Do-it yourself

traditional DC

Romaric David
david@unistra.fr
University of Strasbourg and RENATER
September 11th, 2014

UNIVERSITÉ DE STRASBOURG

RENATER
CONNECTING KNOWLEDGE

Plan

- Introduction
- Airflow control
- Monitoring
- PUE
- Conclusion

Introduction (1/3)

- HPC Center of University of Strasbourg (France)
- http://hpc.unistra.fr
- 300 HPC Servers, Linux, Infiniband Network, ...
- Cold Aisle and Hot Aisle confinment for high-density racks (35kw / rack)
- 28°C in front of the servers
- Free-Chilling
- Server room totally re-designed for HPC

Introduction (2/3)

Introduction (3/3)

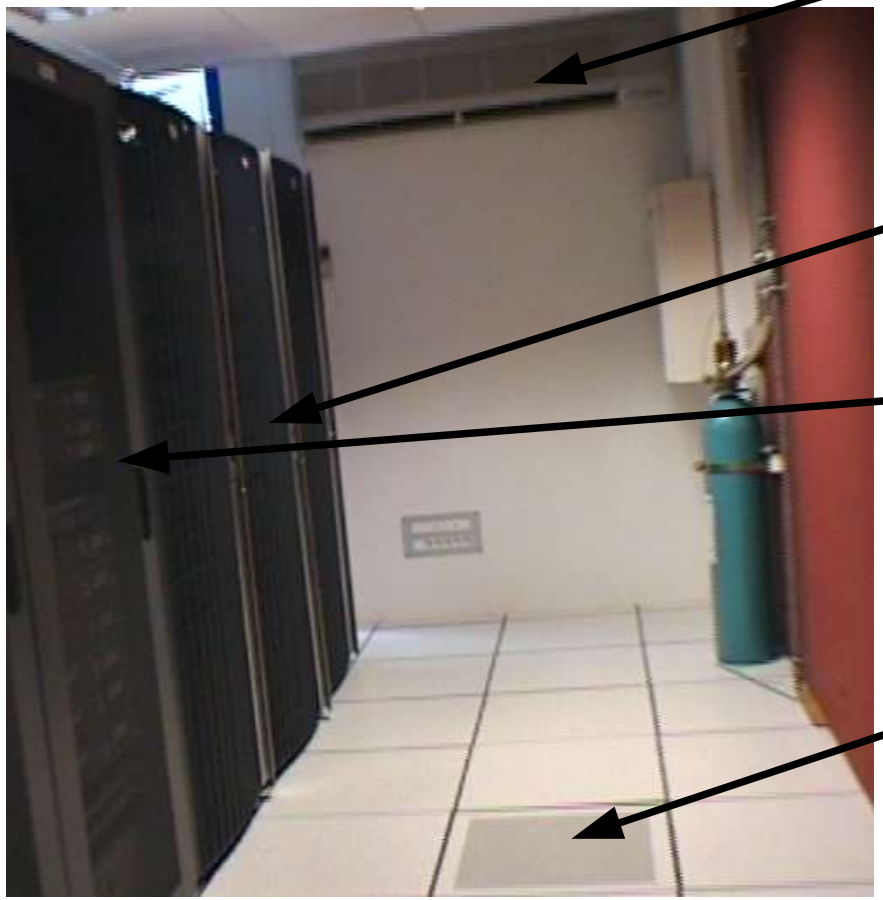
- This talk is about re-configuring an existing machine-room into an up-to-date Datacenter
- It is based on:
  - our experience in realizing datacenters
  - several visits of datacenters in France (thanks to the hosts)
- This talk focuses on practical improvements you can realize

Plan

- Introduction
- Airflow control
- Monitoring
- PUE
- Conclusion

Your old / current DC

- Let's bet your current DC looks like:



Airflow

- Most probably hot air and cold air mix everywhere in the room :

Here hot air comes back to front because of a missing filler

Courtesy Bernard Boutherin (LPSC – CNRS - France)

- Problem: hot air arriving on the front-side of the servers.
- Cooling more expensive (fan speed increases)
- Solution: a plastic filler (max several €)

Airflow: panels

- Anywhere in the room, you should control airflow in order to minimize the quantity of hot air recirculating in front of the servers without going through the cooling
- This is usally done by aisle confinement
- You can approximate aisle confinment by putting (metal) panels around your racks. α 100 €
- In progress in Peugeot, Montbeliard, France, based on a idea from Bernard Barrois

May convince your security staff that this is fire-safe !

