SeDuCe

a Testbed for research on thermal and power management in datacenters

Jonathan Pastor
Jean-Marc Menaud
IMT Atlantique - Nantes

Outline

- Context
- The SeDuCe testbed
- Experimentation example with the testbed
- Future work
- Conclusion

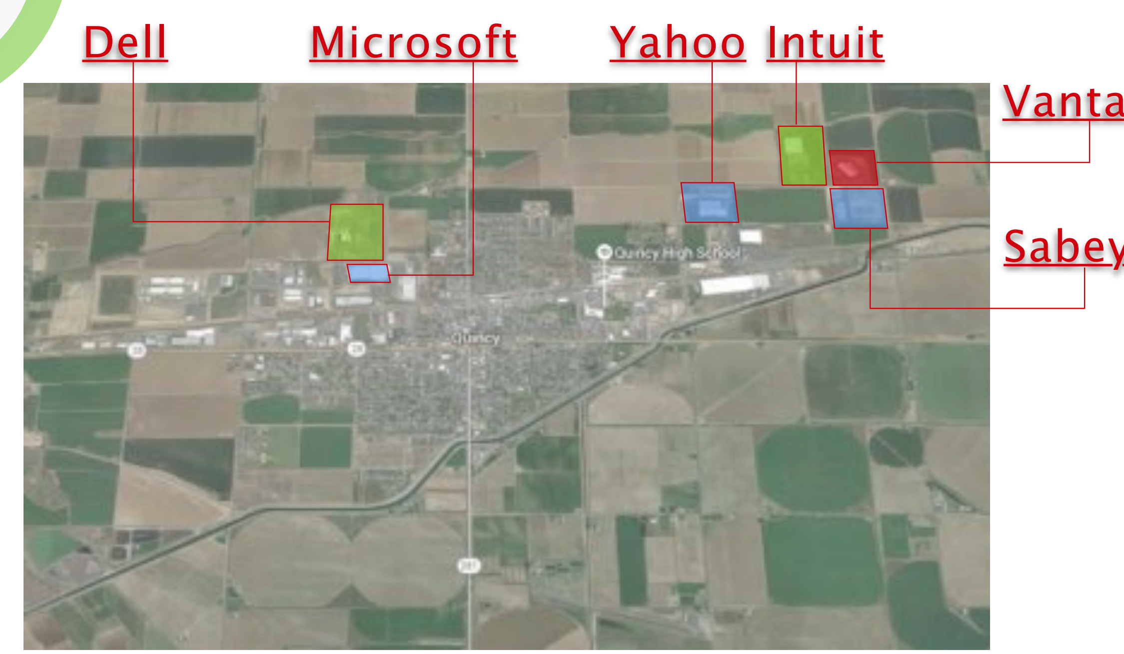
Context

Datacenters

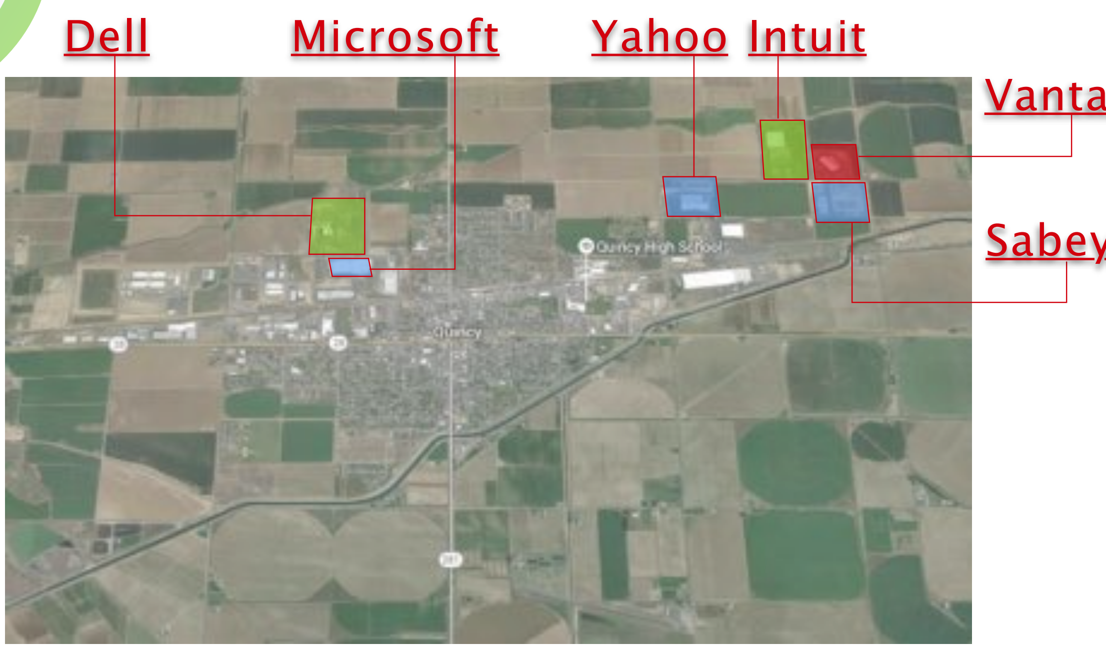
Datacenters

Datacenters

Datacenters



Datacenters



Open challenges

- Software (large scale, fault tolerance, network latency)
  - *Fog computing*
- Energetic (power distribution / cooling)

[2]

Open challenges

- Software (large scale, fault tolerance, network latency)
  - *Fog computing*
- Energetic (power distribution / cooling)

[2]

Open challenges

- Software (large scale, fault tolerance, network latency)
  - *Fog computing*
- Energetic (power distribution / cooling)

	% US electrical production
2000	0.8%
2005	1.5%
2014	2.2% ?

Electrical consumption of US datacenters [1]

Amortized Cost	Component	Sub-Components
~45%	Servers	CPU, memory, storage systems
~25%	Infrastructure	Power distribution and cooling
~15%	Power draw	Electrical utility costs
~15%	Network	Links, transit, equipment

Table 1: Guide to where costs go in the data center. ***[2]***

Few approaches

- An effort has been made to improve energy efficiency of components
- Choice of areas with affordable cooling
- Use renewable energies
- Reuse heat produced by computers

Few approaches

- An effort has been made to improve energy efficiency of components
- Choice of areas with affordable cooling
- Use renewable energies
- Reuse heat produced by computers

Few approaches

- An effort has been made to improve energy efficiency of components

- Choice of areas with affordable cooling

- Use renewable energies

- Reuse heat produced by

Few approa[…]

- An effort has been made to[…] efficiency of components
- Choice of areas with afford[…]
- Use renewable energies
- Reuse heat produced by computers

Few approaches

- An effort
  efficiency
- Choice o
- Use rene
- Reuse he

Experimental research on energy in datacenters

- Datacenters consume a lot of energy (power supply of hardware, cooling, …) **[1], [2]**
- A lot of the research on energy in DCs is based on simulations : few public testbeds offer monitoring of energy consumption of their servers (Grid'5000 proposes *Kwapi*)
- As far as we know, no public testbed provide thermal monitoring of servers
- Energy and Temperature are two related physical quantities
- Lack of a testbed that proposes both thermal and energetic monitoring of its servers

The SeDuCe testbed

G5K + SeDuCe = Ecotype

- Grid’5000 is a french scientific testbed that provides bare metal computing resources to researchers in Distributed Systems.
- Grid’5000 is a distributed infrastructure composed of 8 sites hosting clusters of servers
- SeDuCe is a testbed hosted in Nantes and integrated with Grid’5000
- SeDuCe aims at easing the process of conducting experiments that combine both thermal and power aspect of datacenters

Ecotype

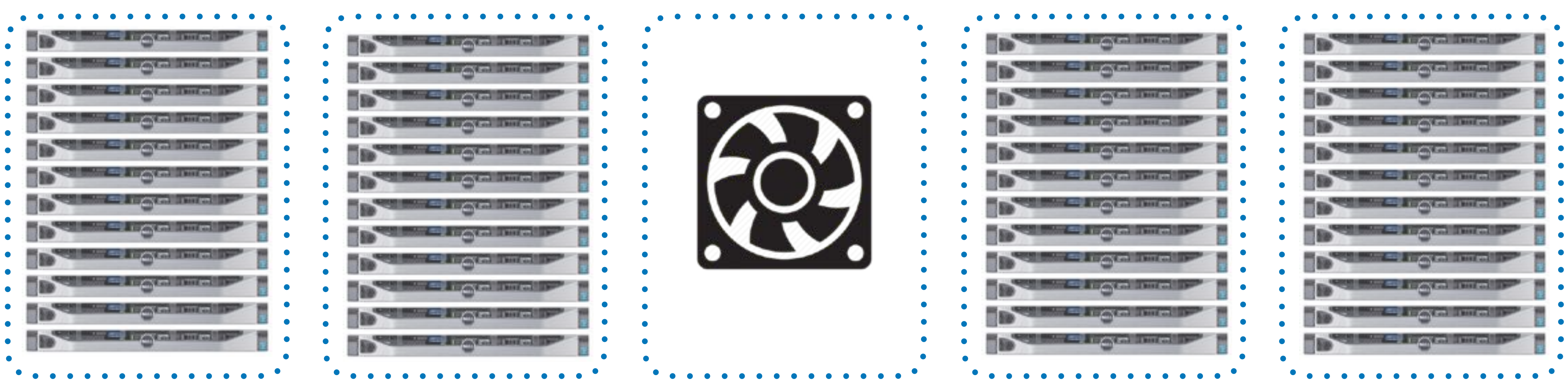
- *Ecotype* is the new Grid'5000 cluster hosted at IMT Atlantique in Nantes

- 48 servers based on Dell R630 designed to operate at up to 35°C

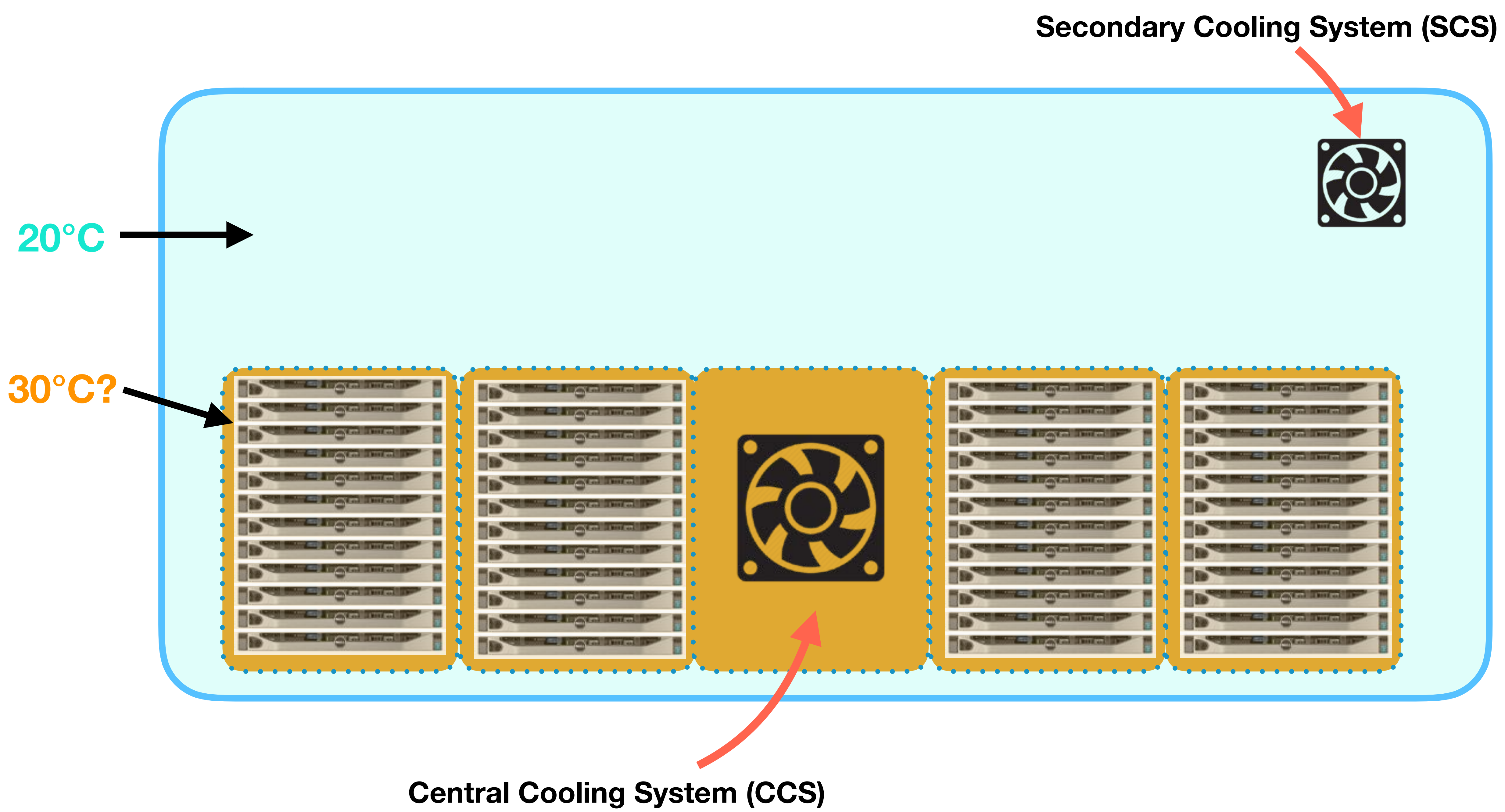
  2x10 cores (2x20 threads), 128GB RAM, 400GB SSDs

