

# Chameleon, CloudLab, Grid'5000: What will the ultimate testbed look like?

Lucas Nussbaum  
lucas.nussbaum@loria.fr



# Grid'5000, Chameleon, CloudLab

## ▶ **Grid'5000**

- ◆ Established testbed (9 sites, 1000 machines, 500+ users/y since '05)
- ◆ <https://www.grid5000.fr/>

## ▶ Two recent projects (2014 ~ 2017; NSF funding for 10 M\$ each)

### ◆ **Chameleon**

- ★ <https://www.chameleoncloud.org/>
- ★ Codebase: CHI (CHameleon Infrastructure)  
CHI = OpenStack + Grid'5000 tools + custom developments
- ★ Currently two sites, 317 nodes (+ KVM Cloud)

### ◆ **CloudLab**

- ★ <http://www.cloudlab.us>
- ★ Emulab codebase as a foundation
- ★ Two three sites, 315 (arm64) + 100 + 100 (x86) nodes

# Grid'5000, Chameleon, CloudLab

## ▶ **Grid'5000**

- ◆ Established testbed (9 sites, 1000 machines, 500+ users/y since '05)
- ◆ <https://www.grid5000.fr/>

## ▶ Two recent projects (2014 ~ 2017; NSF funding for 10 M\$ each)

### ◆ **Chameleon**

- ★ <https://www.chameleoncloud.org/>
- ★ Codebase: CHI (CHameleon Infrastructure)  
CHI = OpenStack + Grid'5000 tools + custom developments
- ★ Currently two sites, 317 nodes (+ KVM Cloud)

### ◆ **CloudLab**

- ★ <http://www.cloudlab.us>
- ★ Emulab codebase as a foundation
- ★ Two three sites, 315 (arm64) + 100 + 100 (x86) nodes

**How do they compare: design choices? features?  
What does it tell us about the future?**

Disclaimer: to the best of my knowledge, and as of January 2016

# Bare-metal reconfiguration

## Absolute requirement for such testbeds:

- ▶ Customize the software environment
- ▶ Enable experiments at all software layers
  - ◆ Including virtualization technologies  $\leadsto$  control of the host OS
- ▶ Testbeds = Meta-Clouds

## Available on all three testbeds:

- ▶ **Grid'5000**: Kadeploy
- ▶ **Chameleon**: OpenStack Ironic
- ▶ **CloudLab**: Emulab's Frisbee

Similar performance (10 nodes in  $\ll$  10 mins), but different technical choices

# Resources description and selection

## Requirements:

- ▶ Fine-grained, complete descriptions
- ▶ Accurate, up-to-date, verified
- ▶ Versioned, archived

Users can **navigate descriptions** and **select resources** matching their experiment's requirements

# Resources description and selection

## Requirements:

- ▶ Fine-grained, complete descriptions
- ▶ Accurate, up-to-date, verified
- ▶ Versioned, archived

Users can **navigate descriptions** and **select resources** matching their experiment's requirements

## Status:

- ▶ **Grid'5000 and Chameleon**: reference API + g5k-checks
- ▶ **CloudLab**:
  - ◆ machine-readable description using RSpec *advertisement* format (less detailed than Grid'5000's)
  - ◆ verification: LinkTest and CheckNode

# Grid'5000 / Chameleon: Reference API<sup>1</sup>

- ▶ Describing resources  $\leadsto$  understand results
  - ◆ Covering nodes, network equipment, topology
  - ◆ Machine-parsable format (JSON)  $\leadsto$  scripts
  - ◆ Archived, versioned (*State of testbed 6 months ago?*)

```
"processor": {
  "cache_l2": 8388608,
  "cache_l1": null,
  "model": "Intel Xeon",
  "instruction_set": "",
  "other_description": "",
  "version": "X3440",
  "vendor": "Intel",
  "cache_l1i": null,
  "cache_l1d": null,
  "clock_speed": 2530000000.0
},
"uid": "graphene-1",
"type": "node",
"architecture": {
  "platform_type": "x86_64",
  "smt_size": 4,
  "smp_size": 1
},
"main_memory": {
  "ram_size": 17179869184,
  "virtual_size": null
},
"storage_devices": [
  {
    "model": "Hitachi HDS72103",
    "size": 298023223876.953,
    "driver": "ahci",
    "interface": "SATA II",
    "rev": "JPFO",
    "device": "sda"
  }
]
```

---

<sup>1</sup>David Margery et al. "Resources Description, Selection, Reservation and Verification on a Large-scale Testbed". In: *TRIDENTCOM. 2014*.