Airflow: plastic curtains

- To create low-cost hot / colds aisles in an existing room, you can as well:
  - add plexigas roofs across racks;
  - on both sides, close the aisles using plastic curtains.
- Next slide: server room of the Institut Pierre et Simon Laplace, Paris (thanks to Franck Corsini):

Airflow: plastic curtains

- Cost: 5300€ for the metal frame and the curtains, for 9 racks (600 € / rack)
- Company: ODC (Optimal Data Center)

Airfrlow: what about free-cooling ?

- In order to be able to go for free-cooling, you need:
  - a source of fresh air **from** the outside;
  - an air output, **to** the outside;
  - on the outside, input and output should not be too close one from the other (several meters)
  - depending on the fresh air temperature, you may want to mix hot air and cold air in order to heat the cold air (winter....)
- A reminder: Ashrae A2 (2011, 3 years ago...) allows 35°C 100% of the year in front of the servers

Airfrlow: what about free-cooling ?

- On 1st try, you can only try and collect and get rid of the hot air;
- You can put an air-extractor on the back-side of your racks
- Done at Observatoire de Paris (thanks to Emmanuel Halbwachs):
- Based on:
  http://www.apc.com/products/resource/include/techspec_index.cfm?base_sku=ACF400

Airfrlow: what about free-cooling ?

- Cost: around 2500€ for the air-extractor
- Should be able to extract 17kw of hot air
- Input air comes from the computer room (leave door open !)

Airfrlow: what about free-cooling ?

- You can then try and inject air from the outside
- Made by the Ciment team in Grenoble (Françoise Roch, Olivier Richard, Christian Seguy, Pierre Neyron)
- https://2011.jres.org/archives/90/paper90_poster.pdf
- http://wiki-oar.imag.fr/index.php/FrigID_construction
- For the box around the servers, they used a shower cabin bought in a general-purpose store !

Airfrlow: what about free-cooling ?

Professional fan

- Cost: 3000€ (HPC machine was 75 k€)
- Watch out for insulation!
- Observed front-side temperature: 33°C max

Airfrlow: what about free-cooling ?

- The previous installation has no heat exchanger. You can add one by upgrading the fan to a real exchanger
- Made in a bigger installation by **Bernard Boutherin** (CNRS, LPSC, Grenoble) **since 2008**
- http://informatique.in2p3.fr/?q=node/290
- Cost: 60k€ for 80kw IT. Refund in less than 5 years
- Very few air-condition failures
- Free-Cooling used 84% of time

Airfrlow: what about free-cooling ?

- Pictures:

Heat exchanger

Motorized registers for air in/out/mix

Filters (2 stages, coarse and fine grain)