- 5 Air tight racks based on *Schneider Electrics IN-ROW*

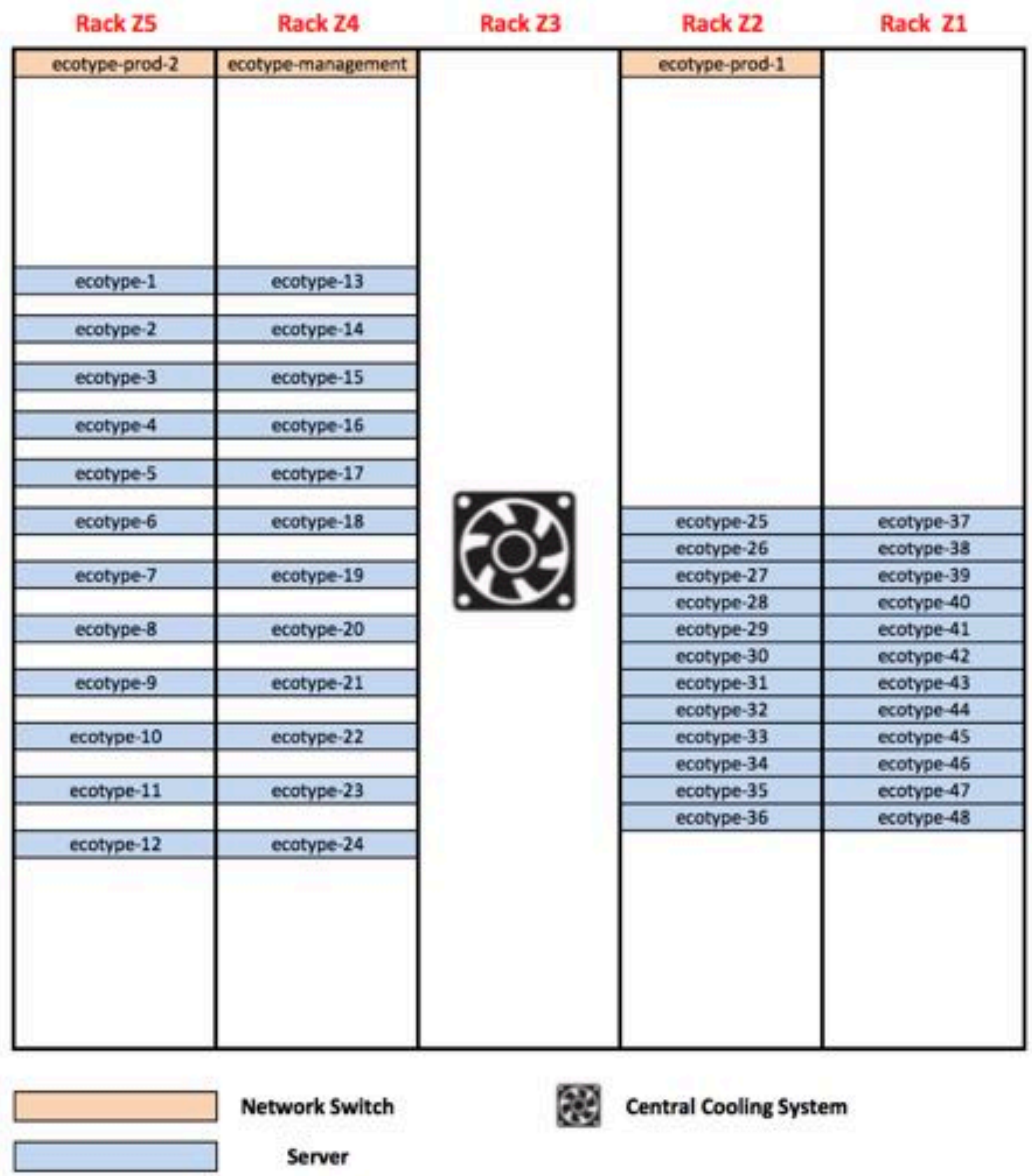
- *Servers are monitored with temperature sensors and wattmeters*



Room architecture



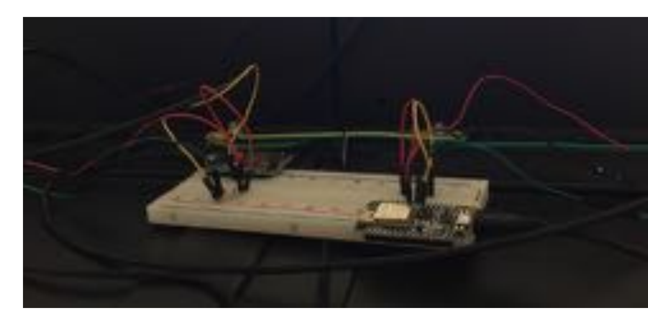
Room architecture



Thermal and power monitoring

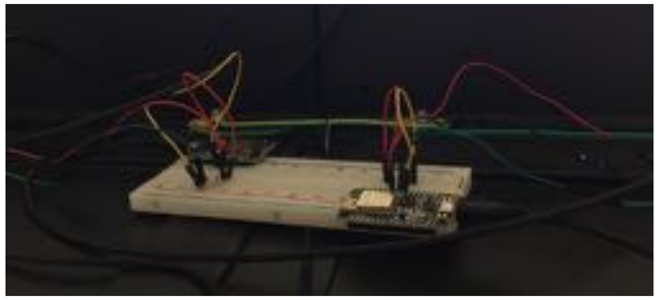
- The energy consumption of each element of the testbed is monitored (one record per second)

- Each sub component of the CCS (fans, condensator, …) is monitored
- Temperature of servers is monitored (one record per seconds)



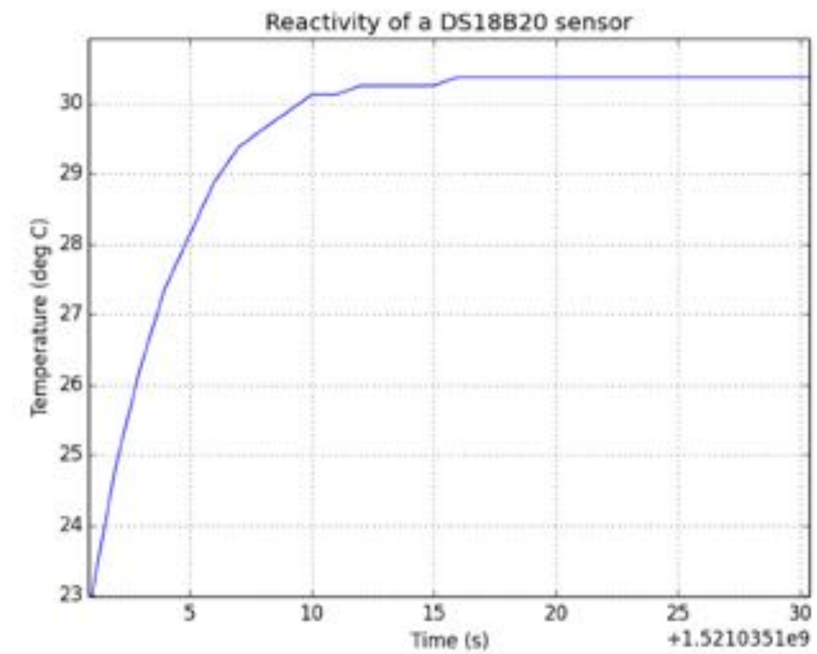
Temperature sensors

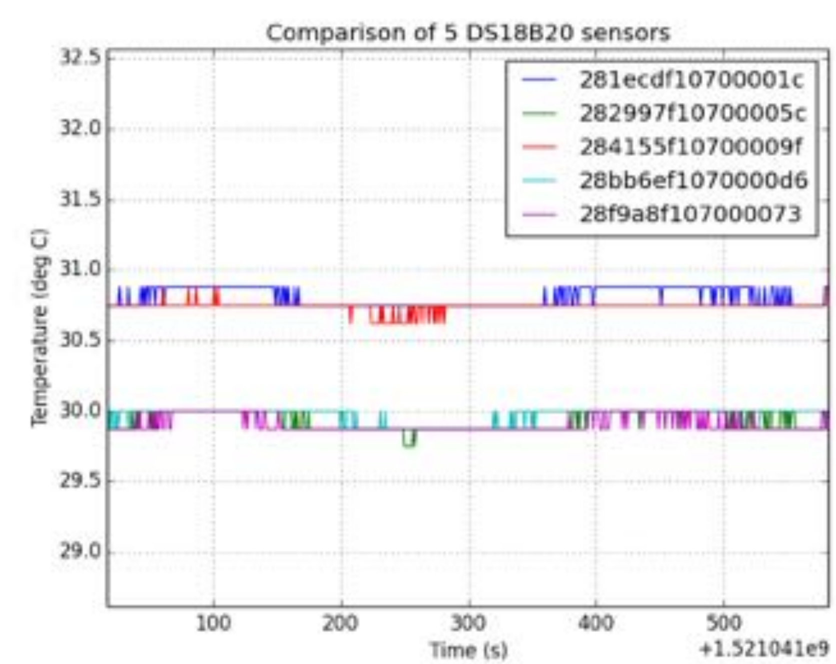
- Based on DS18B20 (unit cost: 3$)
- 96 sensors installed on 8 buses
- Each bus is connected to an arduino (oneWire protocol)
- Arduinos push data to a web service
- Thermal inertia : they fit in environment where temperature changes smoothly



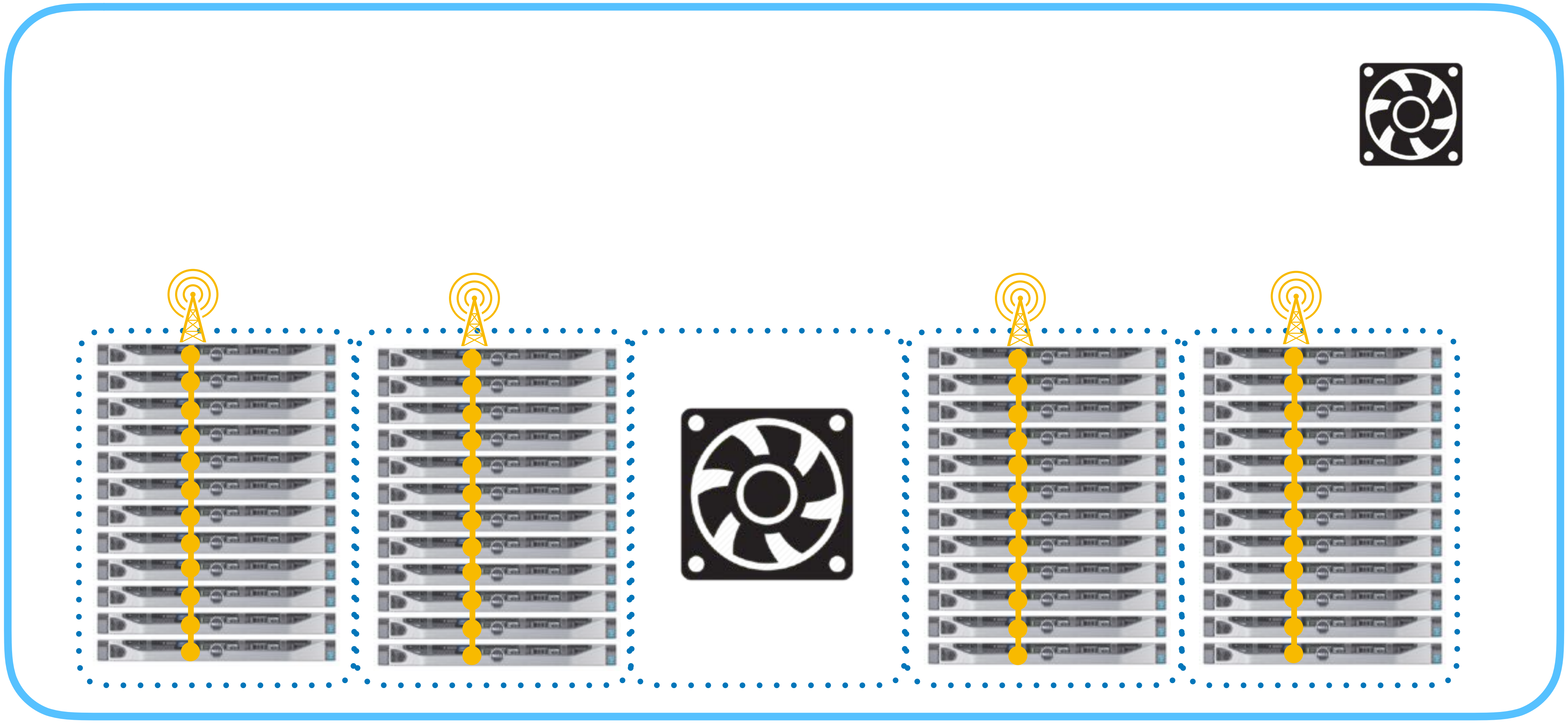
Temperature sensors

- Based on DS18B20 (unit cost: 3$)
- 96 sensors installed on 8 buses
- Each bus is connected to an arduino (oneWire protocol)
- Arduinos push data to a web service
- Thermal inertia : they fit in environment where temperature changes smoothly