# Grid'5000 / Chameleon: Reference API<sup>1</sup>

- ▶ **Describing** resources  $\leadsto$  understand results
  - ◆ Covering nodes, network equipment, topology
  - ◆ Machine-parsable format (JSON)  $\leadsto$  scripts
  - ◆ Archived, versioned (*State of testbed 6 months ago?*)
- ▶ **Verifying** the description
  - ◆ Avoid inaccuracies/errors  $\leadsto$  wrong results
  - ◆ Could **happen frequently**: maintenance, broken hardware (e.g. RAM)
  - ◆ Our solution: **g5k-checks**
    - ★ Runs at node boot (or manually by users)
    - ★ Acquires info using OHAI, ethtool, etc.
    - ★ Compares with Reference API

```
"processor": {
  "cache_l2": 8388608,
  "cache_l1": null,
  "model": "Intel Xeon",
  "instruction_set": "",
  "other_description": "",
  "version": "X3440",
  "vendor": "Intel",
  "cache_l1i": null,
  "cache_l1d": null,
  "clock_speed": 2530000000.0
},
"uid": "graphene-1",
"type": "node",
"architecture": {
  "platform_type": "x86_64",
  "smt_size": 4,
  "smp_size": 1
},
"main_memory": {
  "ram_size": 17179869184,
  "virtual_size": null
},
"storage_devices": [
  {
    "model": "Hitachi HDS72103",
    "size": 298023223876.953,
    "driver": "ahci",
    "interface": "SATA II",
    "rev": "JPFO",
    "device": "sda"
  }
],
```

---

<sup>1</sup>David Margery et al. "Resources Description, Selection, Reservation and Verification on a Large-scale Testbed". In: *TRIDENTCOM. 2014*.



# Grid'5000 / Chameleon: Reference API<sup>1</sup>

- ▶ **Describing** resources  $\leadsto$  understand results
  - ◆ Covering nodes, network equipment, topology
  - ◆ Machine-parsable format (JSON)  $\leadsto$  scripts
  - ◆ Archived, versioned (*State of testbed 6 months ago?*)
- ▶ **Verifying** the description
  - ◆ Avoid inaccuracies/errors  $\leadsto$  wrong results
  - ◆ Could **happen frequently**: maintenance, broken hardware (e.g. RAM)
  - ◆ Our solution: **g5k-checks**
    - ★ Runs at node boot (or manually by users)
    - ★ Acquires info using OHAI, ethtool, etc.
    - ★ Compares with Reference API
- ▶ **Selecting** resources on Grid'5000
  - ◆ OAR database filled from Reference API

```
oarsub -p "wattmeter='YES' and gpu='YES' and eth10g='Y'"
```

```
"processor": {
  "cache_l2": 8388608,
  "cache_l1": null,
  "model": "Intel Xeon",
  "instruction_set": "",
  "other_description": "",
  "version": "X3440",
  "vendor": "Intel",
  "cache_l1i": null,
  "cache_l1d": null,
  "clock_speed": 2530000000.0
},
"uid": "graphene-1",
"type": "node",
"architecture": {
  "platform_type": "x86_64",
  "smt_size": 4,
  "smp_size": 1
},
"main_memory": {
  "ram_size": 17179869184,
  "virtual_size": null
},
"storage_devices": [
  {
    "model": "Hitachi HDS72103",
    "size": 298023223876.953,
    "driver": "ahci",
    "interface": "SATA II",
    "rev": "JPF0",
    "device": "sda"
  }
],
```

<sup>1</sup>David Margery et al. "Resources Description, Selection, Reservation and Verification on a Large-scale Testbed". In: *TRIDENTCOM. 2014*.

# Resources selection on Chameleon

Filter nodes using the options below, then generate a reservation script to reserve those nodes.