Plan

- Introduction
- Airflow control
- Monitoring
- PUE
- Conclusion

Monitoring

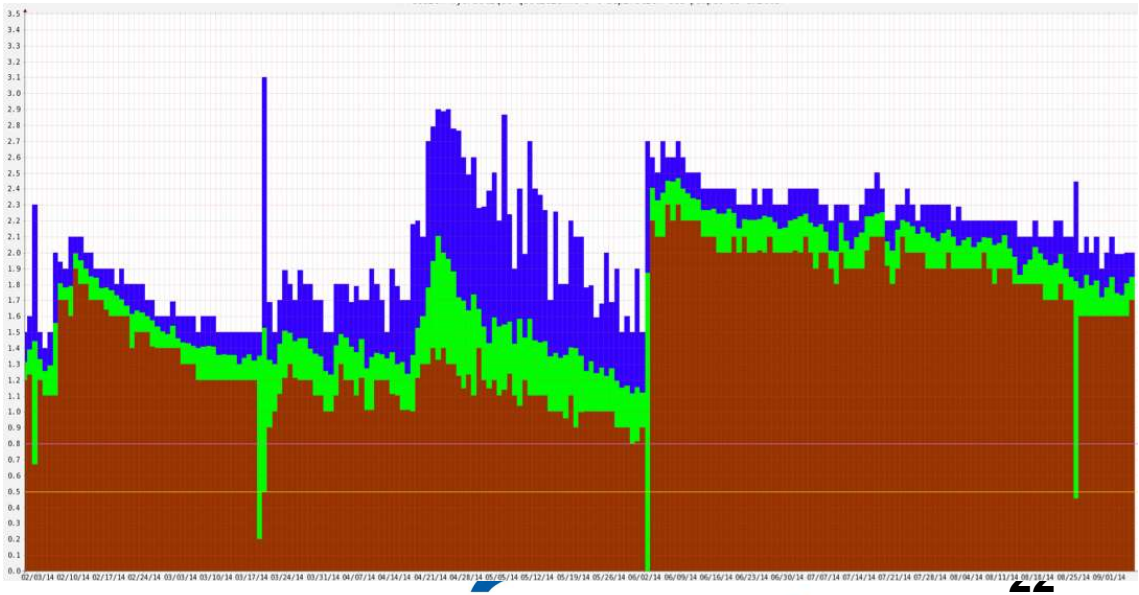
- Some Datacenter vendors offer complete all-in-one monotoring solutions
- You can get good approximate using standard tools, namely SNMP, IPMI, Ganglia
- Monitoring is about extracting environmental data from your Datacenter:
  - Power consumption
  - Temperatures
  - Pressures
  - ...

Monitoring

- When it comes to power consumption, you can use a monitored / managed PDU (≈ 30 € per plug)
- For instance:
  http://www.apc.com/products/resource/include/techspec_index.cfm?base_sku=AP7900
- Ask for the MIB!
- Optionnal temperature sensors available
- IPMI allows monitoring for free, too (temperature, wattage)

Monitoring – Out of IT

- Even if you bought a modern cooling system, you may not use all of its (hidden) features
- You can probably query your system via snmp, and retrieve very interresing environmental values : pressures, water temperatures
- Here is what we get: (average pressures, for diagnostic purposes)
- Think about parsing web interfaces ouput, too



Monitoring

- You can retrieve a lot of useful information for free, using very simple tools and queries
- Several software are available in order to gather data
- You can easily build your own data-center management software!

Plan

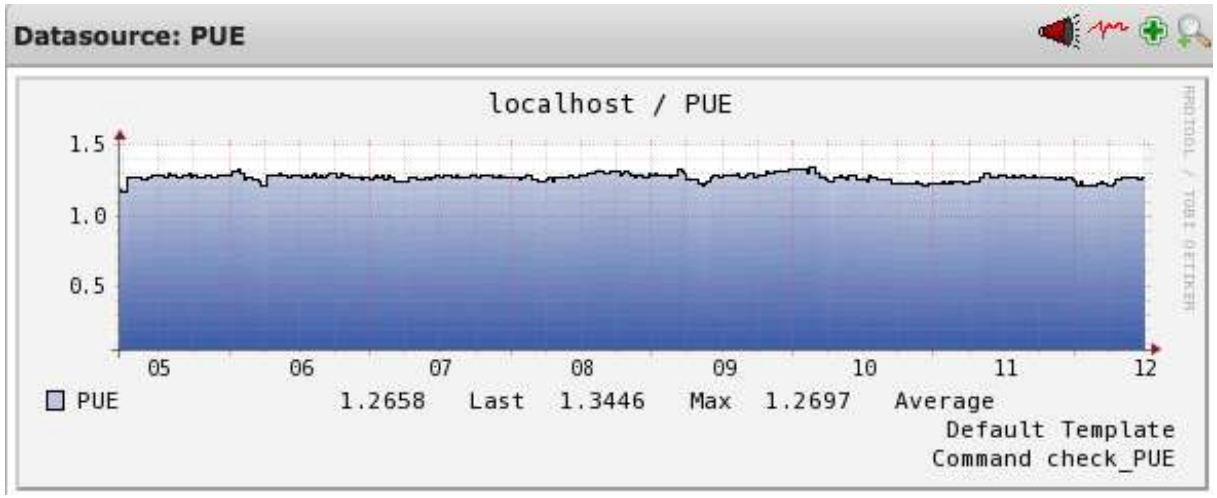
- Introduction
- Airflow control
- Monitoring
- PUE
- Conclusion

