OVHcloud

Cloud materiality
What mitigation strategy ?

GREGORY LEBOURG
GLOBAL ENVIRONMENTAL DIRECTOR

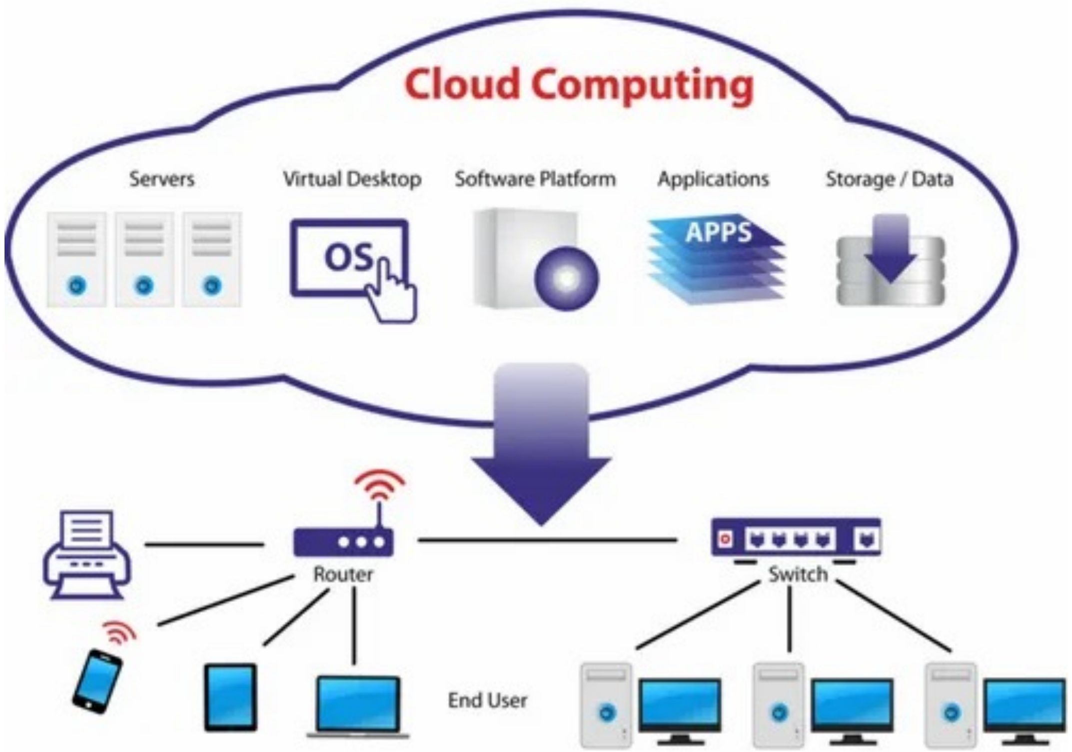
Eco-ICT 2024 autumn school
October 7, 2024

Agenda

01 Definition and scope covered

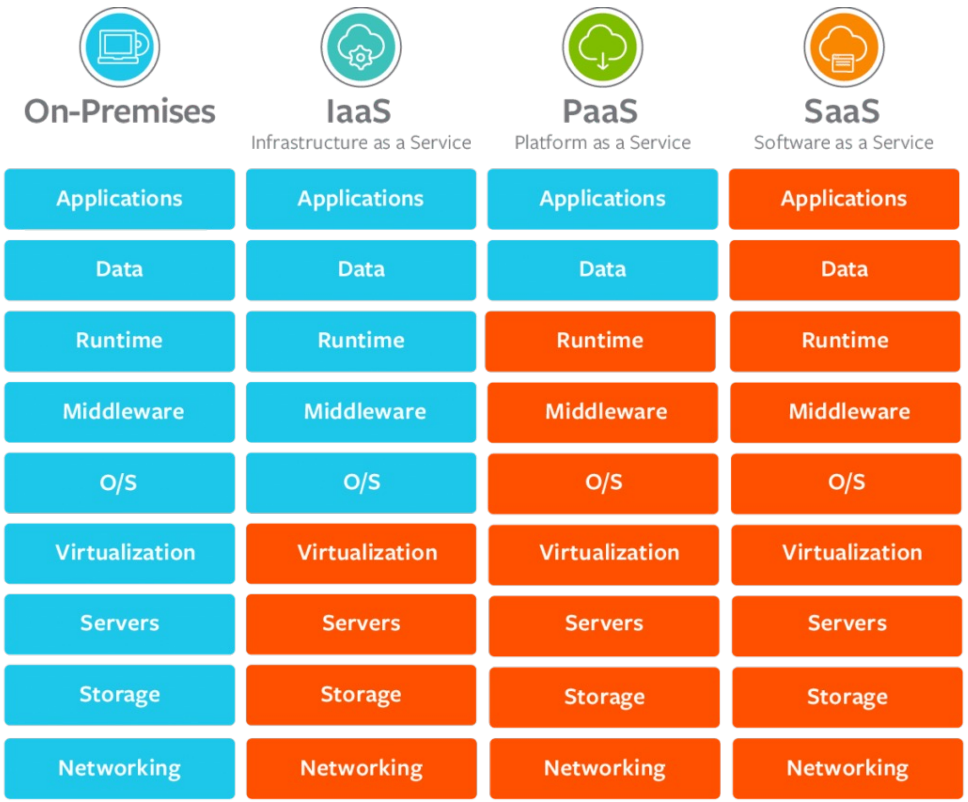
What is Cloud ?

Cloud refers to a system that provides access to IT services (storage, computing, software) through the internet or private networks

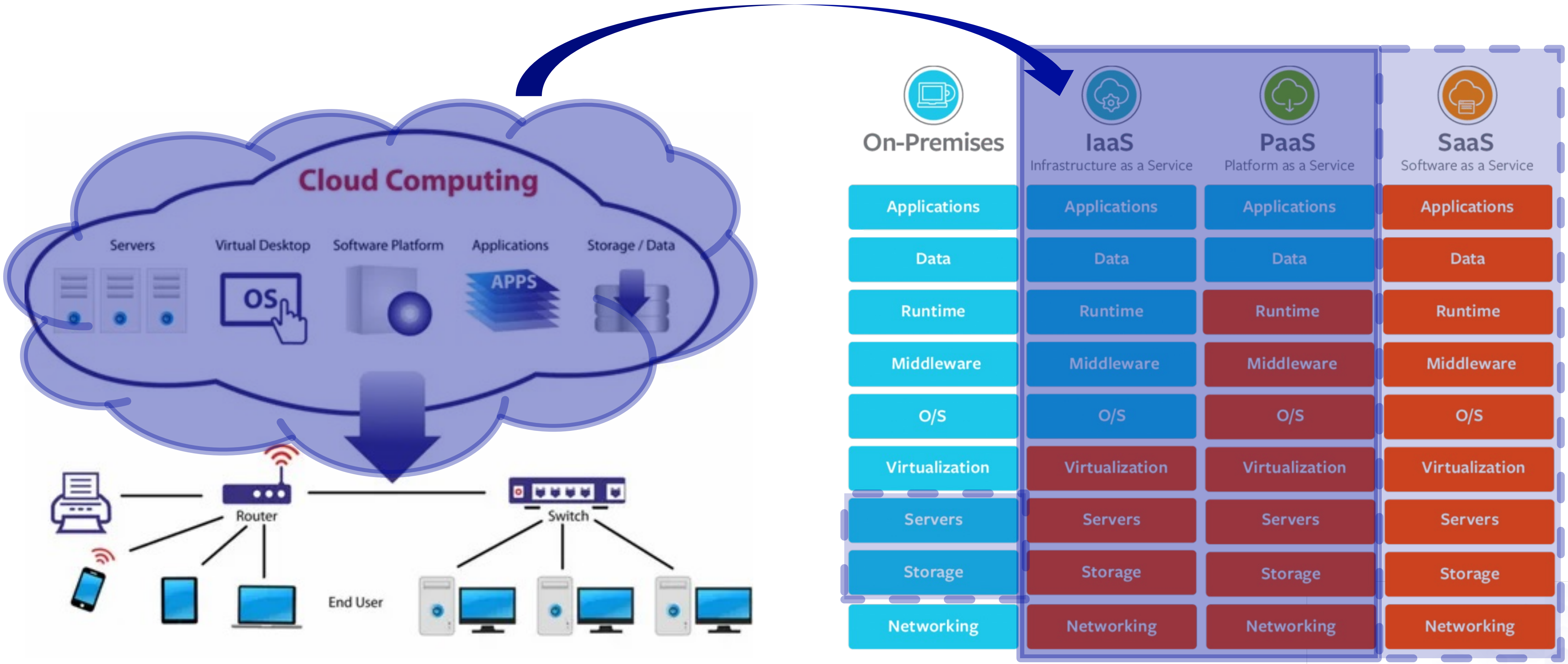


The Cloud segmentation

The Cloud is typically divided in 4 segments that cover all or part of the value chain