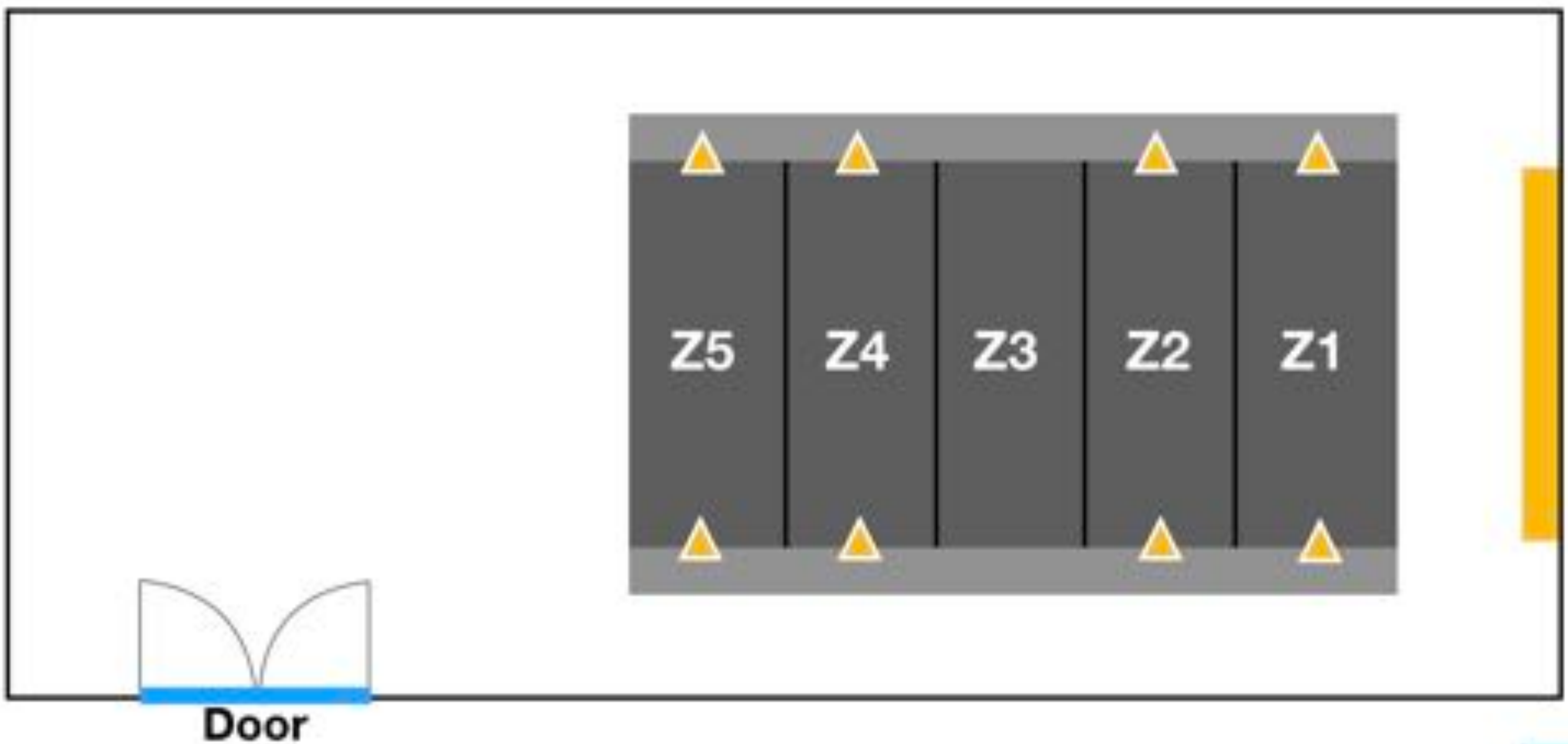




Temperature sensors



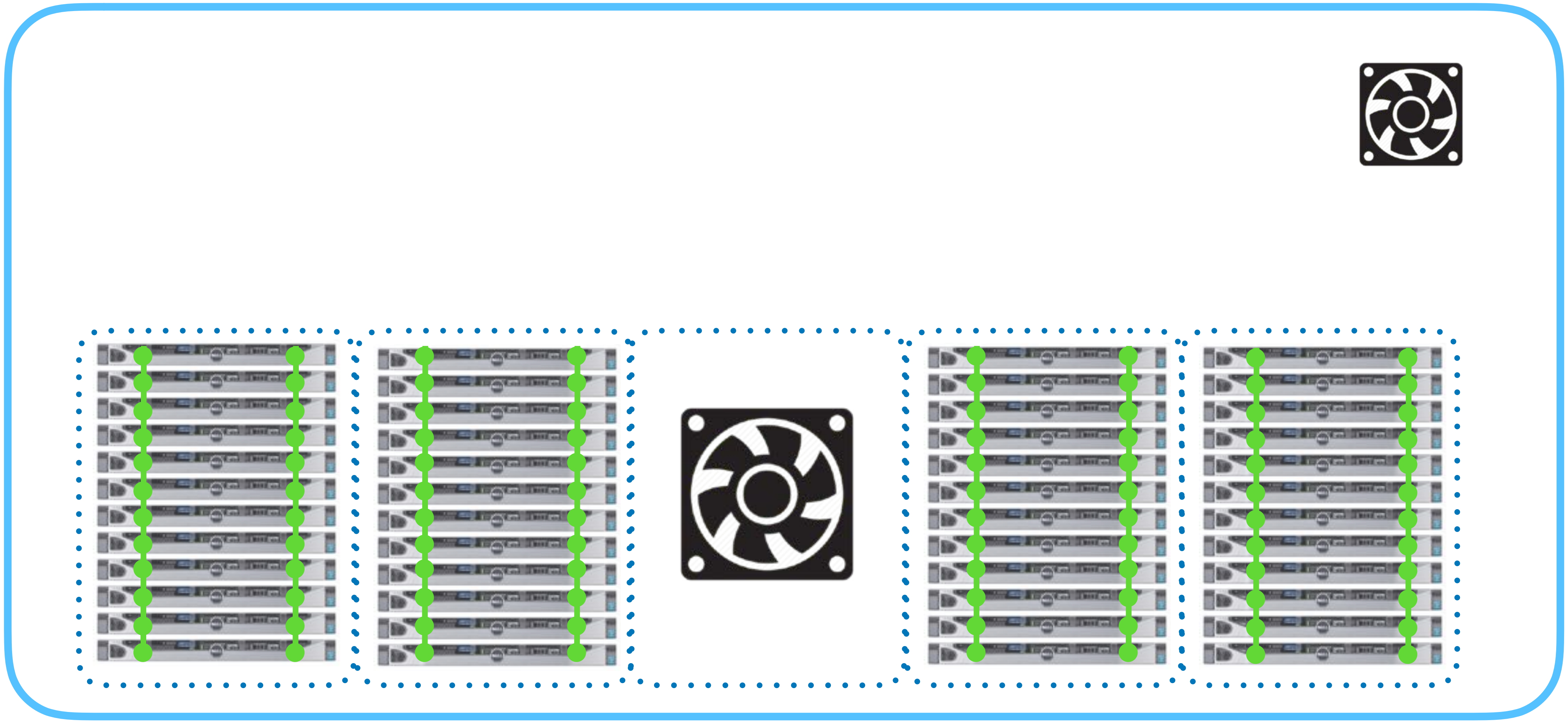
Temperature sensors



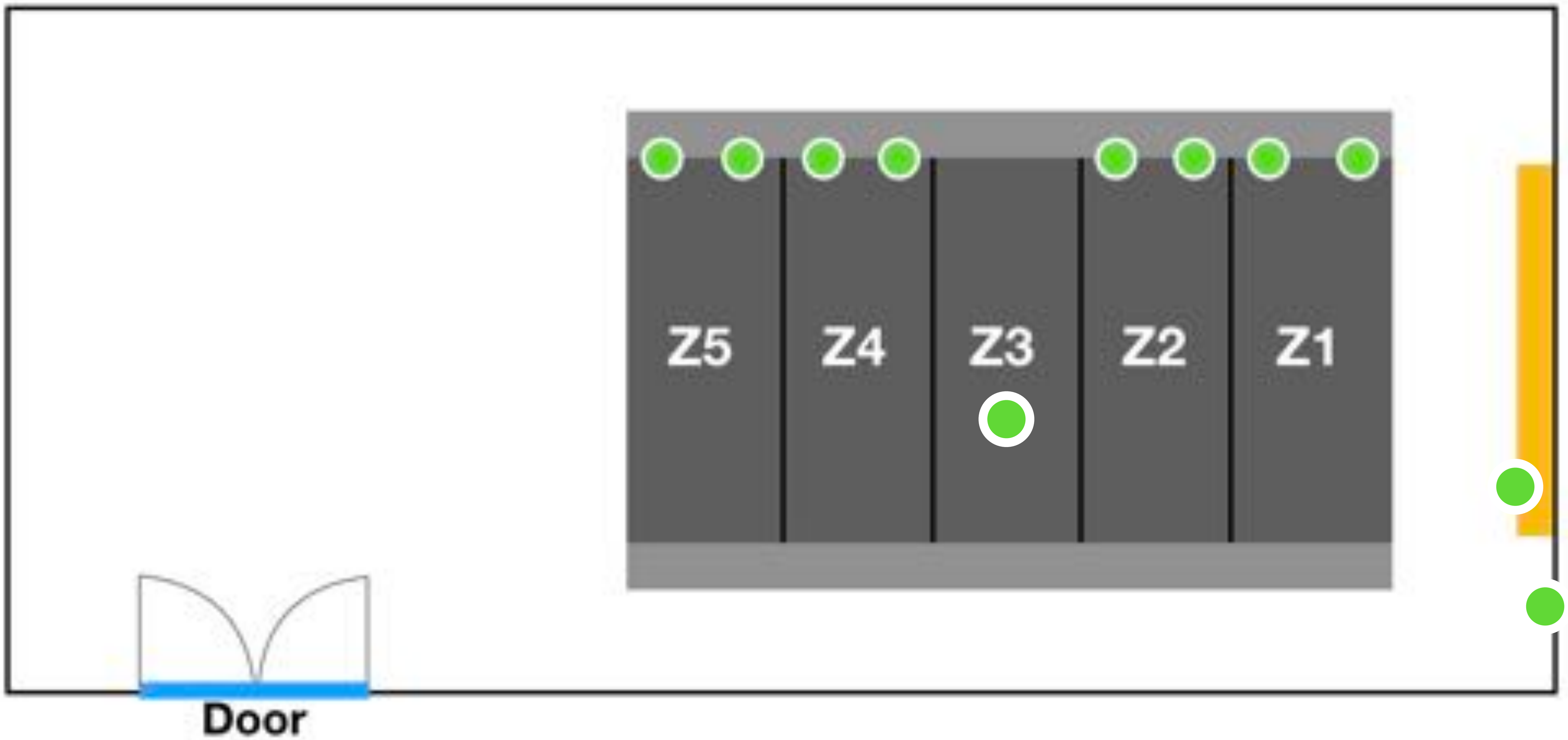
Power monitoring

- Wattmeters integrated in APC PDUs
- Each server has 2 power outlets and is connected to 2 PDUs
- 1 record per outlet per second
- PDUs are connected to a management network
- Network switches, cooling systems (fans, condensator) are also monitored (PDUS, Flukso, Socometers)

Wattmeters

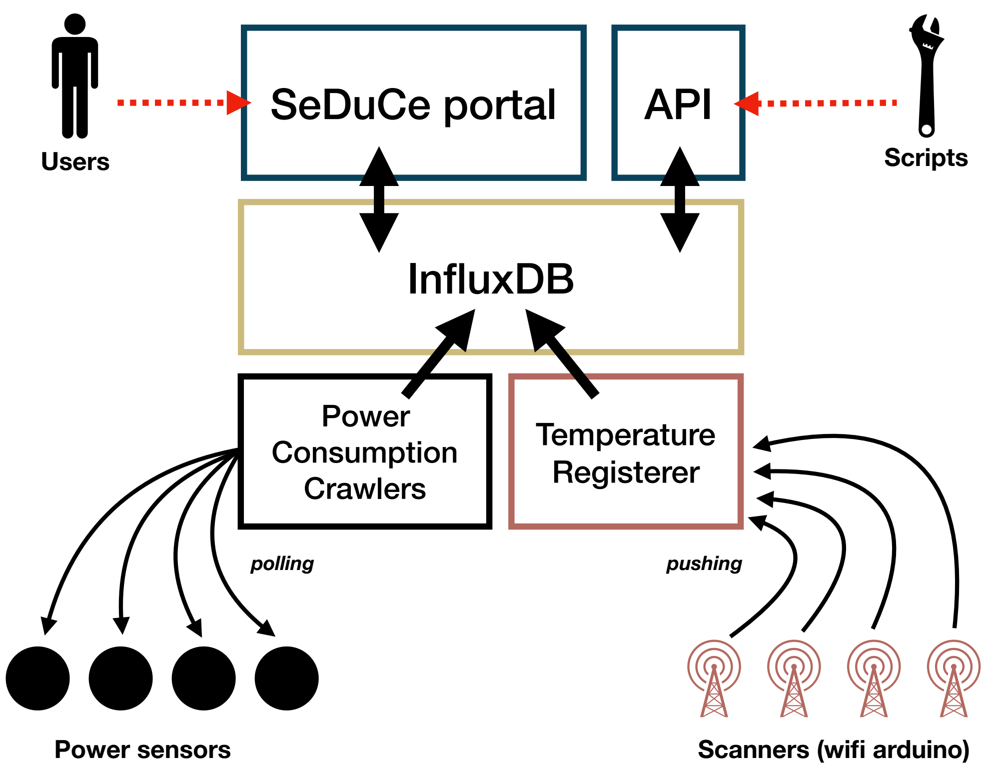


Wattmeters

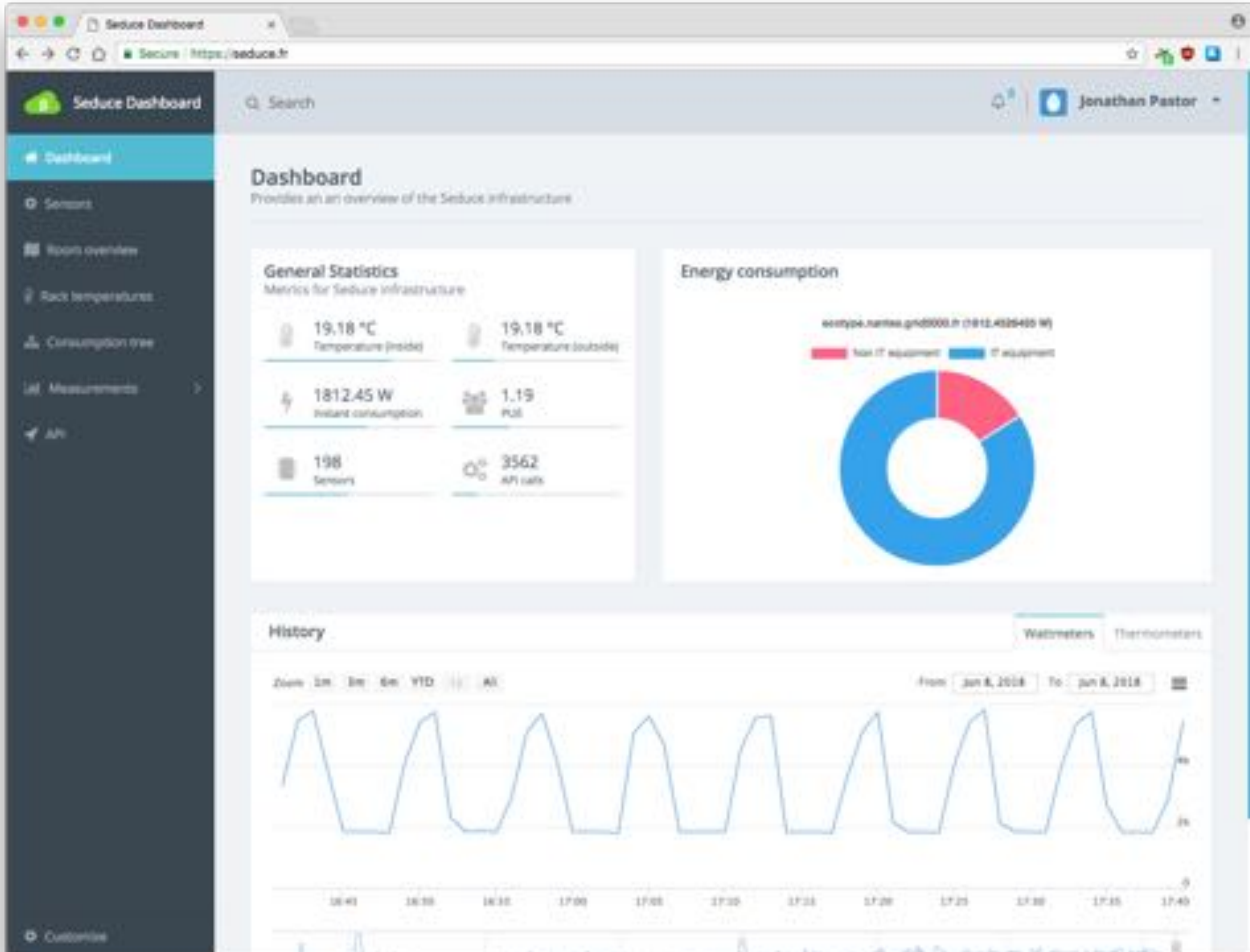


Architecture of the SeDuCe platform

- Arduinos push data to a web service (temperature registerer)
- Power consumption crawlers poll data from PDUs and other power monitoring devices
- Data is stored in InfluxDB (time serie oriented database)
- Users can access to data of the testbed via:
  - a web dashboard: https://seduce.fr
  - a documented Rest API: https://api.seduce.fr
- Dashboard and API fetch data from InfluxDB