Power Usage Effectiveness

- In order to compute PUE, you need to measure:
  - Total IT Power, either wattage (W) or overal consumption (kWh)
  - Total Facilities Power
- We now have enough data to get an estimate of the PUE
- If one data is missing, you can try and deduce it:
  - In Strasbourg, we measure the total Datacenter power. Data source: CTM / modbus
  - We substract from it the sum of the cooling system. Data source: Snmp on the chiller / exchangers

Power Usage Effectiveness

- PUE can then be integrated into Ganglia / Nagios



- Don't hesitate to communicate on the PUE

Plan

- Introduction
- Airflow control
- Monitoring
- PUE
- Conclusion

Conclusion

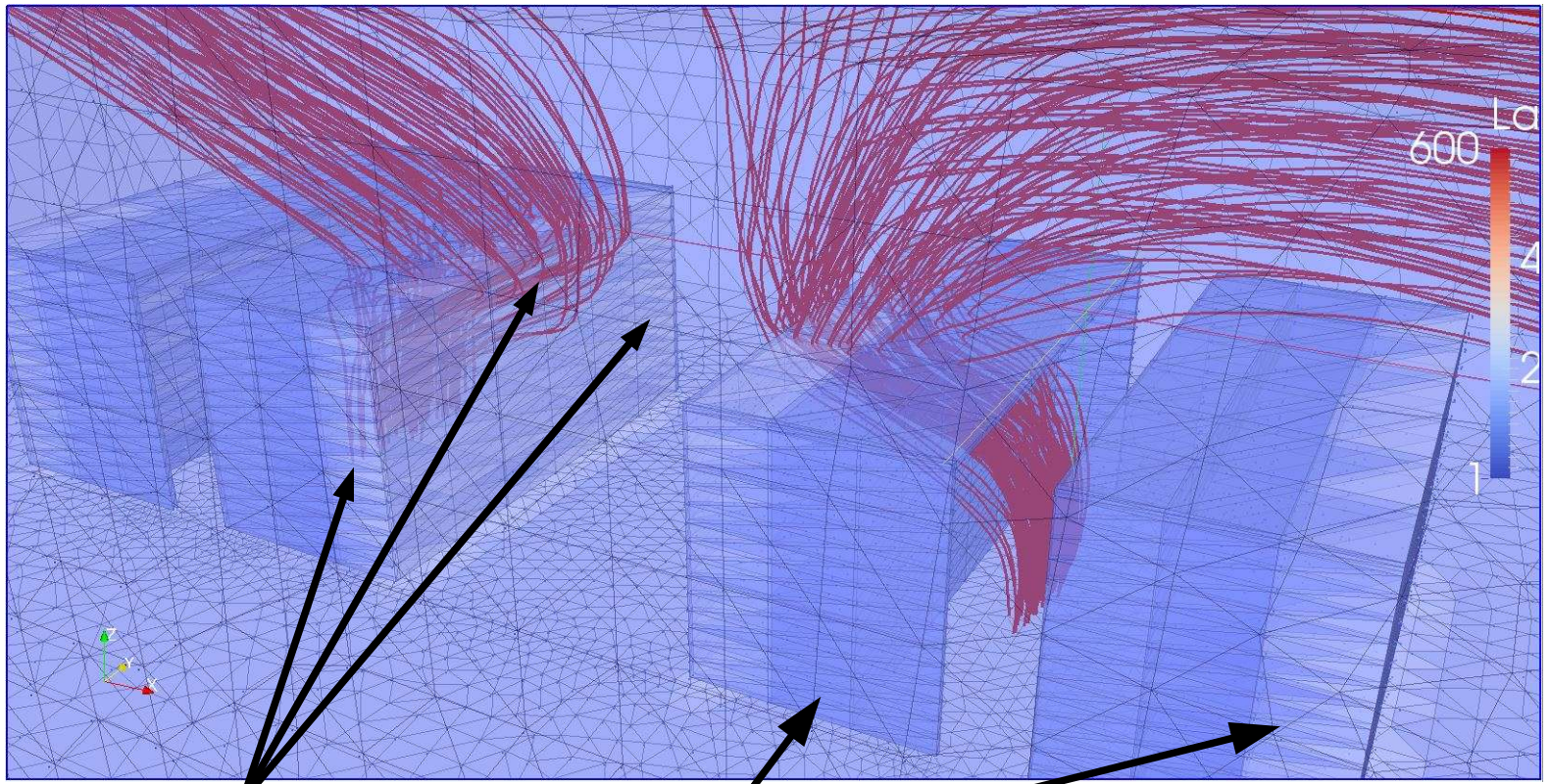
- There are many solutions in order to optimize your Datacenter
- Airflow is the 1st thing to look at
- A bit of time and small amounts of money can really help you transform your DC
- Monitoring will help you to measure the effects of your trial-and error steps
- Keep in mind to give your ideas a try in non-critical rooms (typically : HPC rooms)
- You may be the one who inovate !