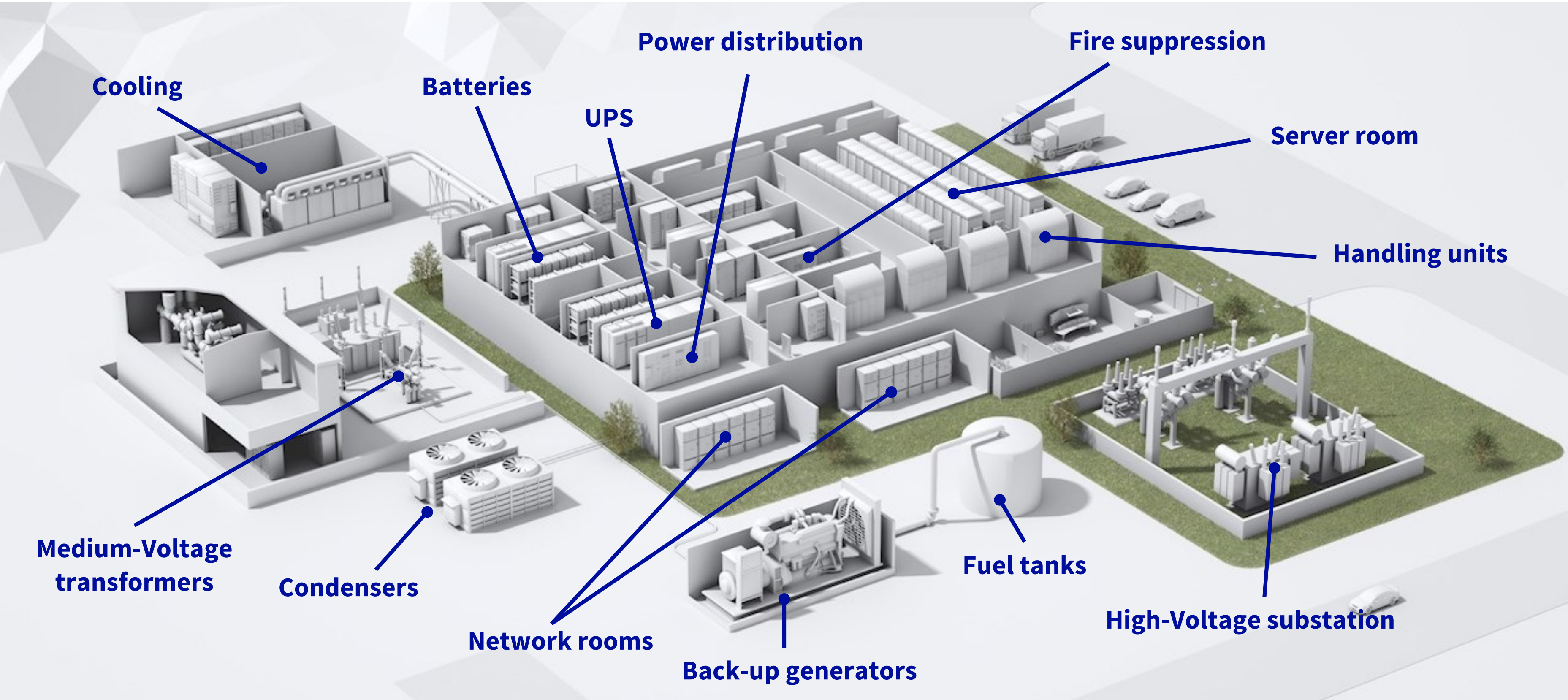
Scope covered in this lecture is the one under the CSP responsibility



02 Physical infrastructure description

The Cloud relies on heavy physical infrastructures

Data centers = land + buildings + industrial gears + networks (energy/water/telco)



Data center architectural standard - TIER I

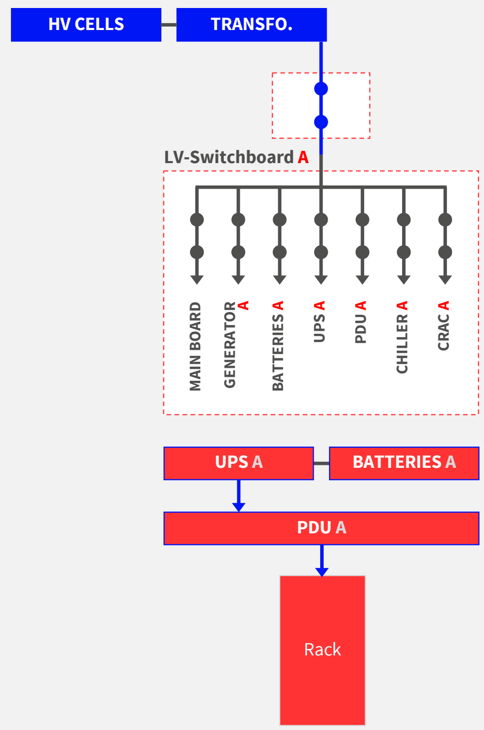
Fundamental requirements

- Non redundant capacity components.
- Non redundant distribution path (single path)..

Operational impacts

- The site is susceptible to disruption from a single unplanned event.
- The site is susceptible to disruption from any planned work activities.
- The site infrastructure maintenance can not be performed without the shutdown of the site infrastructure.

Schematic



Data center architectural standard - TIER II

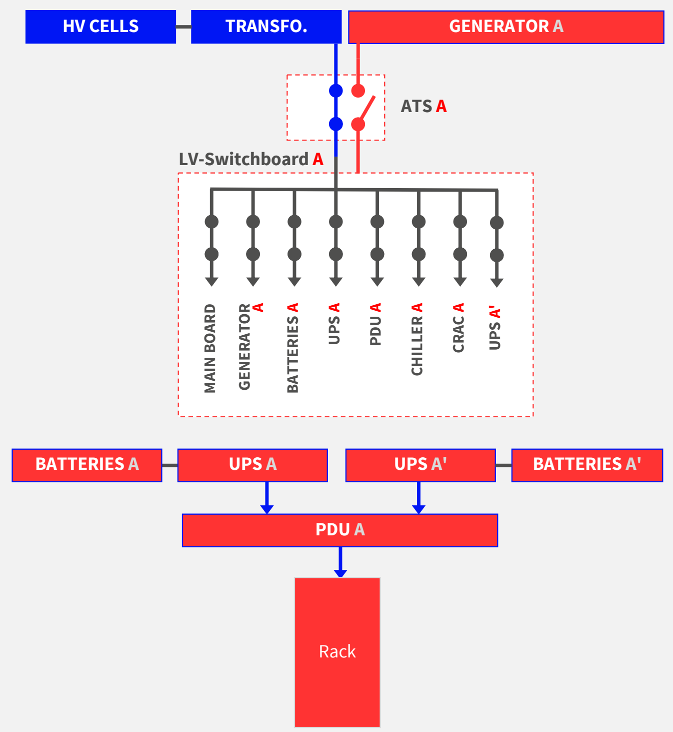
Fundamental requirements

~~› Non~~ redundant capacity components.
- Non redundant distribution path (single path).
- 12 hours of on-site fuel storage for “N” capacity.

Operational impacts

- The site is susceptible to disruption from a single unplanned event. **An unplanned capacity component failure may impact the the computer equipment.**
- The site is susceptible to disruption from any planned work activities.
- The site infrastructure maintenance can not be performed without the shutdown of the site infrastructure.