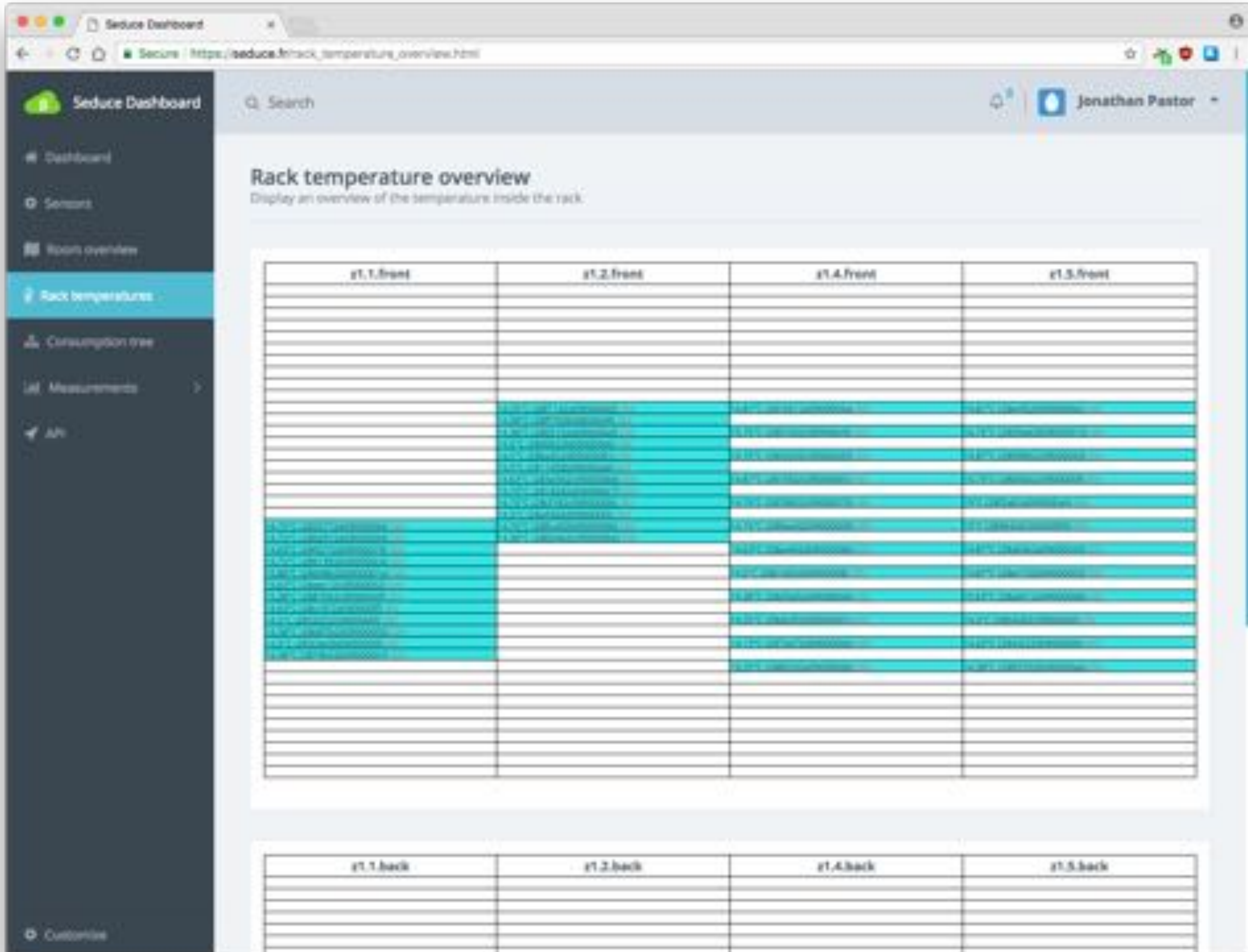


seduce.fr

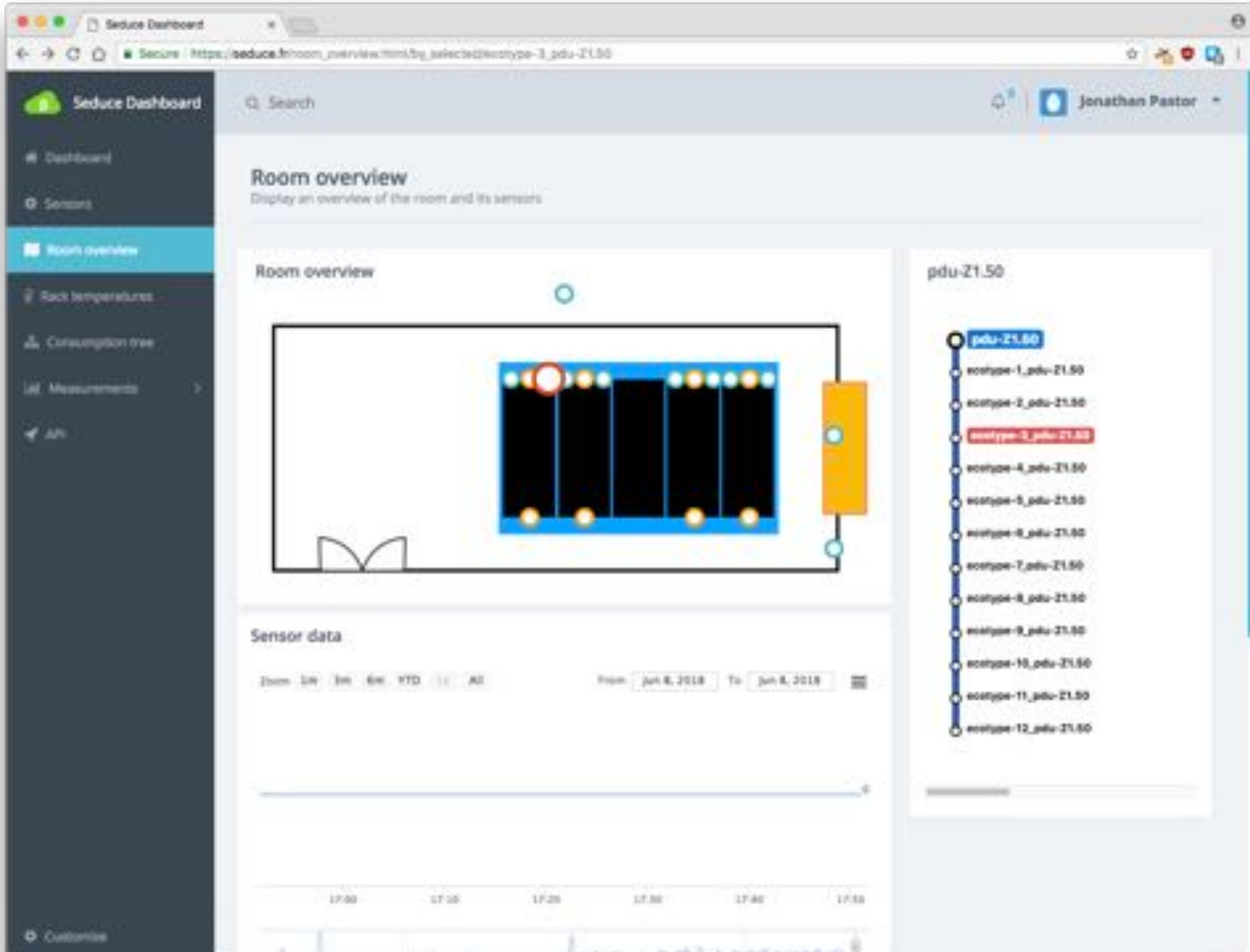


Seduce Dashboard
https://seduce.fr
Seduce Dashboard
Search
Jonathan Pastor
Dashboard
Sensors
Rooms overview
Rack temperatures
Consumption tree
Measurements
API
Customize
Dashboard
Provides an an overview of the Seduce infrastructure
General Statistics
Metrics for Seduce infrastructure
19.18 °C
Temperature (inside)
19.18 °C
Temperature (outside)
1812.45 W
Instant consumption
1.19
PUE
198
Sensors
3562
API calls
Energy consumption
Non IT equipment
IT equipment
History
Wattmeters
Thermometers
Zoom
1m
3m
6m
YTD
All
From
To

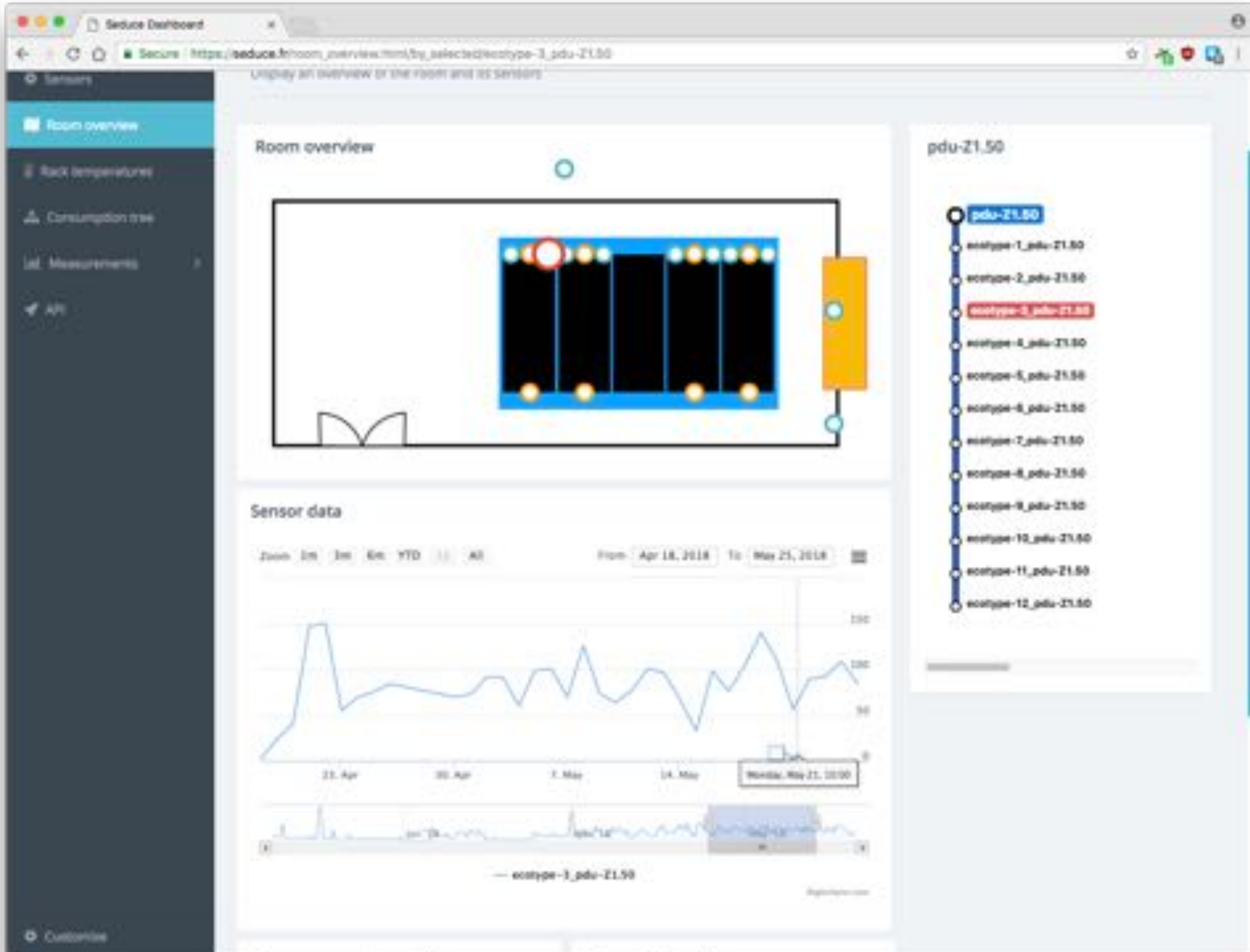
seduce.fr


Seduce Dashboard
https://seduce.fr/rack_temperature_overview.html
Seduce Dashboard
Search
Jonathan Pastor
Dashboard
Sensors
Room overview
Rack temperatures
Consumption tree
Measurements
API
Customize
Rack temperature overview
Display an overview of the temperature inside the rack
z1.1.front
z1.2.front
z1.4.front
z1.5.front
z1.1.back
z1.2.back
z1.4.back
z1.5.back

seduce.fr


Seduce Dashboard
Search
Jonathan Pastor
Room overview
Room overview
pdu-Z1.50
Sensor data

