Turn On/Off Head Turn On/Off Slav...
Reset Headnode Reset Slaves

Power 14 W	Power 15 W	Power 13 W	Power 14 W	Power 12 W	Power 12 W	Power 10 W	Power 11 W	Power 11 W
Board Temp. 28 °C	Board Temp. 27 °C	Board Temp. 27 °C	Board Temp. 27 °C	Board Temp. 27 °C	Board Temp. 28 °C	Board Temp. 29 °C	Board Temp. 28 °C	Board Temp. 28 °C
On / Off	On / Off	On / Off	On / Off	On / Off	On / Off	On / Off	On / Off	On / Off
Reset								
Node 1	Node 2	Node 3	Node 4	Node 5	Node 6	Node 7	Node 8	Node 9 (Head)
Node 10 Power 7 W	Node 11 Power 6 W	Node 12 Power 10 W	Node 13 Power 13 W	Node 14 Power 17 W	Node 15 Power 13 W	Node 16 Power 14 W	Node 17 Power 14 W	Node 18 Power 14 W
Board Temp. 25 °C	Board Temp. 26 °C	Board Temp. 26 °C	Board Temp. 25 °C	Board Temp. 26 °C	Board Temp. 26 °C	Board Temp. 26 °C	Board Temp. 26 °C	Board Temp. 25 °C
On / Off	On / Off	On / Off	On / Off	On / Off	On / Off	On / Off	On / Off	On / Off
Reset								

12.34 Volts

17.82 Amperes

220 Watts

GreenIT : what's next ?

► Power consumption of processes*

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

Several sensors: PerfCounters, CPU%, Memory, CPU temperature, etc.

Two estimators implemented

Inverse model (PE_IC): needs a power meter

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

Linear model (PE_MMC2):

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

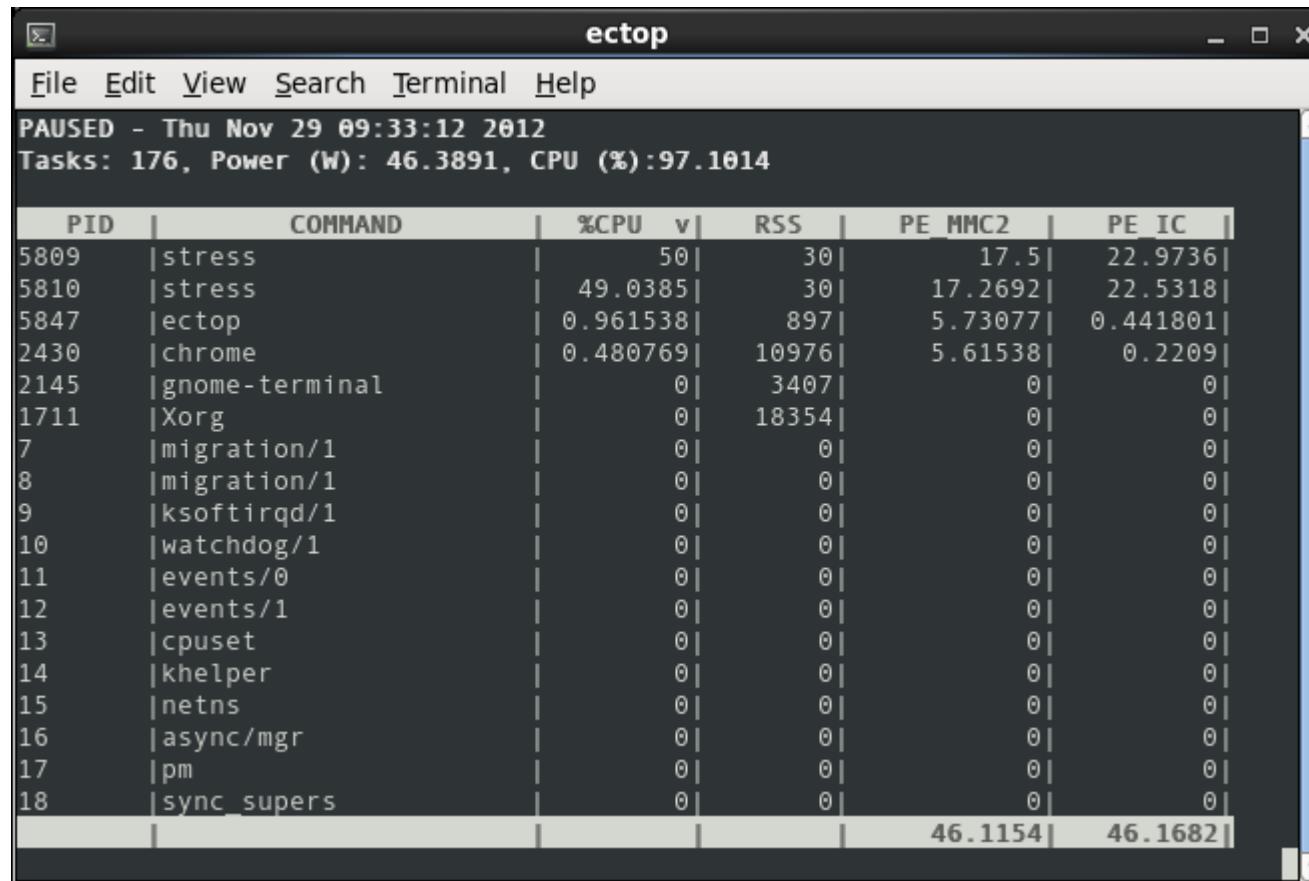
KVM hypervisor ==> each VM is a process thus leading to a possible evaluation of the power consumption of each VM!