Schematic



Data center architectural standard - TIER III

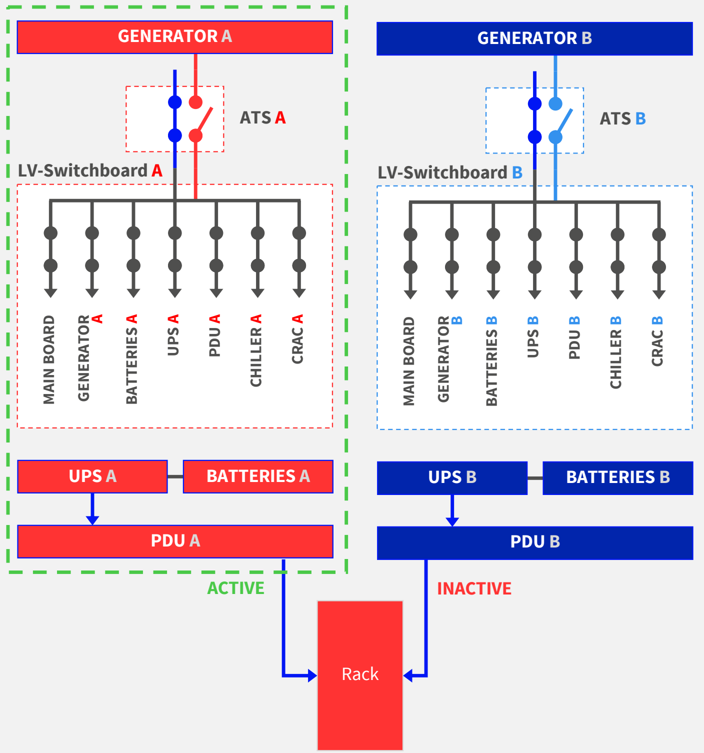
Fundamental requirements

- Redundant capacity components.
- ~~Non redundant~~ **Multiple** distribution paths. **But only one distribution path is required to serve the components at any time.**
- 12 hours of on-site fuel storage for “N” capacity.

Operational impacts

- The site is susceptible to disruption from a single unplanned event.
- The site **is not** susceptible to disruption from any planned work activities.
- The site infrastructure maintenance can ~~not~~ be performed without the shutdown of the site infrastructure. **During maintenance activities, the risk of disruption may be elevated.**

Schematic



Data center architectural standard - TIER IV

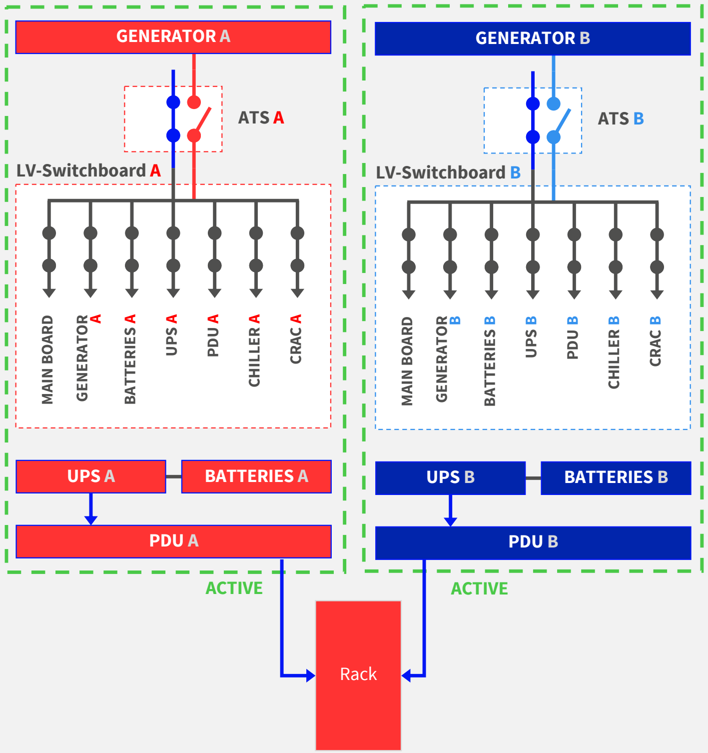
Fundamental requirements

- Redundant capacity components.
- Multiple **active** distribution paths
- 12 hours of on-site fuel storage for “N” capacity.
- **Complementary systems and distribution paths must be physically isolated from one another to prevent any single event from simultaneously impacting both systems or distribution paths.**
- **Continuous cooling is required.**

Operational impacts

- The site **is not** susceptible to disruption from a single unplanned event.
- The site is not susceptible to disruption from any planned work activities.
- The site infrastructure maintenance can be performed without the shutdown of the site infrastructure.
- **During maintenance activity where redundant capacity components shut down, the computer equipment is exposed to an increased risk of disruption in the event of a failure occurs on the remaining path.**

Schematic



Note : After any default, N capacity should be maintained.

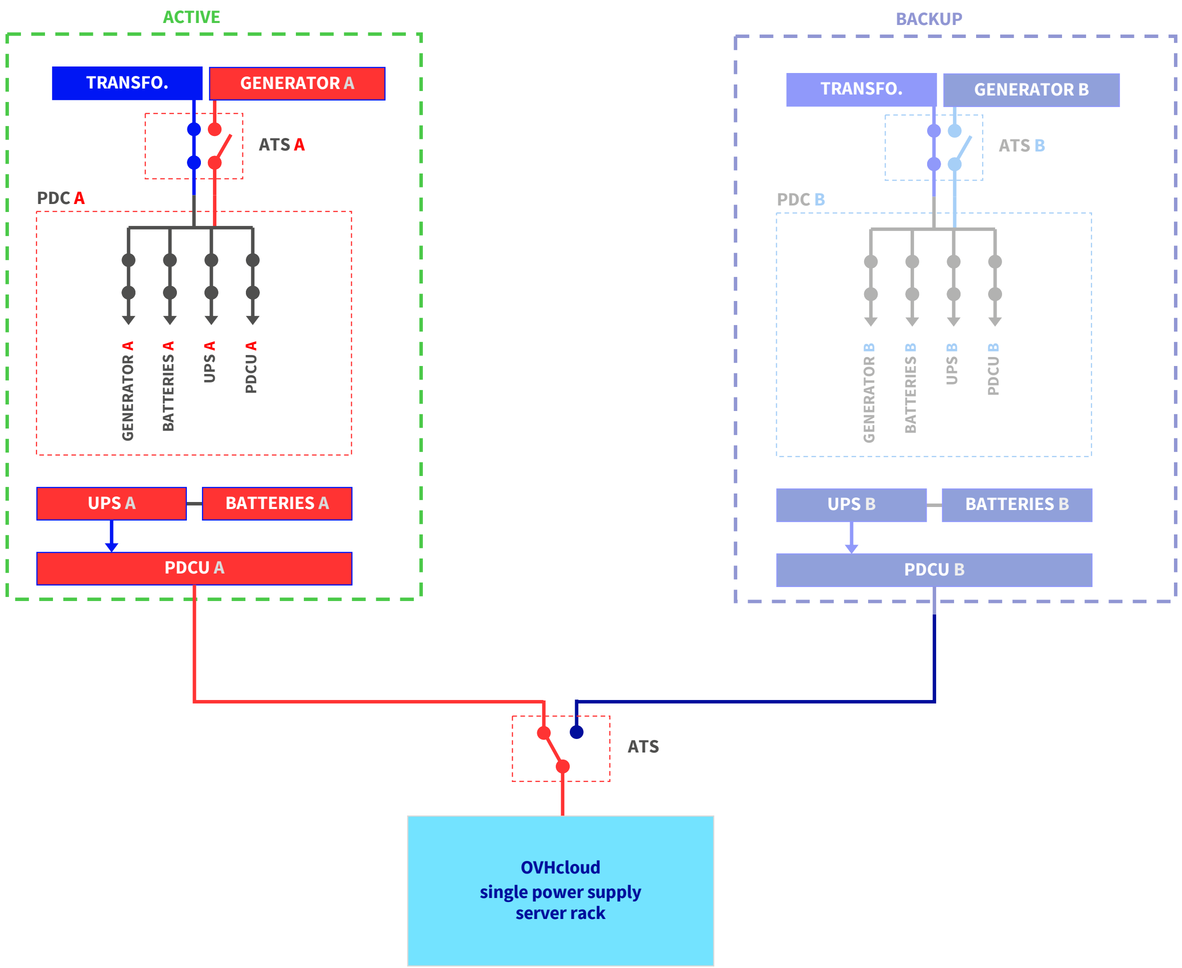
OVHcloud data centers architectural standard - TIER III+