seduce.fr



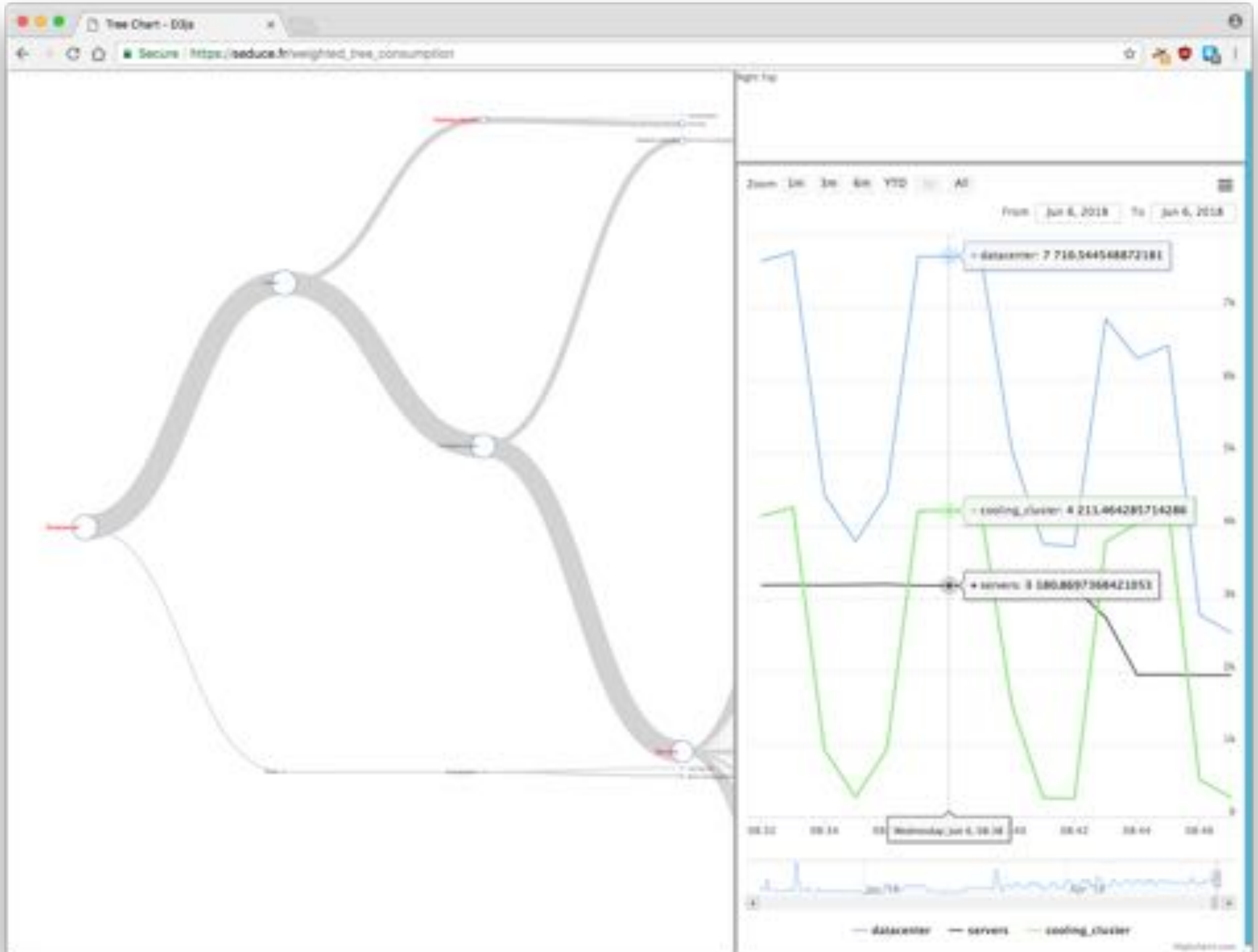
seduce.fr

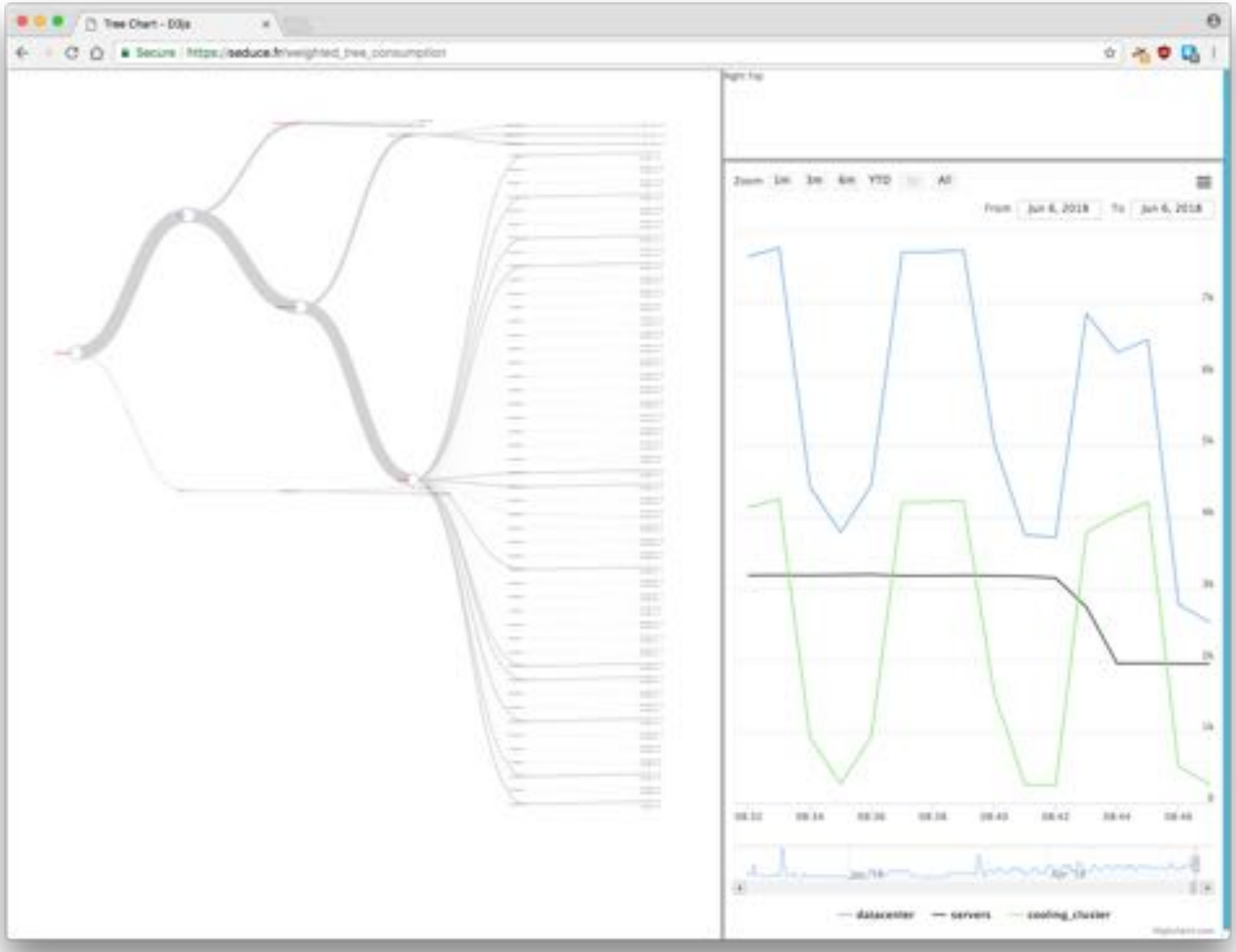
Seduce Dashboard
Room overview
Rack temperatures
Consumption tree
Measurements
API
Customize
Room overview
pdu-Z1.50
Sensor data
ecotype-3_pdu-Z1.50

seduce.fr

Seduce Dashboard
Room overview
Rack temperatures
Consumption tree
Measurements
API
Customize
Room overview
Sensor data
pdu-Z1.50
ecotype-3_pdu-Z1.50

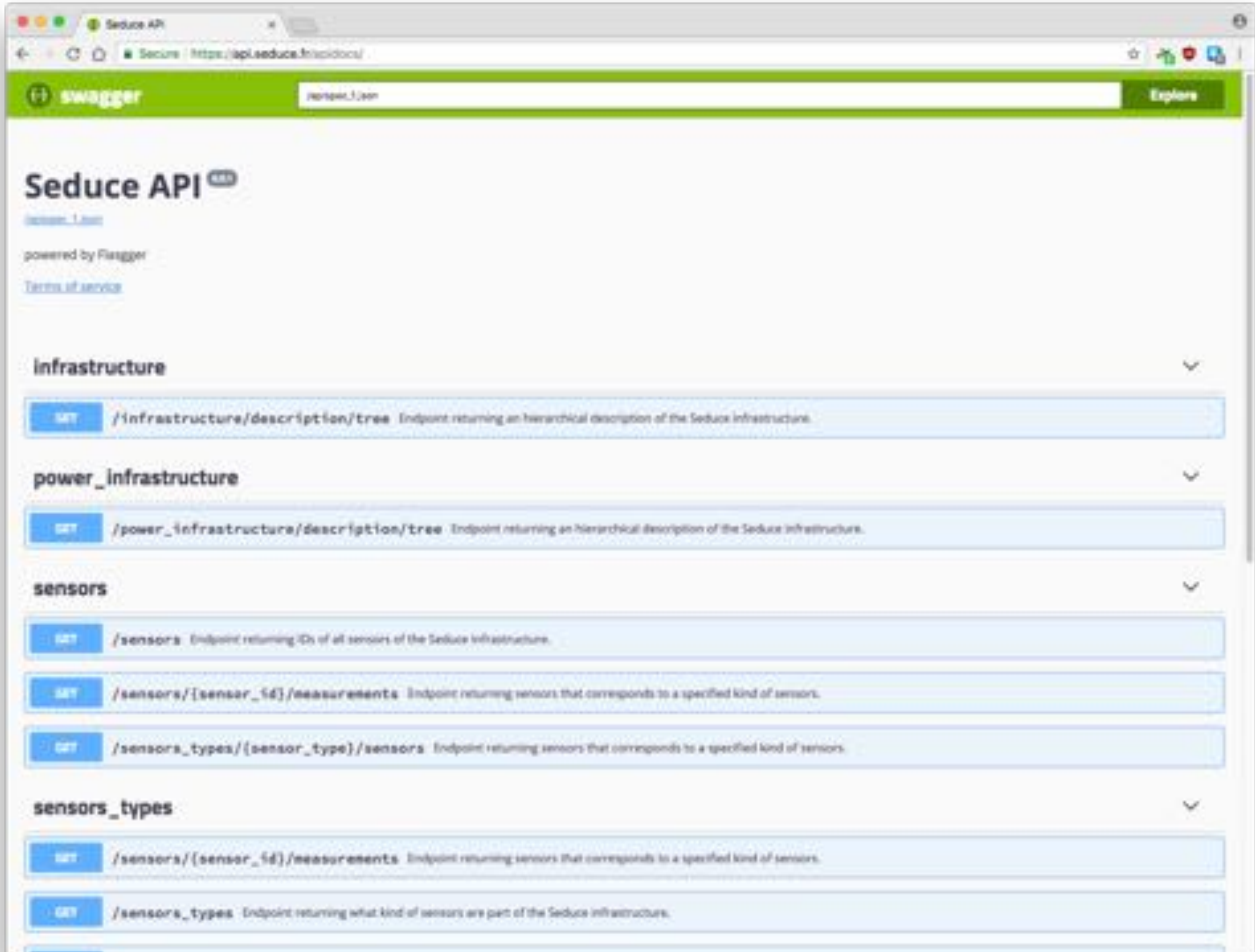
seduce.fr


Tree Chart - D3js
Secure
https://seduce.fr/weighted_tree_consumption
Zoom 1m 3m 6m YTD All
From Jun 6, 2018 To Jun 6, 2018
datacenter: 7 710.5445488872181
cooling_cluster: 4 211.4642857142886
servers: 3 180.8697368421053
datacenter
servers
cooling_cluster


Tree Chart - D3js
Secure
https://seduce.fr/weighted_tree_consumption
Zoom 1m 3m 6m YTD All
From Jun 6, 2018 To Jun 6, 2018
datacenter
servers
cooling_cluster