Plan

- Introduction
- Airflow control
- Monitoring
- PUE
- Conclusion
- Additional slides

Airflow: panels

- Here is the simulation of the airflow in a machine room with panels on the top and on both sides

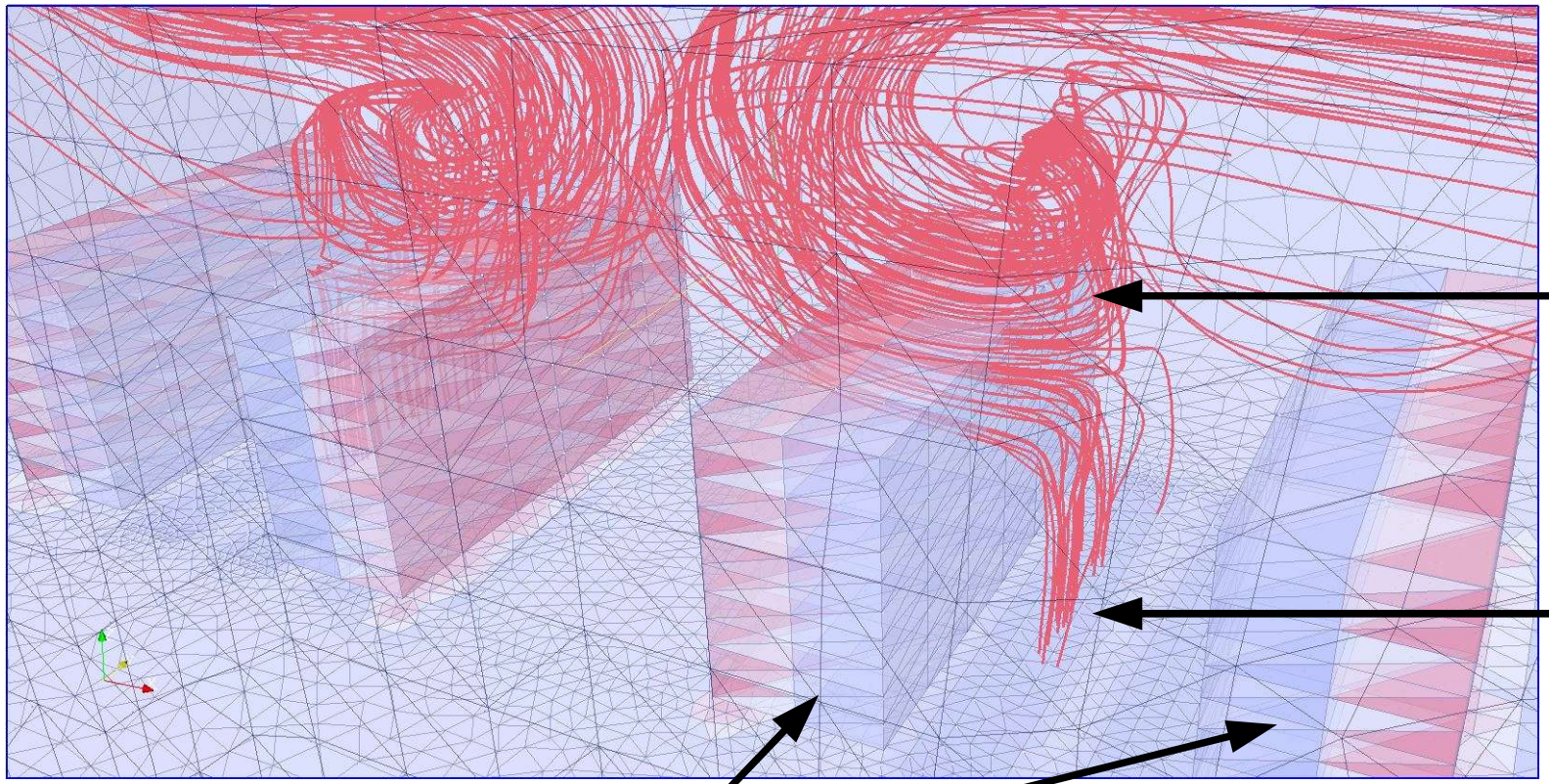


3 panels ($\prod$) on the rear sides of racks

Set of racks

Airflow : panels

- Here is the simulation of the airflow in a machine room without panels on the top and on both sides