Tiers Level	TIER I	TIER II	TIER III	TIER IV
Active capacity components to support IT load	N	N+1	N+1 (Cooling: 1P 2P)	N after any failure (Elec: 1P 2P)
Distribution paths	1	1	1 active 1 alternate (1P)	2 simultaneously active (2P)
Concurrently maintainable	No	No	Yes	Yes (1P 2P)
Fault tolerant	No	No	No	Yes (1P 2P)
Compartmentalization	No	No	No (1P 2P)	Yes (1P 2P)
Continuous cooling	Load density dependent	Load density dependent	Load density dependent	Class A (1P 2P)
On-site fuel	12 hours	12 hours	12 hours	12 hours (1P 2P)
Annual impact of maintenance and/or unplanned outages	28.8 hours	22.0 hours	1.6 hours	26 min
Site availability	99,671%	99,749%	99,982%	99,995%

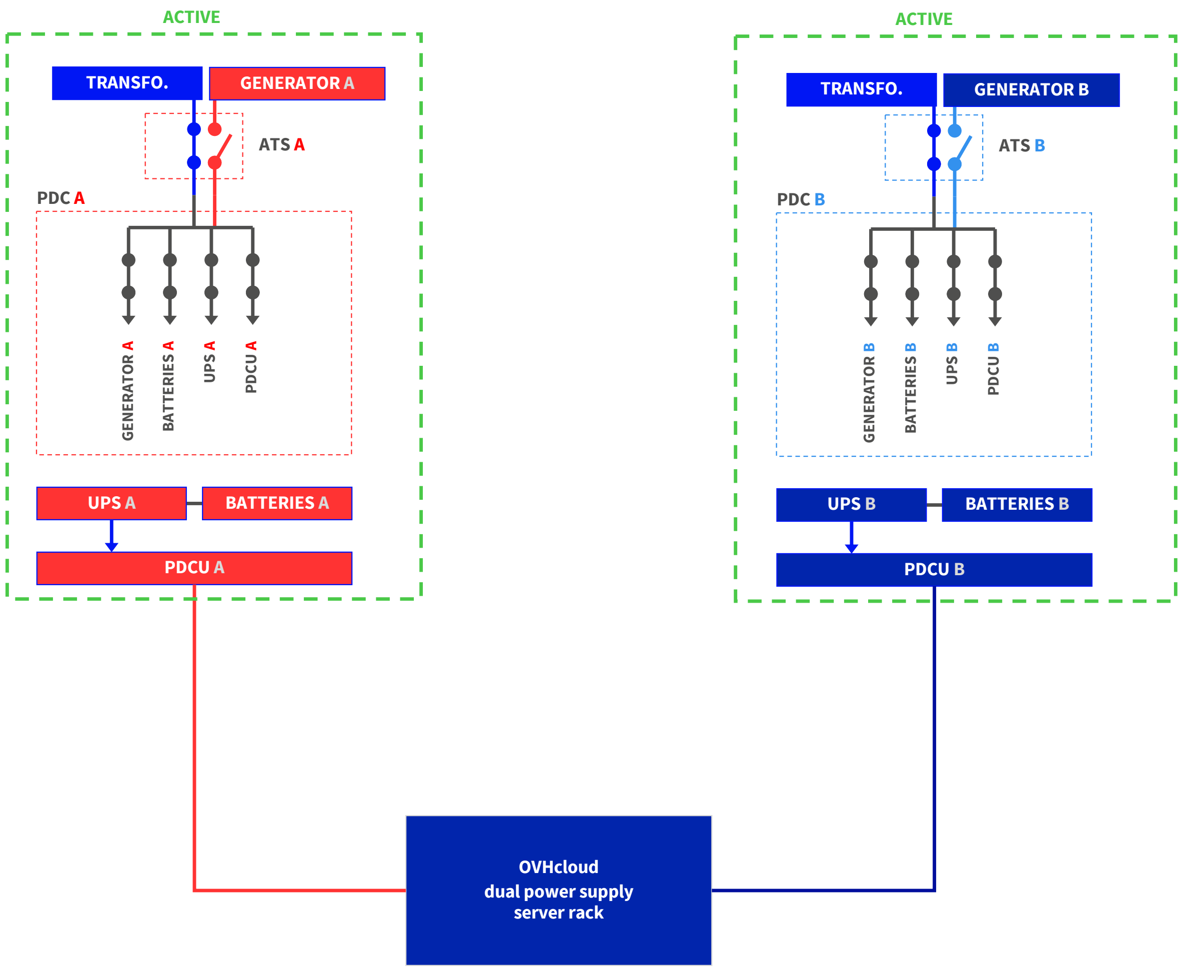
1P Single power supply servers

2P Dual power supply servers

OVHcloud single power supply server racks: N+1 architecture

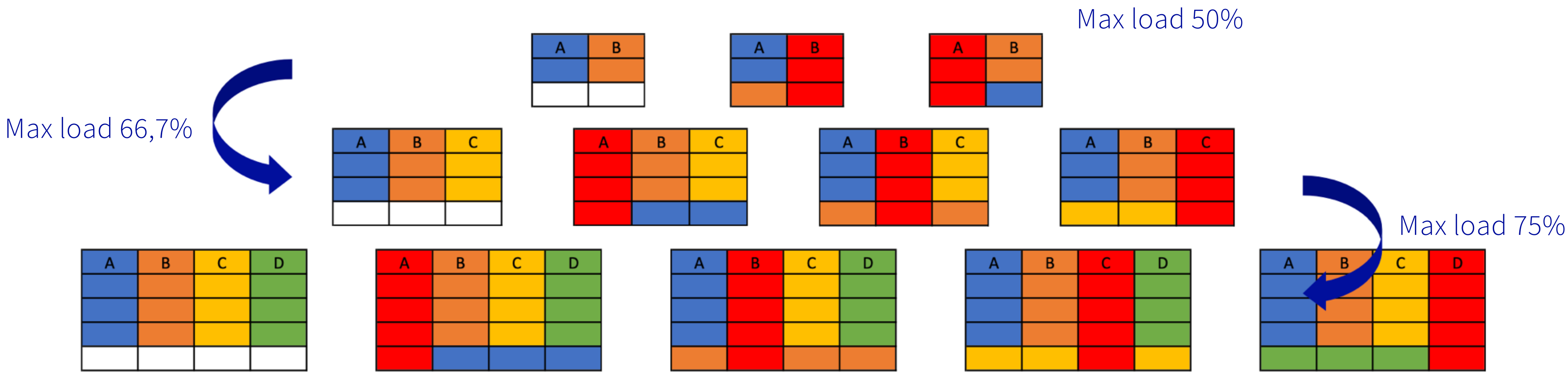


OVHcloud dual power supply server racks: 2N architecture



OVHcloud 2N architecture is optimized (2N A/B/C or A/B/C/D)