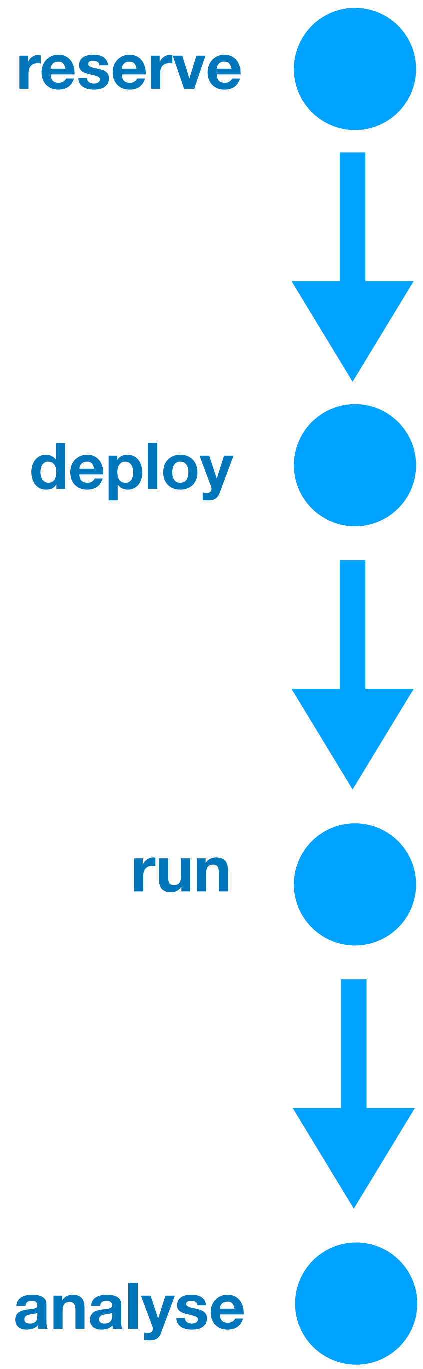
api.seduce.fr

<u>api.seduce.fr</u>



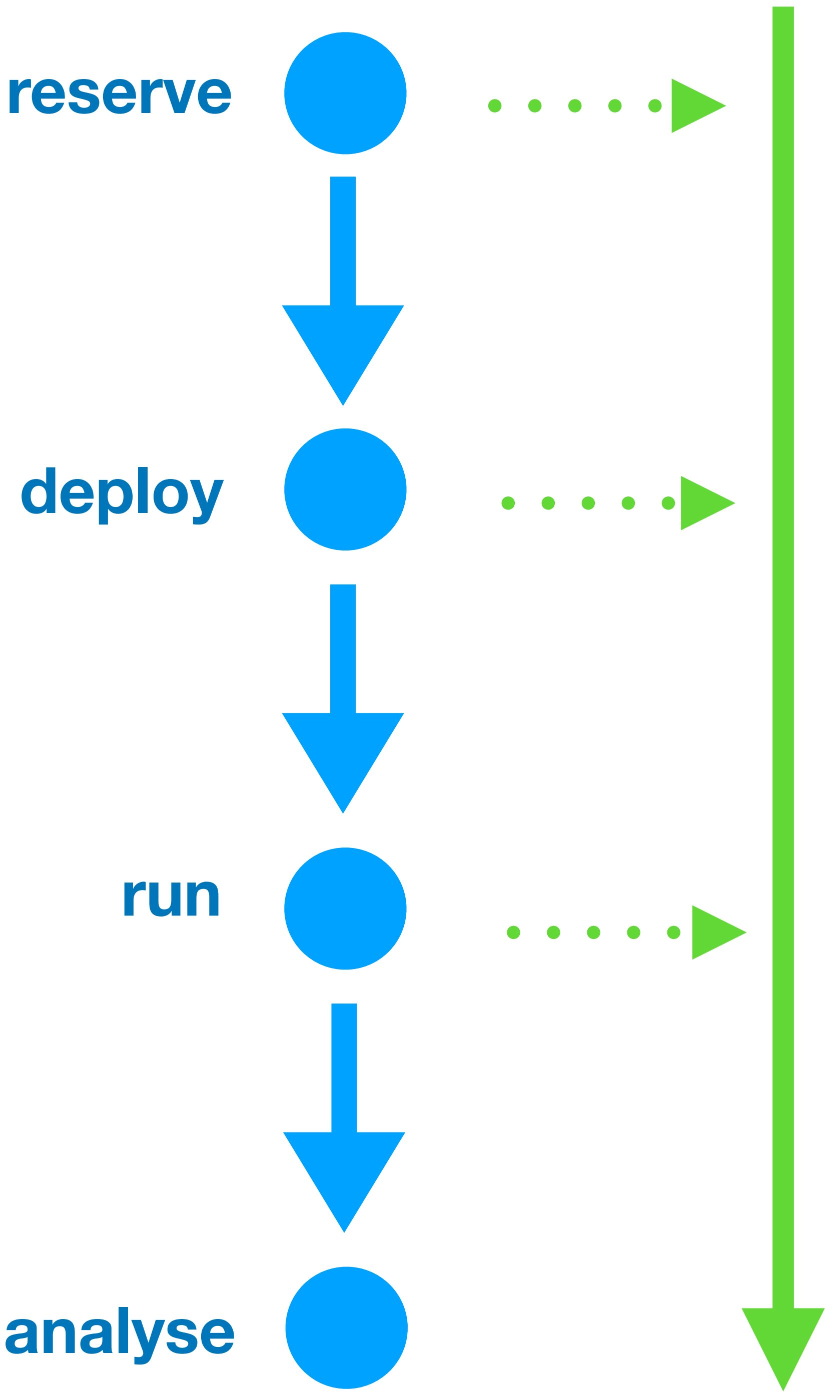
Experimental workflow

- User conduct an Grid’5000 experiment on the ecotype cluster
- In parallel of the experiment, energetic and thermal data become available on the Seduce platform
- It is possible to collect data of a specific time range after the experiment



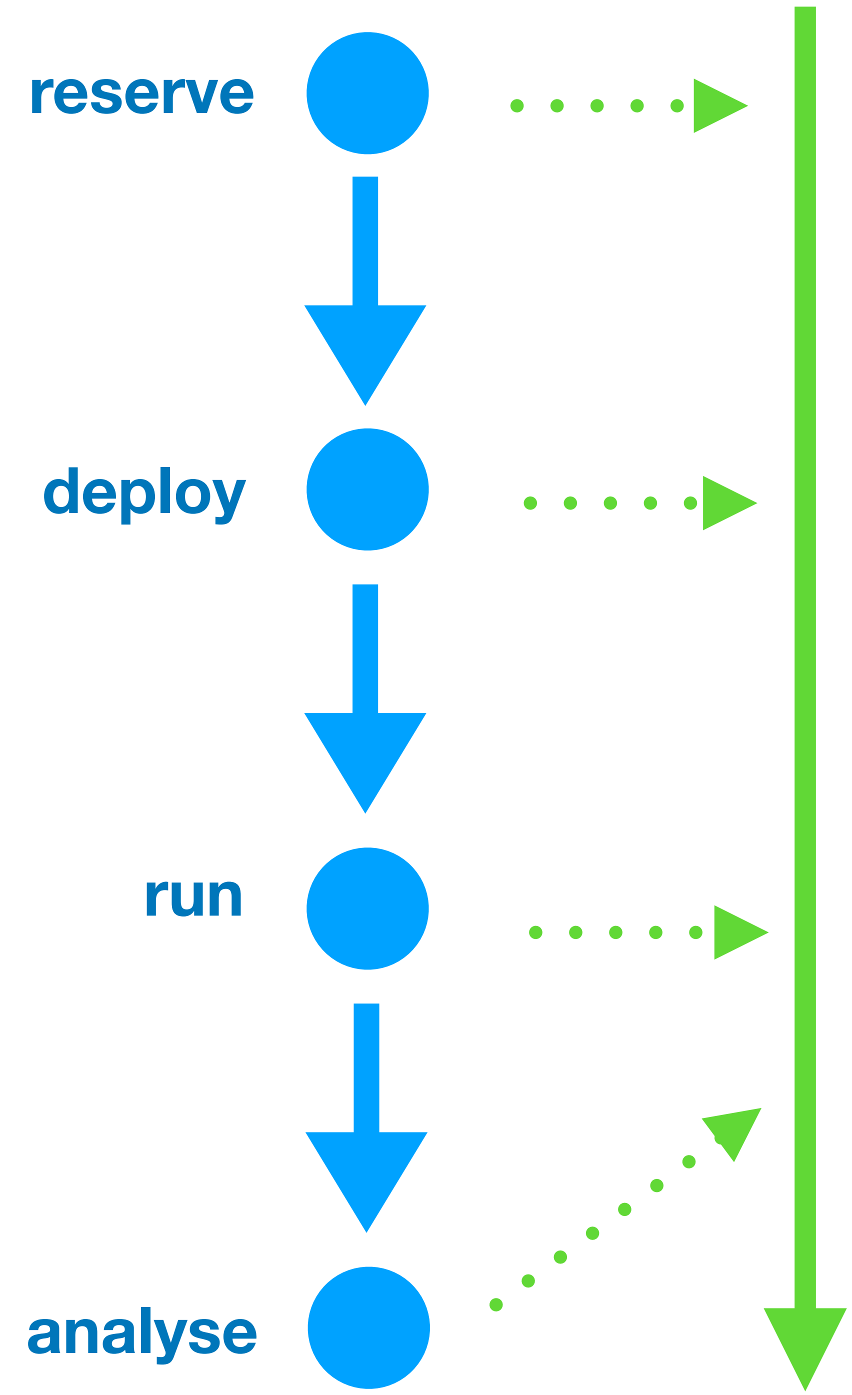
Experimental workflow

- User conduct an Grid'5000 experiment on the ecotype cluster
- In parallel of the experiment, energetic and thermal data become available on the Seduce platform
- It is possible to collect data of a specific time range after the experiment



Experimental workflow

- User conduct an Grid’5000 experiment on the ecotype cluster
- In parallel of the experiment, energetic and thermal data become available on the Seduce platform
- It is possible to collect data of a specific time range after the experiment



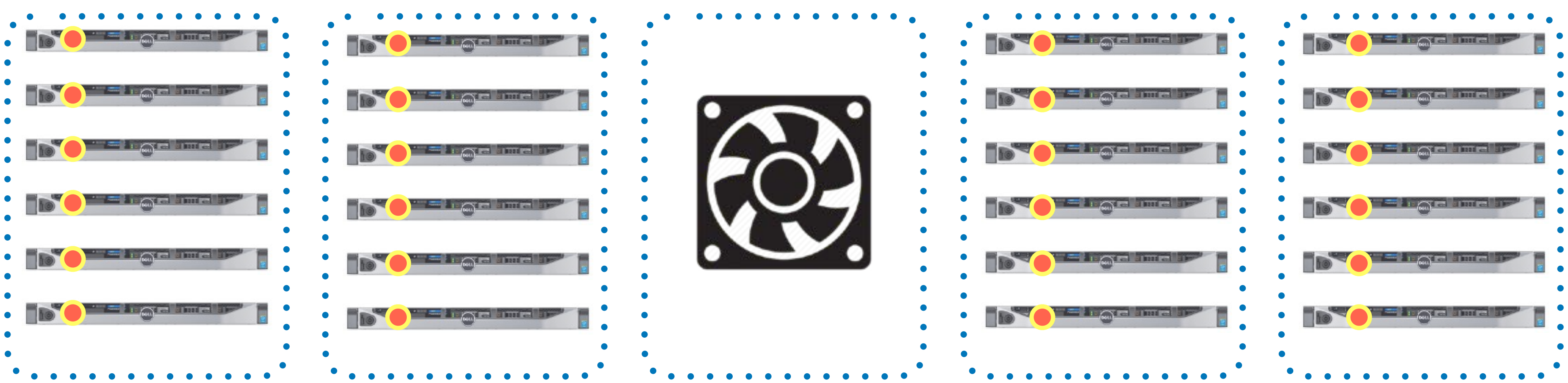
Experimentation example with the testbed

Understand the impact of idle servers

- Idle servers are active servers that don't execute any useful workload
- They consume energy
- They produce heat
- They don't contribute to the cluster
- Impact of idle servers has been studied in a third party publication **[6]**
- We would like to reproduce this observation with our data

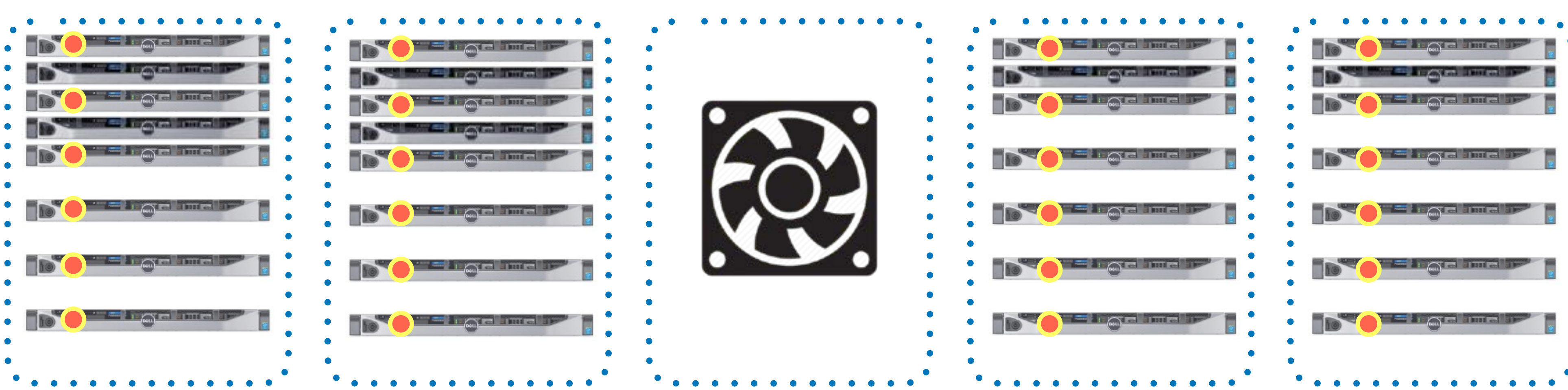
Protocol

- Servers are divided in 3 groups : active, idle, turned off servers
  - actives group : 24 servers
  - idle servers
  - turned off servers : remaining servers
- CPUs of all active servers are stressed
- During one hour, consumption of the CCS is recorded
- Iteratively, we set the number of idle servers to 0, 6, 12, 18, 24 servers
- Each experiment is repeated 5 times. Between 2 experiment, servers are shut down until the temperature is back to 26°C.



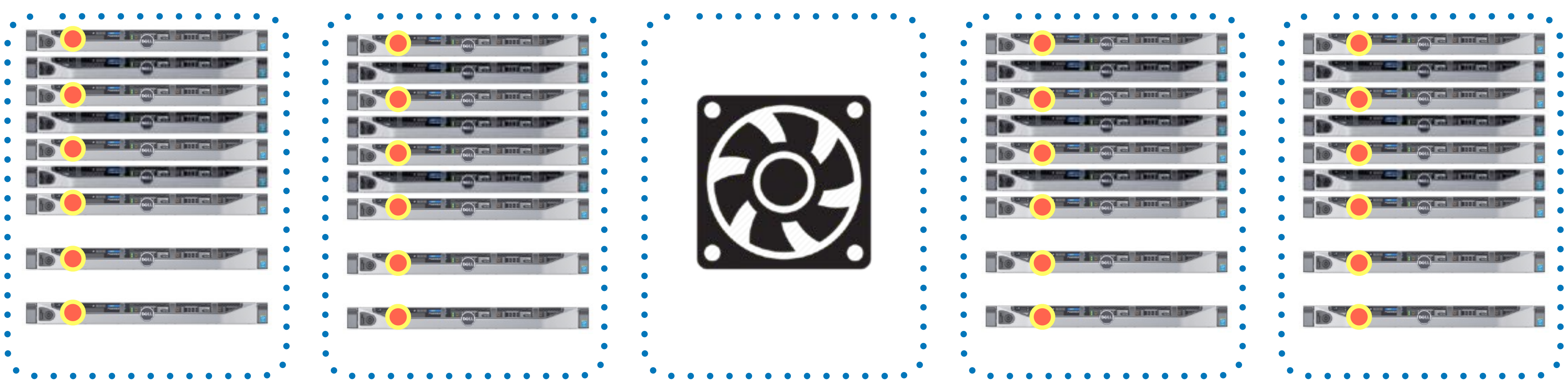
Protocol

- Servers are divided in 3 groups : active, idle, turned off servers
  - actives group : 24 servers
  - idle servers
  - turned off servers : remaining servers
- CPUs of all active servers are stressed
- During one hour, consumption of the CCS is recorded
- Iteratively, we set the number of idle servers to 0, 6, 12, 18, 24 servers
- Each experiment is repeated 5 times. Between 2 experiment, servers are shut down until the temperature is back to 26°C.