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

GreenIT : what's next ?

► Power consumption of processes (cont.)

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



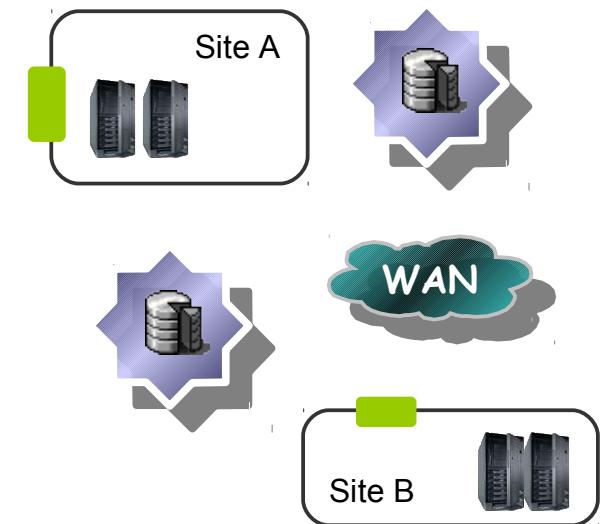
The screenshot shows a terminal window titled "ectop" running on a Linux system. The window title bar includes standard menu options: File, Edit, View, Search, Terminal, and Help. Below the title bar, the status line displays "PAUSED - Thu Nov 29 09:33:12 2012" and "Tasks: 176, Power (W): 46.3891, CPU (%):97.1014". The main area of the window is a table with the following 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	

The Secure Virtual Cloud (SVC) project

- Who : iTrust* company (leader), Bull, IRIT/SEPIA** ...,
- Budget : 14M€ / 3years (grand emprunt).

SVC project has a strong emphasis on security to both VMs and data in the Cloud.



The SVC project focuses on the following points :

- Securing VMs ➔ development of **iKare v2** software,
- Secure distributed filesystem,
- Energy efficiency.

*<http://www.itrust.fr>

**{mzoughi,jorda}@irit.fr

Experiments timespan

Today

Hardening the platform with local users workflows

Opening the platform to industrial collaborations

Scheduling of single and multi-thread applications (network simulator, energy-aware Content Delivery Network simulator ...) over 256 VMs (Scientific Linux 6.3 x64 customised image)

VPN, OpenNebula deployment

- Energy Monitoring
- Auto-routing of public IPs to virtual machines,

Cloud and Inter-Cloud Distributed File System experiments,
Power metering of running VMs, ...

PireCloud platform

- Hadoop@PireCloud
Boosting processing of in flight data (2Moro company / OMNIDATA project)

- DVFS to reduce energy consumption in MPI synchronisation loop

CoolEmAll project (FP7)
Energy-efficiency in Data Center
<http://www.coolemall.eu/>

iTrust Secured Virtual Cloud (SVC) project
Securing both data and VMs in the cloud (14M€ / 3y ---grand emprunt)

Future work

Near future ...

- Deployment of an object oriented distributed file system whose storage will be based on nodes' unused storage space,
- Much more fine-grained power consumption monitoring : a per-core approach,
- Deployment of the per-process power consumption accounting (Leandro's work),
- VMDirac plug-ins integration to OpenNebula,
- Enabling the PireCloud site from Pau as a CloudMIP clone,
- ...

... a bit further

- PireCloud  CloudMIP tighter integration : span of the distributed FS across both sites, proxying Zabbix monitoring to CloudMIP,
- RECS2 and RECS3 integrated made available for energy efficiencies researches,
- Inter-Cloud VM migration scheduler according to pluggable policies,
- ...

Questions ?

Cloud
MIP