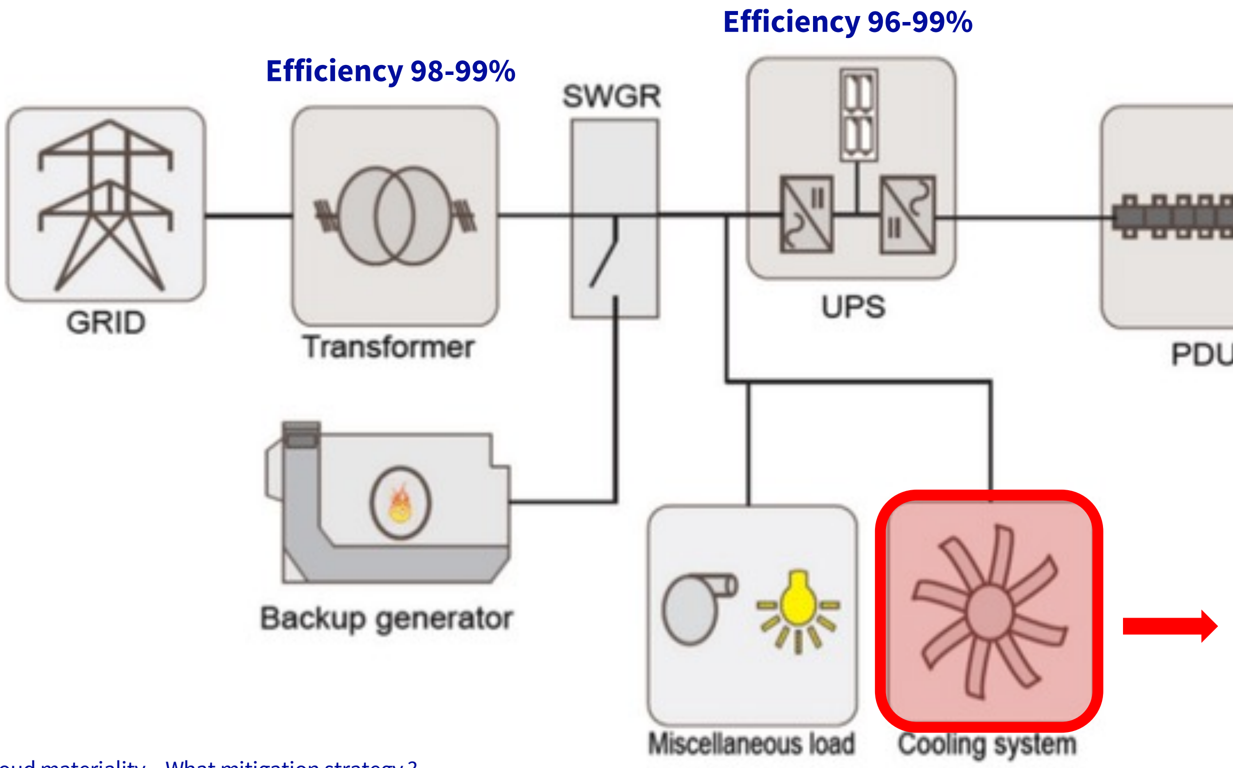
In case of failure, the service continuity is ensured by a **mutualized** back-up line-up



Whatever the architecture it's all about chasing the inefficiency

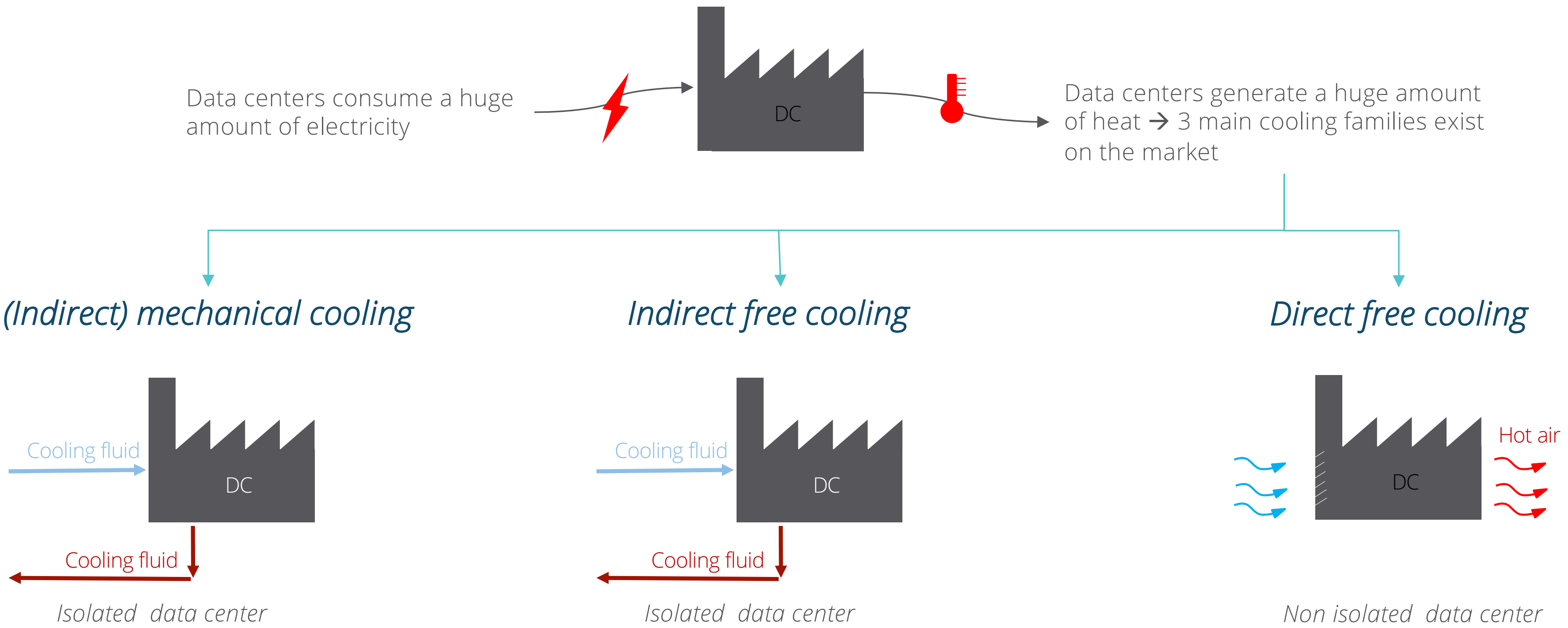
Efficiency depends on the load and on the type of equipment chosen

Transformers class (A,B,C or D) / UPS conversion mode / PSU 80+ certification (Bronze/Silver/Gold/Platinum/Titanium)



Up to 40% of data center electrical consumption

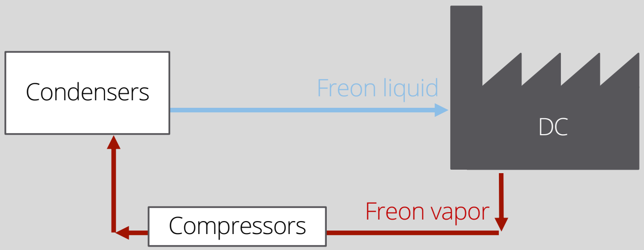
Cooling systems – choosing the right technology



Mechanical cooling only

Isolated data center

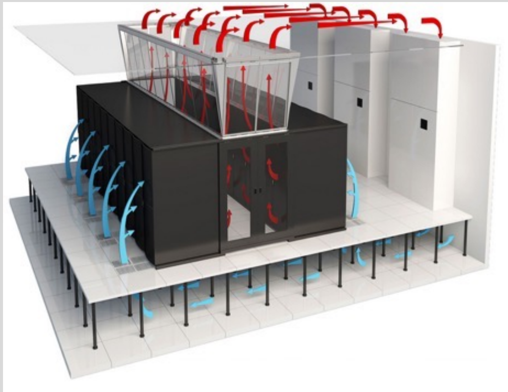
Direct expansion



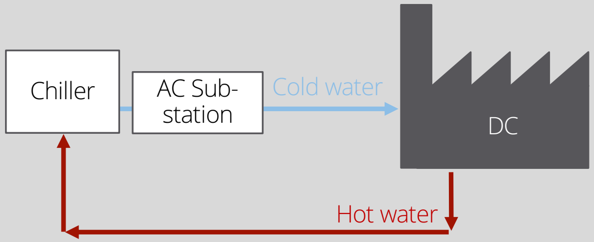
Main equipment:

- *Compressors and condensers*
- *Evaporators / CRAC*
- *Option misting*

Fluid: Freon and air



Classic chilling

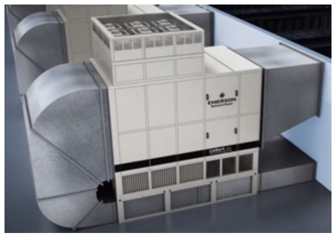


Main equipment:

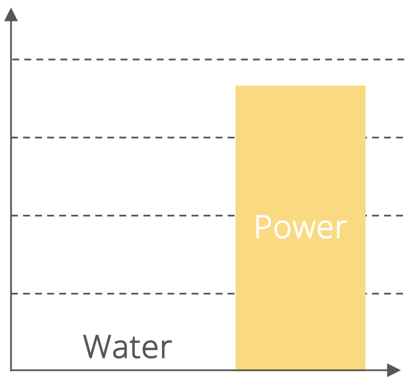
- *Chillers*
- *Air handling units (AHU) / CRAH*
- *Option misting or Cooling tower*