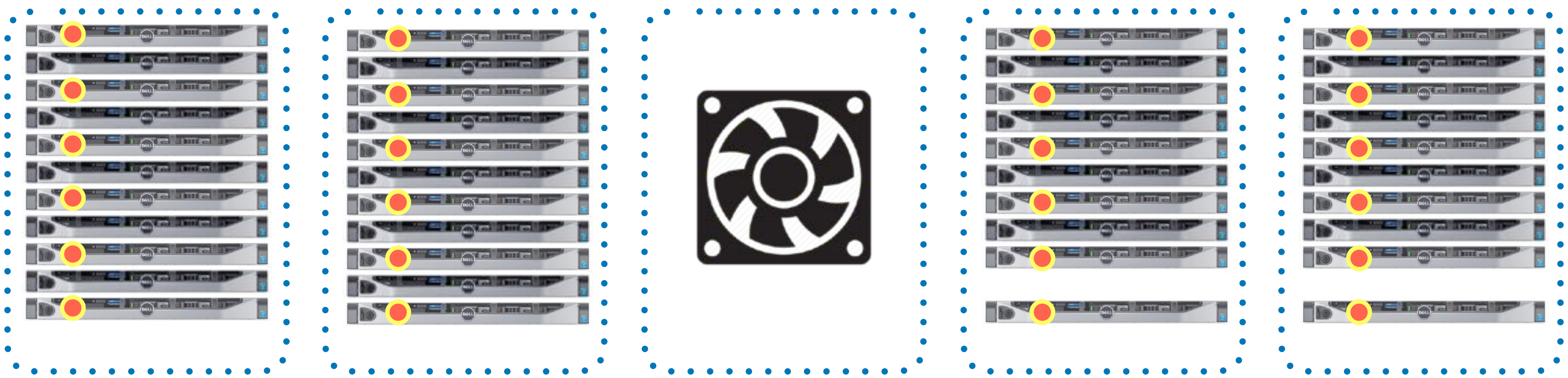
Protocol

- Servers are divided in 3 groups : active, idle, turned off servers
  - actives group : 24 servers
  - idle servers
  - turned off servers : remaining servers
- CPUs of all active servers are stressed
- During one hour, consumption of the CCS is recorded
- Iteratively, we set the number of idle servers to 0, 6, 12, 18, 24 servers
- Each experiment is repeated 5 times. Between 2 experiment, servers are shut down until the temperature is back to 26°C.



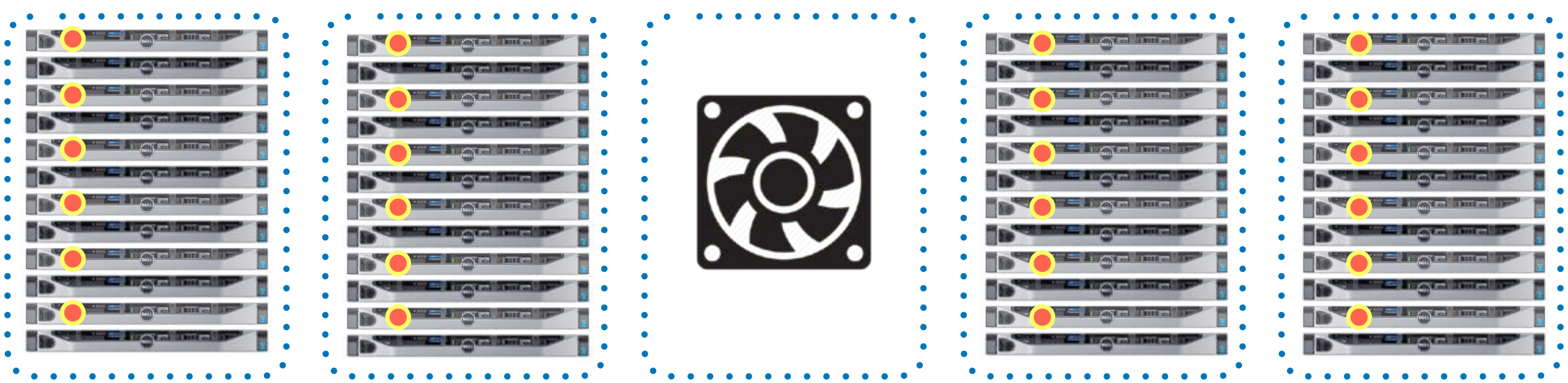
Protocol

- Servers are divided in 3 groups : active, idle, turned off servers
    - actives group : 24 servers
    - idle servers
    - turned off servers : remaining servers
- CPUs of all active servers are stressed
- During one hour, consumption of the CCS is recorded
- Iteratively, we set the number of idle servers to 0, 6, 12, 18, 24 servers
- Each experiment is repeated 5 times. Between 2 experiment, servers are shut down until the temperature is back to 26°C.



Protocol

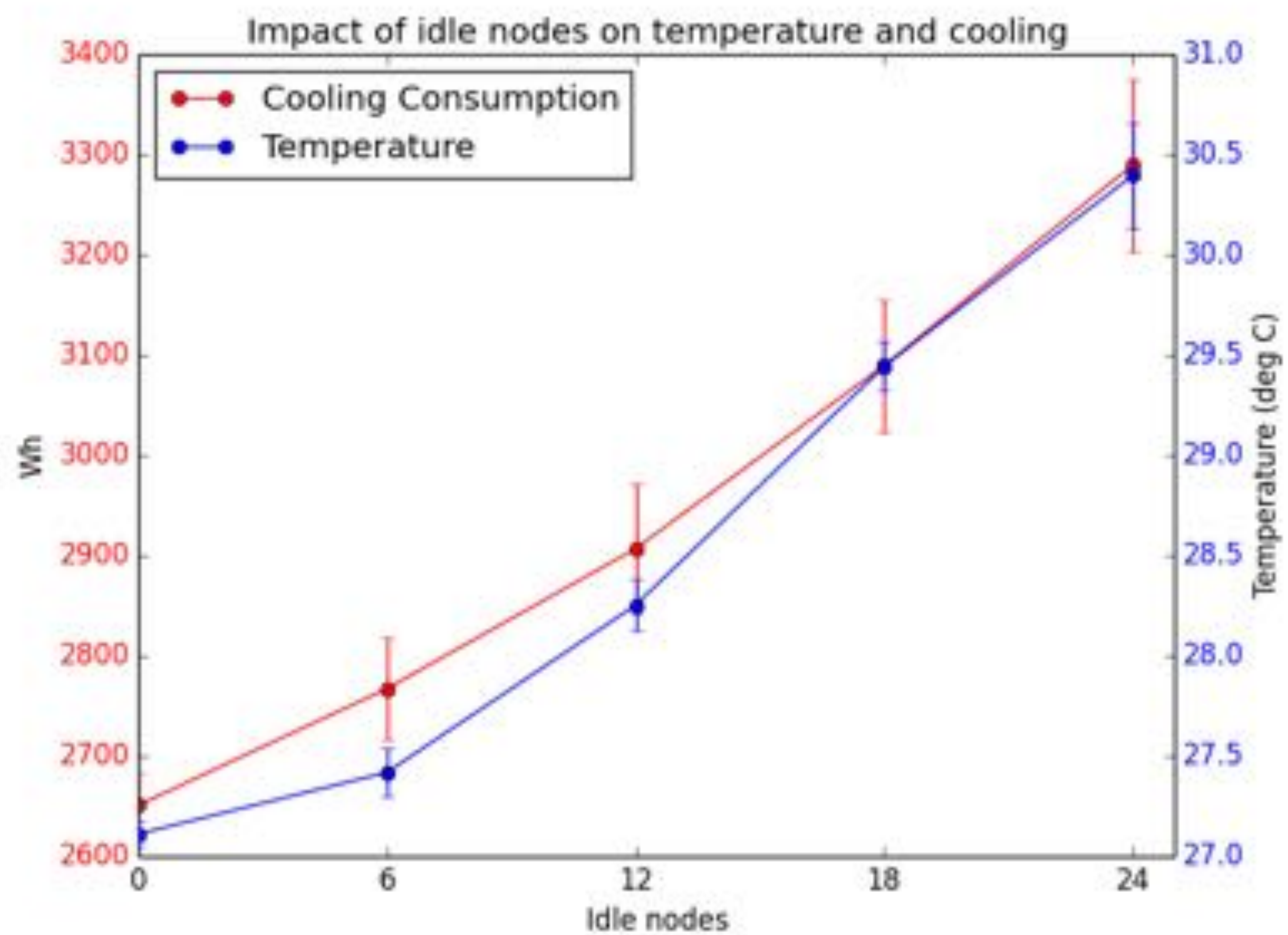
- Servers are divided in 3 groups : active, idle, turned off servers
    - actives group : 24 servers
    - idle servers
    - turned off servers : remaining servers
- CPUs of all active servers are stressed
- During one hour, consumption of the CCS is recorded
- Iteratively, we set the number of idle servers to 0, 6, 12, 18, 24 servers
- Each experiment is repeated 5 times. Between 2 experiment, servers are shut down until the temperature is back to 26°C.