Applied Filters: **Main Memory Humanized Ram Size: 128 GiB** ✖

**317 nodes** filtered from 337 originally.

Compute Nodes (317)     Storage Nodes     With GPU     With Infiniband Support (38)

Site	Cluster	Platform Type	# CPUs	# Cores	RAM Size
<input type="checkbox"/> Tacc (233)	<input checked="" type="checkbox"/> Chameleon (317)	<input checked="" type="checkbox"/> X86 #64 (317)	<input checked="" type="checkbox"/> 2 (317)	<input checked="" type="checkbox"/> 48 (317)	<input checked="" type="checkbox"/> 128 GiB (317)
<input type="checkbox"/> Uc (84)					

**Advanced Filters**

Search by space separated keywords.

**Bios**

- Release Date**
  - 03/09/2015 (317)
- Vendor**
  - Dell Inc. (317)
- Version**
  - 1.2 (317)

**Chassis**

- Manufacturer**
  - Dell Inc. (317)
- Name**
  - PowerEdge R630 (317)

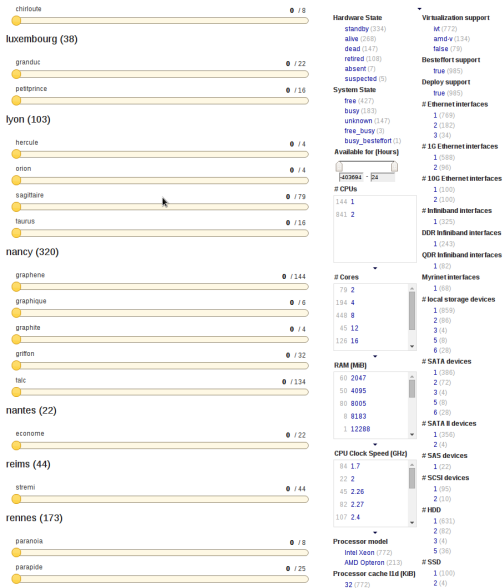
**Monitoring**

... ..

**317 nodes** filtered from 337 originally.

<https://www.chameleoncloud.org/user/discovery/>

# Another selection UI on Grid'5000



# Resources verification on CloudLab

- ▶ **LinkTest<sup>23</sup>**: validate network performance configuration (latency, routing, link loss, bandwidth)
- ▶ **CheckNode<sup>4</sup>**
  - ◆ Similar to G5K-checks
  - ◆ Collects data about CPU, memory, disks, NICs
  - ◆ But no link to RSpec description

---

<sup>2</sup>D.S. Anderson et al. "Automatic Online Validation of Network Configuration in the Emulab Network Testbed". In: *ICAC'06*.

<sup>3</sup><https://wiki.emulab.net/wiki/linktest>

<sup>4</sup><https://wiki.emulab.net/wiki/checknode>

# Resources reservation

## Requirements:

- ▶ Ensure that all users can get a *fair* share of resources
- ▶ Avoid wasting of resources
- ▶ Without preventing large-scale experiments

# Resources reservation

## Requirements:

- ▶ Ensure that all users can get a *fair* share of resources
- ▶ Avoid wasting of resources
- ▶ Without preventing large-scale experiments

## Status:

- ▶ **Grid'5000**: batch scheduler (OAR) with advance reservation
  - ◆ And usage policy:
    - ★ Shared usage during the day
    - ★ Large reservations on nights and week-ends
- ▶ **Chameleon**: leases using OpenStack Blazar
  - ◆ Supports advance reservation
  - ◆ Set of *Best Practices* to promote fairness
  - ◆ Duration limit of one week for reservations (since 12/2015)
- ▶ **CloudLab**: experiments start immediately, default duration of a few hours, can be extended on demand (no advance reservations)

# Network reconfiguration and SDN

**Requirements:** (still to be clarified for Software Defined Networking)

- ▶ Create custom topologies featuring multiple L2 networks
- ▶ Provide network emulation features
- ▶ Higher-level support for creating OpenFlow-managed networks?
- ▶ Reconfigure network switches? Reinstall them  $\rightsquigarrow$  white box switches?

---

<sup>5</sup><http://distem.gforge.inria.fr>

<sup>6</sup><http://cloudlab-announce.blogspot.com/2015/06/using-openflow-in-cloudlab.html>

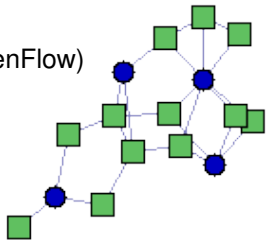
# Network reconfiguration and SDN

**Requirements:** (still to be clarified for Software Defined Networking)

- ▶ Create custom topologies featuring multiple L2 networks
- ▶ Provide network emulation features
- ▶ Higher-level support for creating OpenFlow-managed networks?
- ▶ Reconfigure network switches? Reinstall them  $\leadsto$  white box switches?