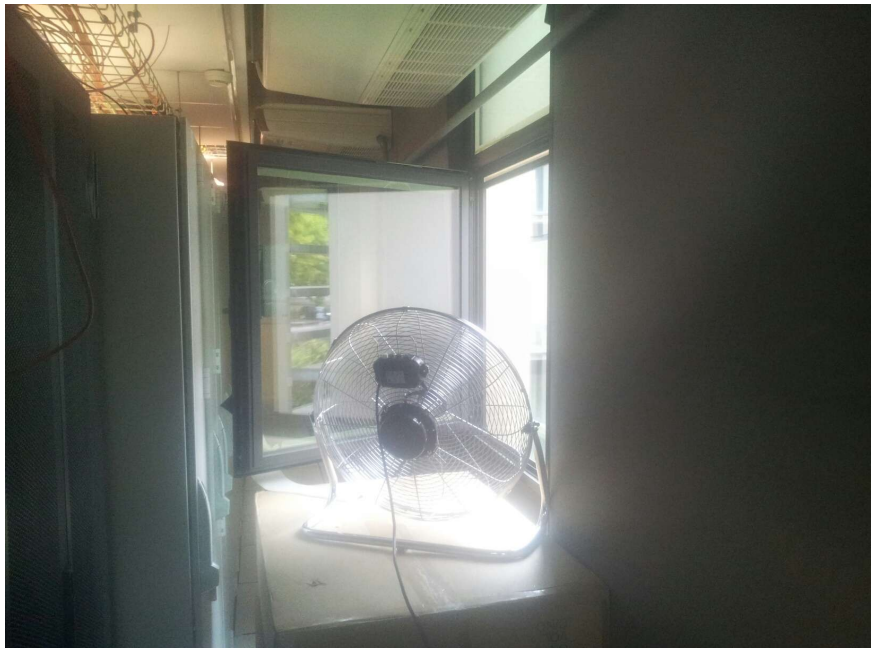


Airflow: panels

- This simulator was written in a project involving:
  - PSA (French Automotive Industry), Bernard Barrois
  - GdS Ecoinfo, Romaric David, and internship of Mohcine El BarhBarh
  - University of Strasbourg/Icube: Yannick Hoarau
- Cost for the panels: several hundred €, and a few screws
- Customized metal panels available on several websites

Airfrlow: what about free-cooling ?