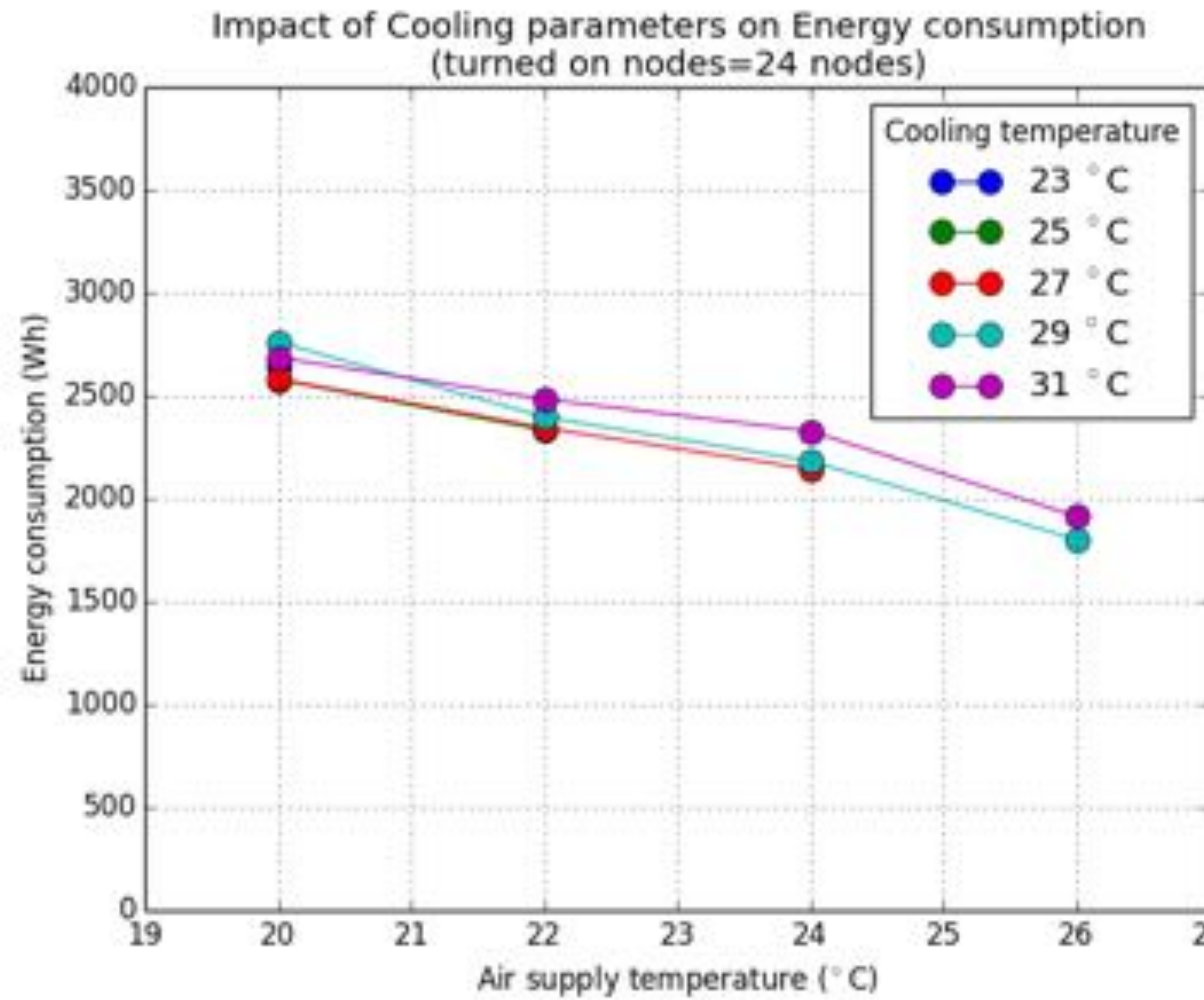


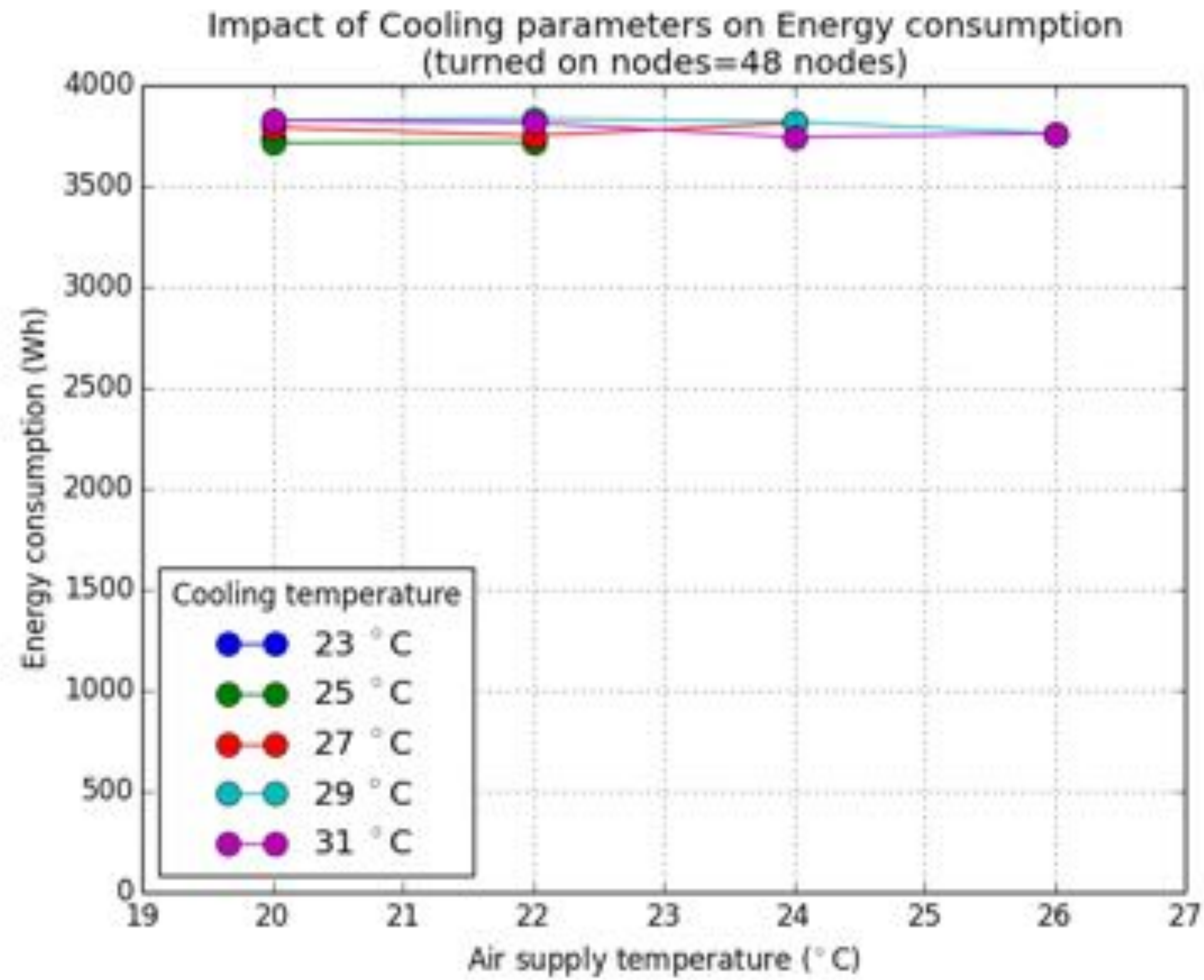
Results

- The number of idle nodes has an impact on the temperature in the hot aisle
- High density of active servers ➔ <u>Hot spots</u>
- Sensors of the CCS detect the hot spots
- CCS has to maintain a temperature target
- Max consumption of the CCS is ~3300Wh

SCS enabled




Impact of Cooling parameters on Energy consumption
(turned on nodes=24 nodes)
Cooling temperature
23 °C
25 °C
27 °C
29 °C
31 °C
Energy consumption (Wh)
4000
3500
3000
2500
2000
1500
1000
500
0
19
20
21
22
23
24
25
26
27
Air supply temperature (°C)



Impact of Cooling parameters on Energy consumption
(turned on nodes=48 nodes)
Energy consumption (Wh)
Air supply temperature (°C)
Cooling temperature
23 °C
25 °C
27 °C
29 °C
31 °C
0
500
1000
1500
2000
2500
3000
3500
4000
19
20
21
22
23
24
25
26
27

Analysis

- Integration with Grid'5000 : experimental workflow mainly rely on tools such as *kadeploy*, *kapower* and *execo*

- Meanwhile, we collected from the API the power consumption and the temperature of elements of the testbed

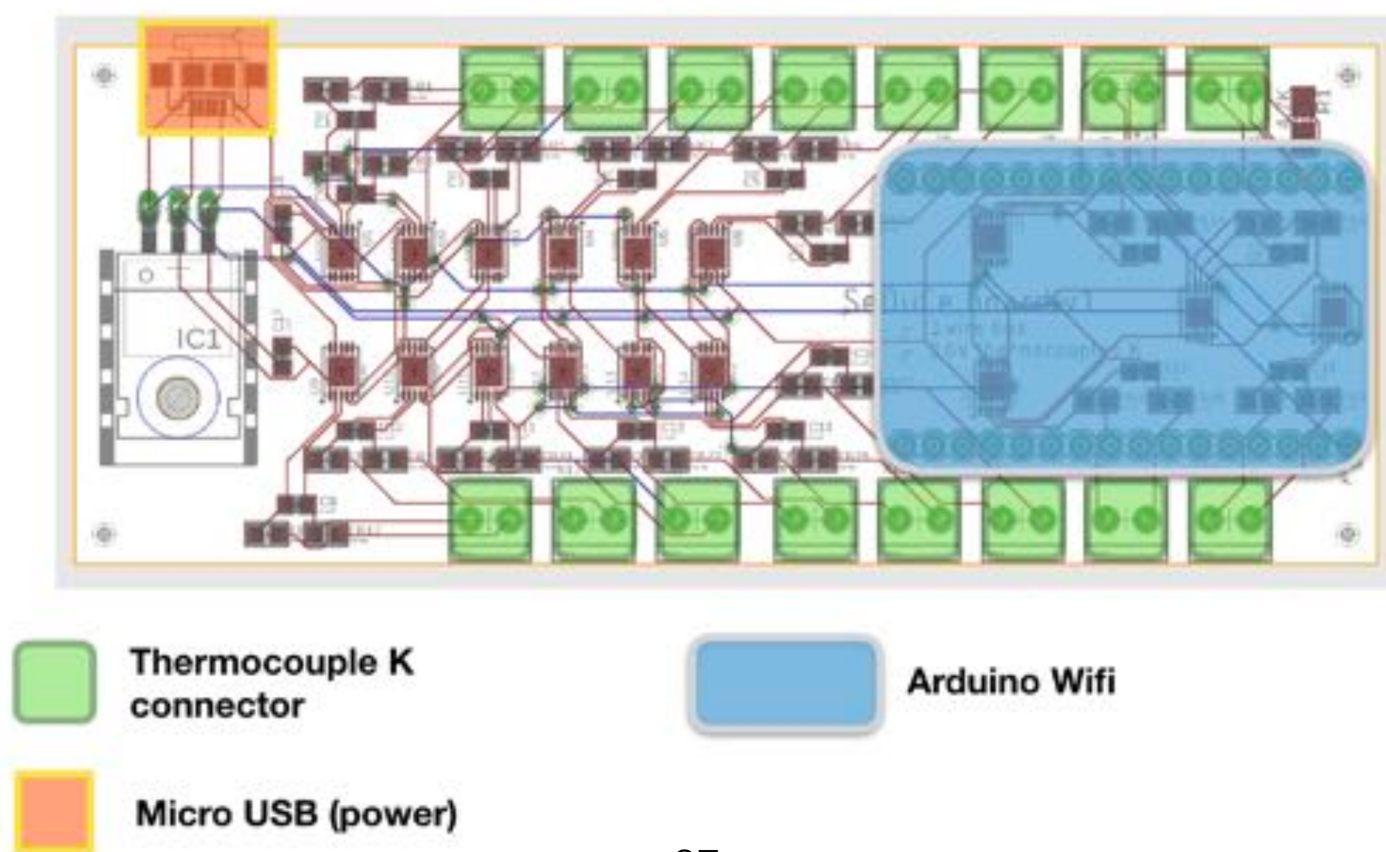
- Thus, the SeDuCe testbed enables to easily perform experiments mixing temperature and energy aspects of a datacenter

- The data collected by SeDuCe sensors seems to be relevant

Future work

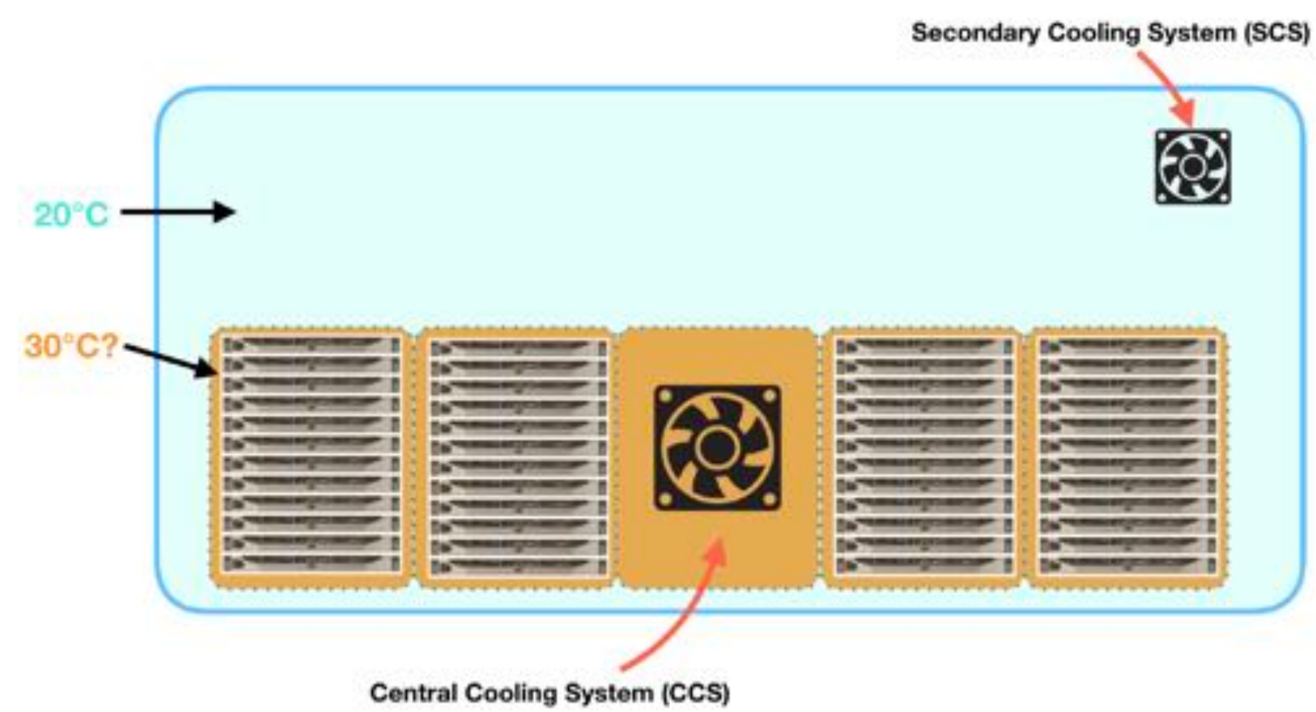
Better temperature sensors

- Collaboration with the Energy Department at IMT Atlantique
- We have designed an electronic card that embeds 16 thermocouples, (*using the <u>AutoDesk Eagle</u> CAD software*)
- We are planning to install these cards within September 2018



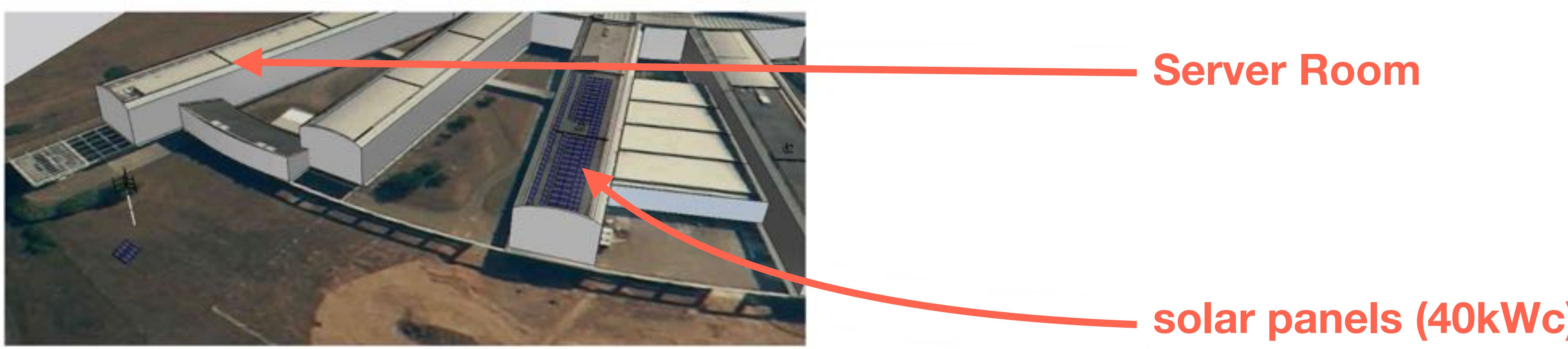
Provide full control on cooling settings

- We plan to let users decide the temperature of the testbed during their experiments
- Few questions need to be answered (ensure that only the user that is experimenting can change cooling parameters, security, prevent incident)



Renewable energies

- In summer 2018, solar panels and batteries will be included to the testbed.
- We will discuss with the manufacturer mid July *to understand what we can do with the solar panels*
- *Ideally*, we plan to let users build experiment where they can decide what to do : use energy, store in battery.
- Enable a wide range of research such as placement policies that takes into account renewable energy.



Conclusion

Conclusion

- SeDuCe is a testbed that enable research activities that mix both thermal and power management in datacenters
- It proposes “Ecotype”, a new Grid’5000 cluster composed of 48 servers
- The temperature and the power consumption of equipments of the testbed are monitored and made available to the testbed’s users
- Future work consists in improving the quality of the temperature sensors and including the renewable energies aspect in the testbed

Questions?

jonathan.pastor@inria.fr

seduce.fr

References

- **[1]** Awada, Uchechukwu & Li, Keqiu & Shen, Yanming. (2014). Energy Consumption in Cloud Computing Data Centers. International Journal of Cloud Computing and Services Science (IJ-CLOSER). 3. . 10.11591/closer.v3i3.6346.

- **[2]** Albert Greenberg, James Hamilton, David A. Maltz, and Parveen Patel. 2008. The cost of a cloud: research problems in data center networks. *SIGCOMM Comput. Commun. Rev.* 39, 1 (December 2008), 68-73. DOI=http://dx.doi.org/10.1145/1496091.1496103

- **[3]** http://www.hardware.fr/articles/965-4/consommation-efficacite-energetique.html

- **[4]** https://www.cs.rutgers.edu/content/parasol-rutgerss-green-udatacenter

- **[5]** https://www.qarnot.com/qrad/

- **[6]** Justin D Moore, Jeffrey S Chase, Parthasarathy Ranganathan, and Ratnesh K Sharma. 2005. Making Scheduling" Cool": Temperature-Aware Workload Place- ment in Data Centers.. In USENIX annual technical conference, General Track. 61–75.