**Status:**

- ▶ **Grid'5000:** KaVLAN (VLAN reconfiguration on switches)  
The creation of topologies is still up to the user (WIP)
  - ◆ + Distem<sup>5</sup> for network emulation
- ▶ **Chameleon:** shared L2 net (more planned w/ OpenFlow)
- ▶ **CloudLab:**
  - ◆ Emulab's network emulation features
  - ◆ OpenFlow access on switches<sup>6</sup>
  - ◆ Interconnection to Internet2's AL2S



<sup>5</sup><http://distem.gforge.inria.fr>

<sup>6</sup><http://cloudlab-announce.blogspot.com/2015/06/using-openflow-in-cloudlab.html>



# Experiment monitoring

**Goal: enable users to understand what happens during their experiment**

▶ **Grid'5000:**

- ◆ System-level sensors with Ganglia (CPU, mem, load, net, processes)
- ◆ Infrastructure-level sensors with Kwapi (network, power)
  - ★ Captured at high frequency (~1 per second)
  - ★ Live visualization
  - ★ REST API and long-term storage

▶ **Chameleon:** System-level sensors with OpenStack Ceilometer

- ◆ Load, disk usage, memory, network traffic, I/O traffic

▶ **CloudLab:** Infrastructure-level sensors for power: WIP<sup>78</sup>

---

<sup>7</sup><http://docs.cloudlab.us/planned.html>

<sup>8</sup><http://groups.geni.net/geni/wiki/GENIFireCollaborationWorkshopSeptember2015/Session6>

# Long-term storage

## Requirements:

- ▶ Store large datasets, experiments results, etc. between experiments

## Status:

- ▶ **Grid'5000**: storage5k (file-based), permanent Ceph (WIP)
- ▶ **Chameleon**: file-based object store (OpenStack Swift)
- ▶ **CloudLab**: yes<sup>9</sup>: file store and block store, with versioning (using ZFS) (the snapshots features are not documented yet)

---

<sup>9</sup><http://cloudlab-announce.blogspot.fr/2015/04/persistent-dataset.html>

# Storage during experiments

## Requirements:

- ▶ Enable experiments with large amounts of data
- ▶ On nodes with large number of disks

## Status:

- ▶ All testbeds have nodes with large numbers of disks
  - ◆ **Grid'5000**: 40 nodes with 4+ HDD, 28 with 5 HDD and 1 SSD
  - ◆ **Chameleon**: 20 nodes with 15 HDD and 1 SSD
  - ◆ **CloudLab**: 10 nodes with 13 HDD and 1 SSD
- ▶ But all require users to transfer their data from long-term storage at the beginning of each experiment (if users want node-local storage)

# Appliances / software stacks deployment

(Conflicting?) requirements:

- ▶ Software stacks **useful to experimenters**
  - ◆ Recent versions (to stay **relevant**)
  - ◆ **Easy to use** (low entry barrier)
  - ◆ Easily **customizable**  $\rightsquigarrow$  replace components
- ▶ **Maintainable** in the long run (despite 6-month release cycles)

---

<sup>10</sup><https://www.chameleoncloud.org/docs/appliances/>

<sup>11</sup><https://www.cloudlab.us/show-profile.php?uuid=aa4c185b-9adc-11e5-9f8c-020cbce80001>

# Appliances / software stacks deployment

(Conflicting?) requirements:

- ▶ Software stacks **useful to experimenters**
  - ◆ Recent versions (to stay **relevant**)
  - ◆ **Easy to use** (low entry barrier)
  - ◆ Easily **customizable**  $\rightsquigarrow$  replace components
- ▶ **Maintainable** in the long run (despite 6-month release cycles)

Status:

- ▶ **Grid'5000**:
  - ◆ OpenStack available: Liberty (with external users)
  - ◆ Ceph deployment tool in beta status
- ▶ **Chameleon**: *marketplace* in preview<sup>10</sup>
- ▶ **CloudLab**: OpenStack *profile* available (Kilo or Juno)<sup>11</sup>. Also Hadoop
  - ◆ GUI, large number of configuration parameters
  - ◆ But not clear how easily it could be customized or extended