- Maintenance free-cooling:
- Airflow helped by fans (server and additional)



UNIVERSITÉ DE STRASBOURG

RENATER
CONNECTING KNOWLEDGE

Airfrlow: what about free-cooling ?

- Block Diagram:

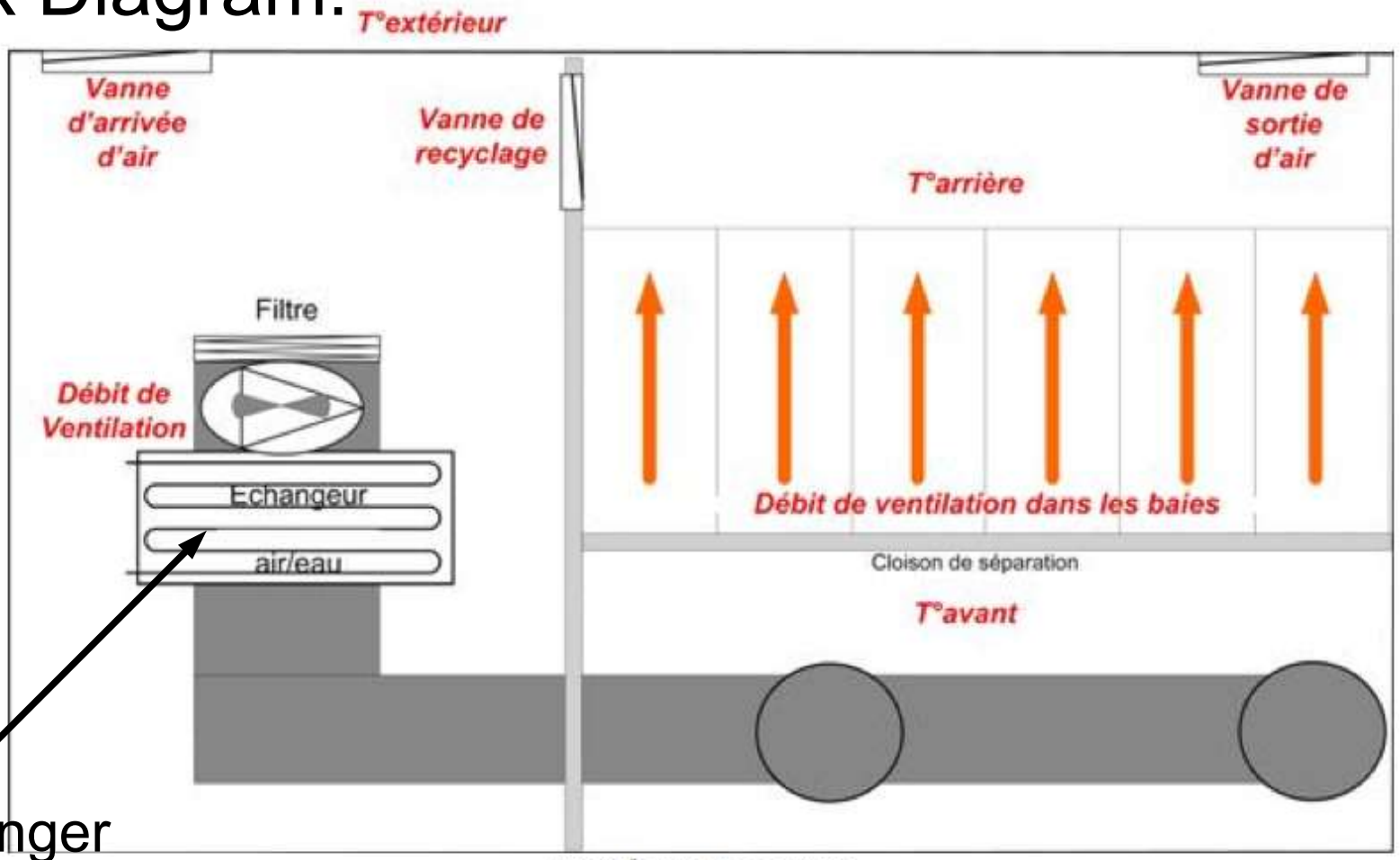


Schéma Ecoclim

Heat exchanger (cold water from an existing network)

Monitoring - Temperature

- IPMI example: a server from 2008 (HP DL165 G5)

```
> ipmitool -I lan -H [host] -U [user] -f [file containing IPMI user
   password]  sdr type Temperature
CPU0 Diode       | 20h | ok  |  3.0 | 48.50 degrees C
CPU1 Diode       | 22h | ok  |  3.1 | 51 degrees C
Power Ambient    | 24h | ok  |  0.0 | 16 degrees C
CPU0 Prochot     | 07h | ok  |  3.0 | Limit Not Exceeded
CPU1 Prochot     | 08h | ok  |  3.1 | Limit Not Exceeded
```

Monitoring - Temperature

- IPMI example: a server from 2013 (Supermicro)

```
> ipmitool -I lan -H [host] -U [user] -f [file containing IPMI user
   password]  sdr type Temperature
CPU1 Temp         | 01h | ok  |  3.1 | 68 degrees C
CPU2 Temp         | 02h | ok  |  3.2 | 62 degrees C
System Temp       | 11h | ok  |  7.1 | 35 degrees C
Peripheral Temp   | 12h | ok  |  7.2 | 56 degrees C
IB Temp           | 13h | ok  |  7.5 | 55 degrees C
PCH Temp          | 0Ah | ok  |  7.3 | 57 degrees C
P1-DIMMA1 TEMP    | B0h | ok  | 32.64 | 40 degrees C
P1-DIMMB1 TEMP    | B4h | ok  | 32.65 | 39 degrees C
P1-DIMMC1 TEMP    | B8h | ok  | 32.68 | 41 degrees C
P1-DIMMD1 TEMP    | BCh | ok  | 32.69 | 40 degrees C
P2-DIMME1 TEMP    | D0h | ok  | 32.72 | 34 degrees C
P2-DIMMF1 TEMP    | D4h | ok  | 32.73 | 35 degrees C