Fluid: Freon, water and air

Chiller



Air treatment station

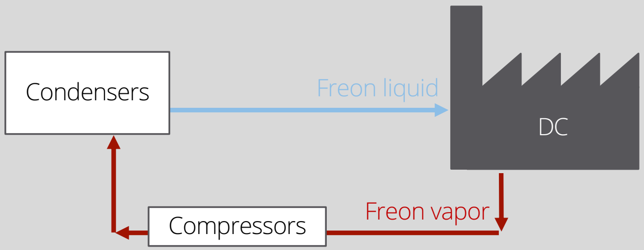


Criteria	
Temperature control	✓
Relative humidity control	✓
Dust control	✓
Maintenance	✓
Efficiency (WUE)	✓
Efficiency (PUE)	X
Cost	X
Environmentally friendly	X
Electrical consumption	X

Mechanical cooling associated with evaporative cooling

Isolated data center

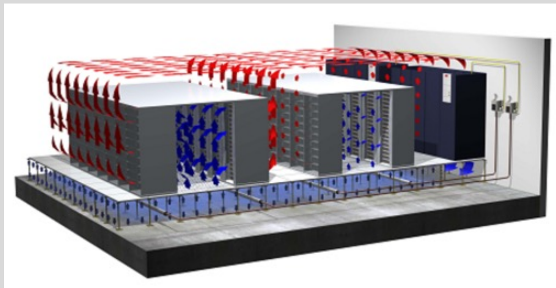
Direct expansion



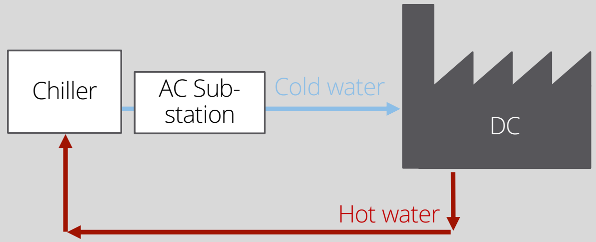
Main equipment:

- *Compressors and condensers*
- *Evaporators / CRAC*
- *Option misting*

Fluid: Freon and air



Classic chilling



Main equipment:

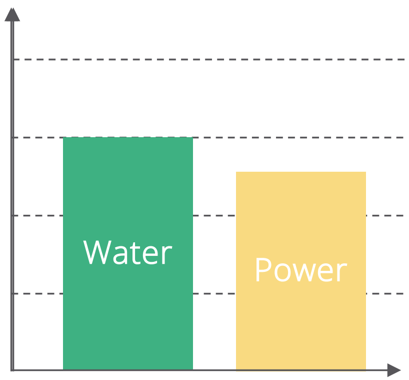
- *Chillers*
- *Air handling units (AHU) / CRAH*
- *Option misting or Cooling tower*

Fluid: Freon, water and air

Chiller

+

Wet cooling tower

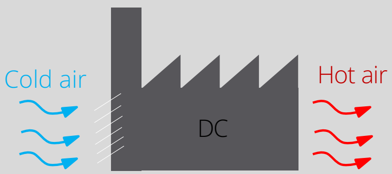


Criteria	
Temperature control	✓
Relative humidity control	✓
Dust control	✓
Maintenance	✓
Efficiency (WUE)	X
Efficiency (PUE)	X
Cost	X
Environmentally friendly	✓
Electrical consumption	X

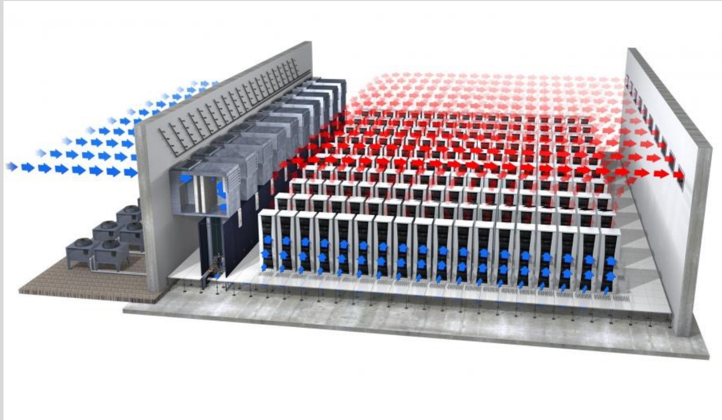
Direct free cooling

Non isolated data center

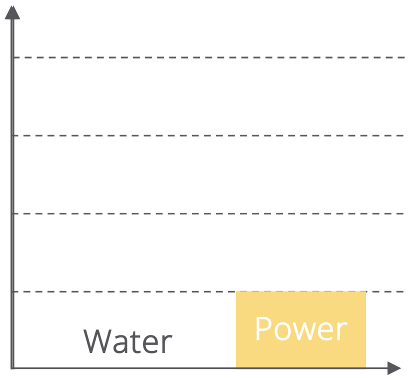
Direct free cooling



Non isolated data center



Ventilation system

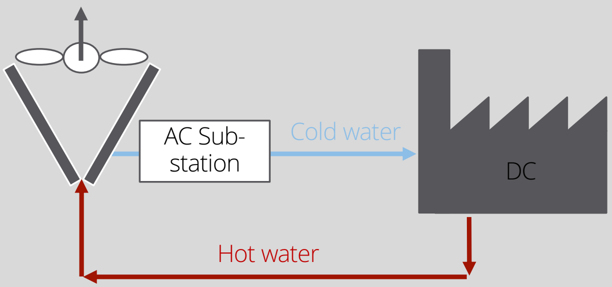


Criteria	
Temperature control	X
Relative humidity control	X
Dust control	X
Maintenance	✓
Efficiency (WUE)	✓
Efficiency (PUE)	✓
Cost	✓
Environmentally friendly	✓
Electrical consumption	✓

Indirect free cooling with evaporative cooling

Isolated data center

Indirect free cooling

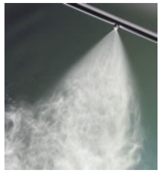


Main equipment:
- *Compressors and condensers*
- *Evaporators / CRAC*
- *Option misting*

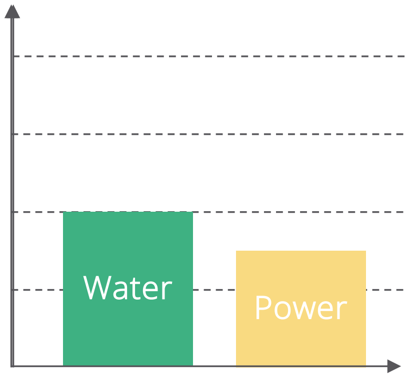
Fluid: Freon and air

Dry coolers

Misting



Humid media

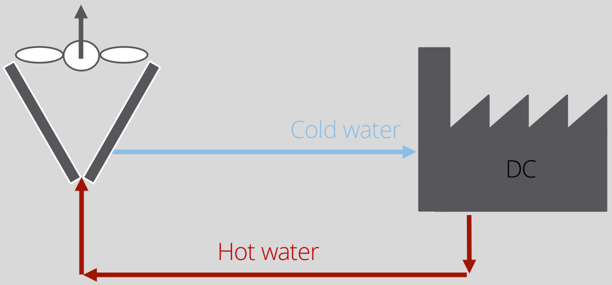


Criteria	
Temperature control	✓
Relative humidity control	✓
Dust control	✓
Maintenance	✓
Efficiency (WUE)	✓
Efficiency (PUE)	✓
Cost	✓
Environmentally friendly	✓
Electrical consumption	X

Indirect free cooling with evaporative cooling (OVHcloud proprietary DLC)

Isolated data center

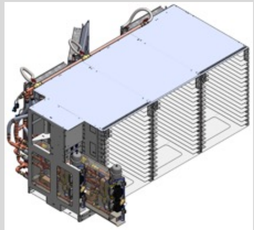
Indirect free cooling

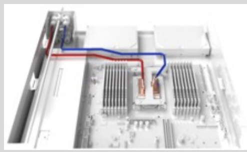


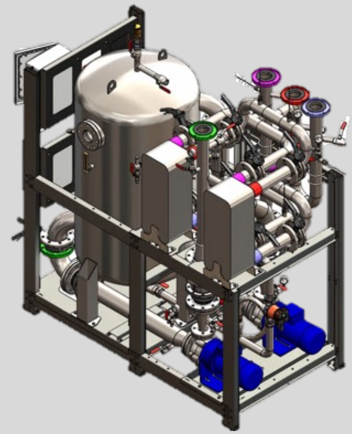
Main equipment:

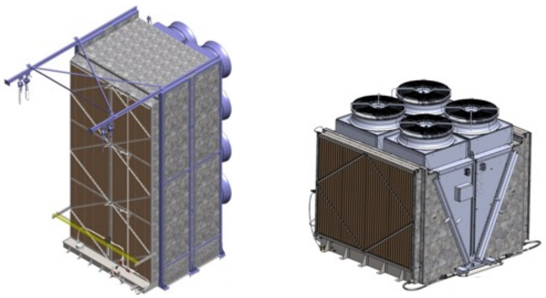
- *Built to purpose racks*
- *Cooling modules*

Fluid: water and air

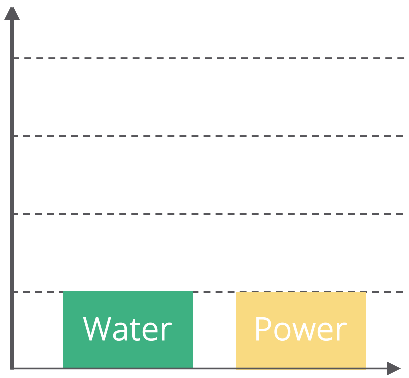








Dry coolers with humid media



Criteria	
Temperature control	✓
Relative humidity control	✓
Dust control	✓
Maintenance	✓
Efficiency (WUE)	✓
Efficiency (PUE)	✓
Cost	✓
Environmentally friendly	✓
Electrical consumption	✓

OVHcloud Cooling Design Resiliency

2 **Water Cooling Modules**

(supplying cold water to the "fridge doors" at the back of the rack)

- 2N electrical supply
- 2N water pumps

3 **"Fridge doors"**

= 30% of server's heat dissipation

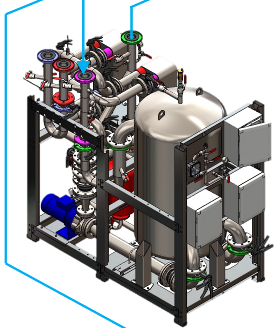
(fans + heat exchanger = air cooling system, refreshing the warm air from servers, then supplying tepid water to the Mini Water Cooling Modules)

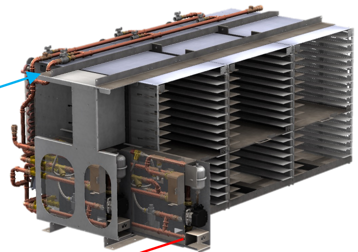
- 50% fans on source A, 50% on source B

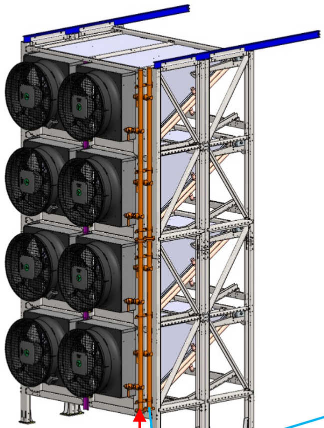
1 **Dry Coolers**

(cooling the warm water generated by servers' cooling system, using outside fresh air)

- 50% fans on source A, 50% on source B
- N+1







4 **Mini Water Cooling Modules**

= 70% of server's heat dissipation

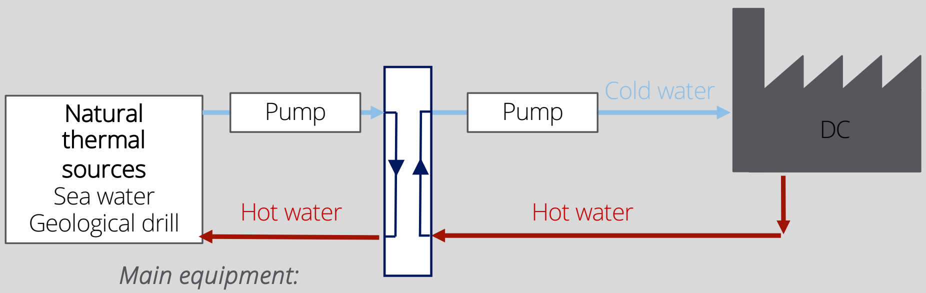
(closed-circuit water cooling system (plate heat exchanger) supplying tepid water to cool the servers' CPUs or GPUs)

- 2N

Indirect free cooling with natural water cooling

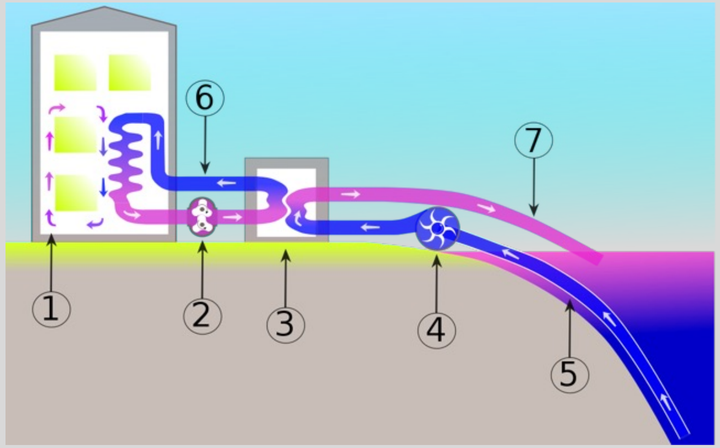
Isolated data center

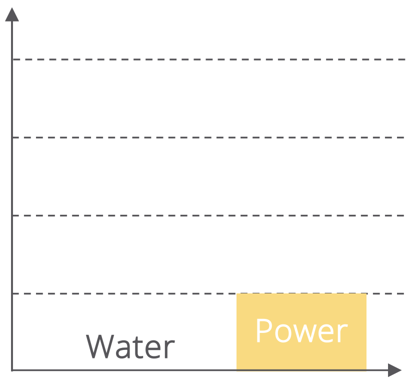
Indirect free cooling



Main equipment:

- *Pumping station*
- *Heat exchanger*





Criteria	
Temperature control	✓
Relative humidity control	✓
Dust control	✓
Maintenance	✓
Efficiency (WUE)	✓
Efficiency (PUE)	✓
Cost	✓
Environmentally friendly	✓
Electrical consumption	✓

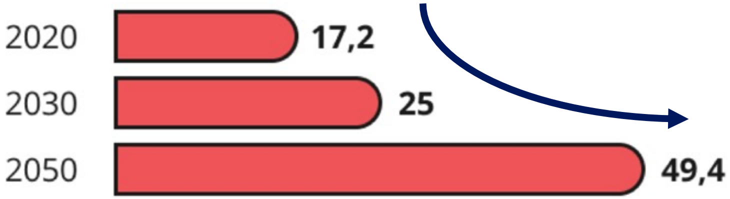
03

Environmental impact and its measurement

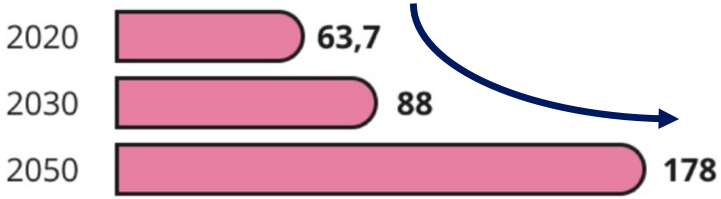
ICT sector impact on the environment is multifactorial and exponential

Projections show that needs will more than double across all segments*

Carbon footprint
(million tonnes of CO_2 eq.)

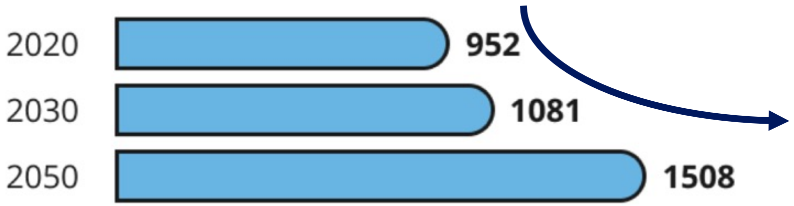


Resources used (Abiotic resources/biomass/soil movement/water/air)
(million tonnes)



Energy consumption
(in TWh)

Metal and mineral consumption (Metal/mineral resources as antimony eq.)
(tonnes Sb eq.)