---

<sup>10</sup><https://www.chameleoncloud.org/docs/appliances/>

<sup>11</sup><https://www.cloudlab.us/show-profile.php?uuid=aa4c185b-9adc-11e5-9f8c-020cbce80001>

# Federation

## Three levels of federation:

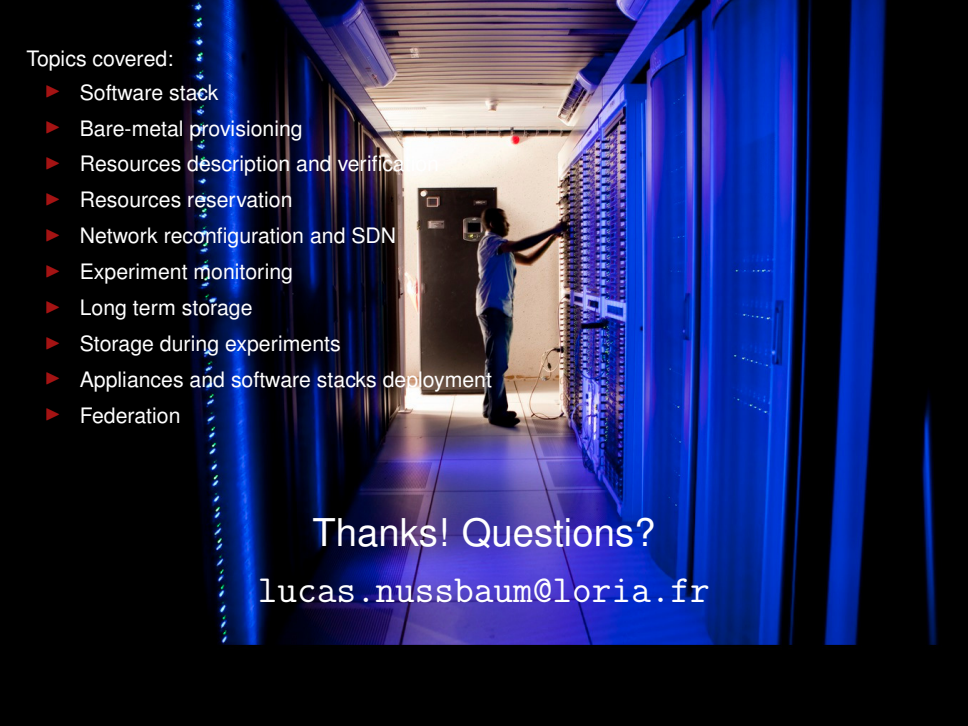
- ▶ Identity (same account to access several testbeds)
- ▶ API (same API to access several testbeds)
- ▶ Data plane (cross-testbed experiments possible)

## Status:

- ▶ Chameleon and Cloudlab: identity federation with GENI
- ▶ Grid'5000 and Chameleon: common resources description API
- ▶ CloudLab: compatible with the GENI APIs
- ▶ CloudLab: cross-sites experiments with Internet2's AL2S

# Conclusions

- ▶ For a long time, Grid'5000 was quite unique
- ▶ Now: **competition, with different ideas, strategies, backgrounds**
  - ◆ No testbed is feature-complete
  - ◆ CloudLab & Grid'5000: *in-house developments*  
vs Chameleon: *stand on off-the-shelf components, improve*  
~> 😊 faster   😞 flexible enough?
  - ◆ CloudLab: *roots in the Emulab network testbed*  
vs Grid'5000 & Chameleon: *roots in HPC*
- ▶ Progress during next months/years will be interesting to follow!

A photograph of a server room with blue ambient lighting. A person in a light blue shirt and dark pants is standing in the center, reaching into a server rack on the right. The room is filled with rows of server racks on both sides, and the floor is a light-colored tile. The lighting is primarily blue, with some white light from the ceiling fixtures.

Topics covered:

- ▶ Software stack
- ▶ Bare-metal provisioning
- ▶ Resources description and verification
- ▶ Resources reservation
- ▶ Network reconfiguration and SDN
- ▶ Experiment monitoring
- ▶ Long term storage
- ▶ Storage during experiments
- ▶ Appliances and software stacks deployment
- ▶ Federation

Thanks! Questions?

`lucas.nussbaum@loria.fr`