P2-DIMMG1 TEMP    | D8h | ok  | 32.76 | 33 degrees C
P2-DIMMH1 TEMP    | DCh | ok  | 32.77 | 35 degrees C
```

Monitoring - Temperature

- Getting the temperature server by server can give you quite a good overview of the temperature of your room
- May need a bit of scripting to integrate into ganglia/nagios
- You may select the sensor closest from the server air inlet
- The same standard Ipmi command can be used to query servers from different vendors

Monitoring – Power consumption

- Some proprietary commands or extensions can retrieve the power consumption

Example from Supermicro

```
./SMCIPMITool 192.168.4.92 login password pminfo
[SlaveAddress = 78h] [Module 1]
 Item                          |                Value
 ----                          |                -----
 Input Voltage                 |              226.5 V
 Input Current                 |               2.17 A
 Main Output Voltage           |              12.07 V
 Main Output Current           |              34.75 A
 Main Output Power             |                416 W
 Input Power                   |                472 W
```

- To get the consumption of the whole datacenter, you need to sum server by server

Monitoring – Out of IT

- To get more accurate measurements of power consumption without a managed PDU, you may be able to query the electric distribution modules (do-it-yourself Centralized Technical Management)
- This is usually done through the modbus protocol
- You need the *modbus address* in order to query the device