**Source – ADEME-Arcep 2023 assessment of the digital environmental footprint in France*

The 4 main environmental impact of the ICT

Some (stunning) numbers

2.1-3.9% (2020)	ICT sector contribution to GHG emissions in 2020 (x3 by 2050)
-1000 BC - Today	As much minerals to be extracted by 2050 (circa x2 for ICT vs 2020)
2.5% (2020)	ICT sector demand to global energy demand (x2 by 2050)
4.2-6.6 billions m^3	AI water consumption projection by 2027 (⇔ France)
1400+ (2024)	# of Data centers in Europe (+100 by 2026)

Greenhouse gases emissions (aka carbon footprint)

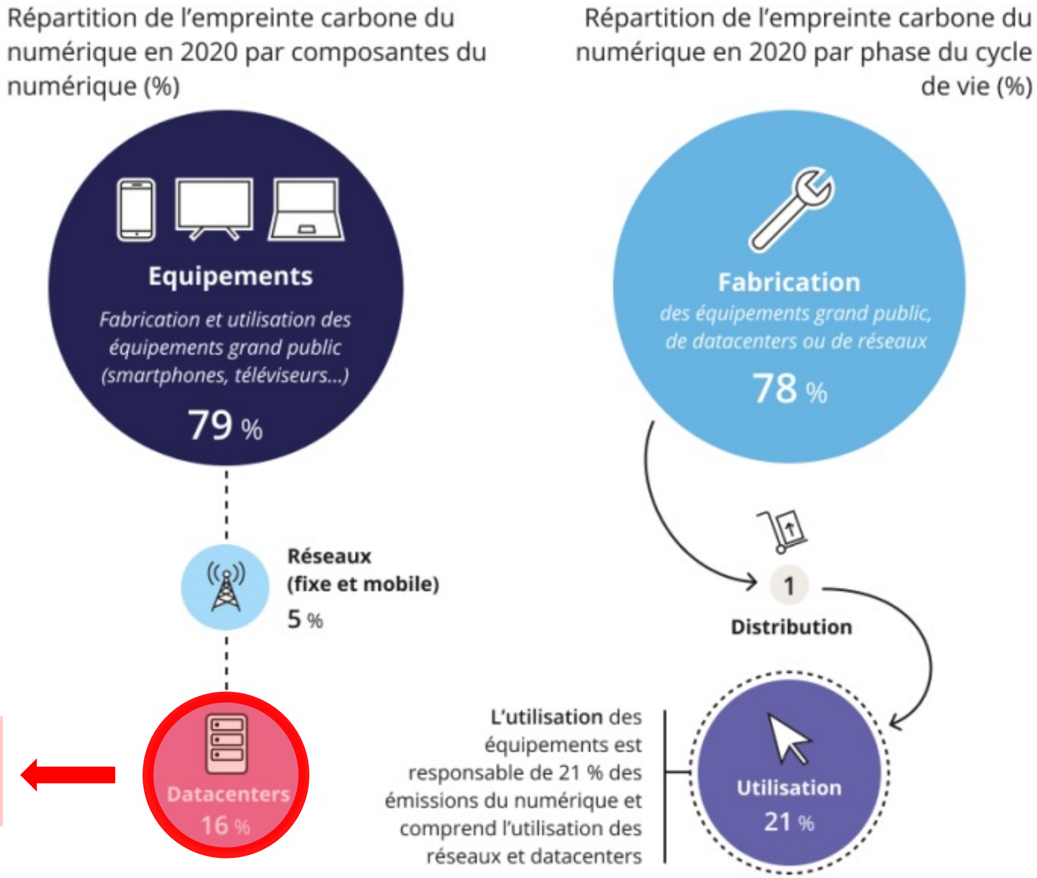
Indicator : measurement of the amount (mass) of greenhouse gases released into the atmosphere

Since not all greenhouse gases have the same global warming potential (GWP), it was decided to compare them against a standard: carbon dioxide equivalent (CO2e)

By convention : 1 molecule of CO2 has a warming potential of 1 over 100 years
For example, 1 molecule of methane (CH4) has a warming potential of 28 over the same period

The carbon balance (tCO2e) can be calculated using annual flow vision (GHG protocol) or life cycle analysis (LCA)

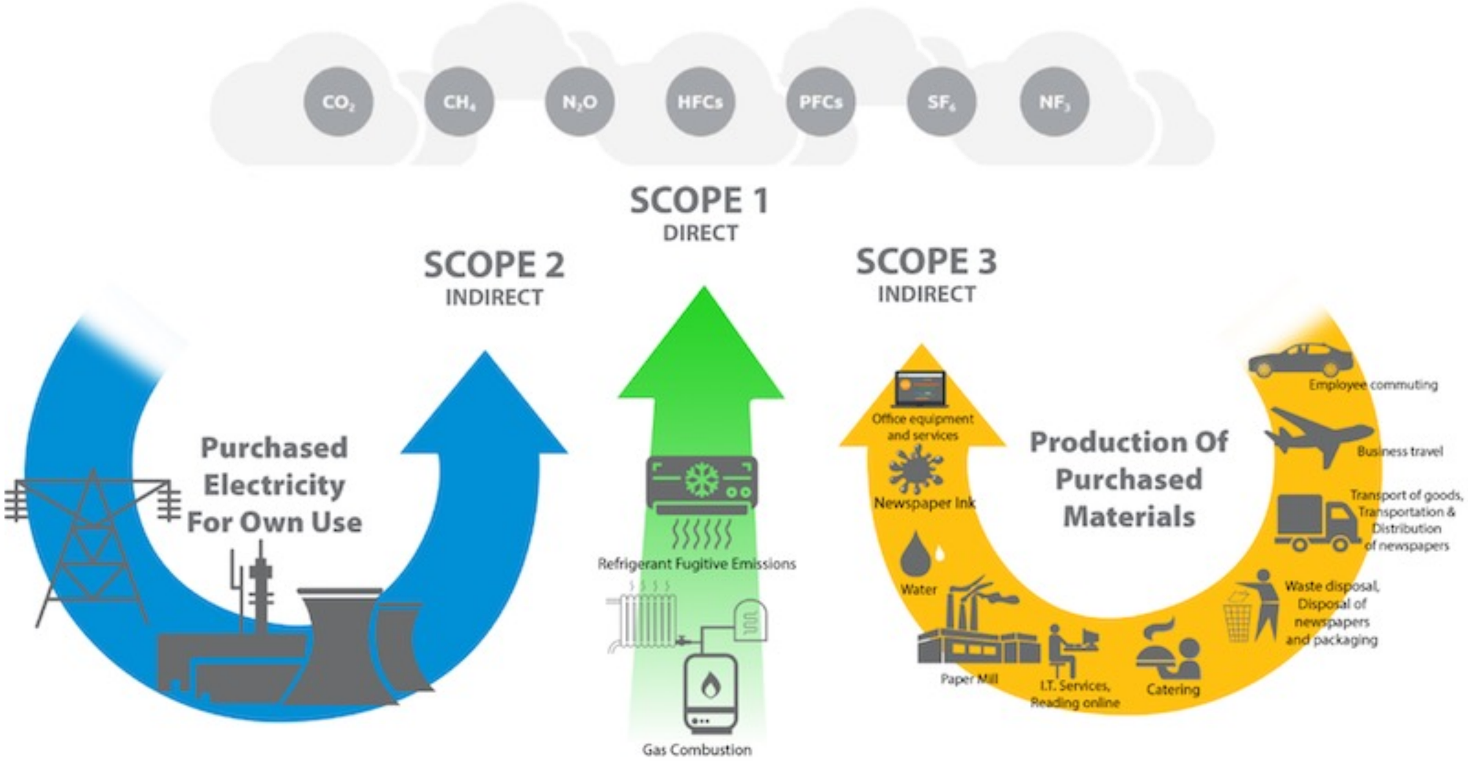
ICT carbon footprint by components and by cycle



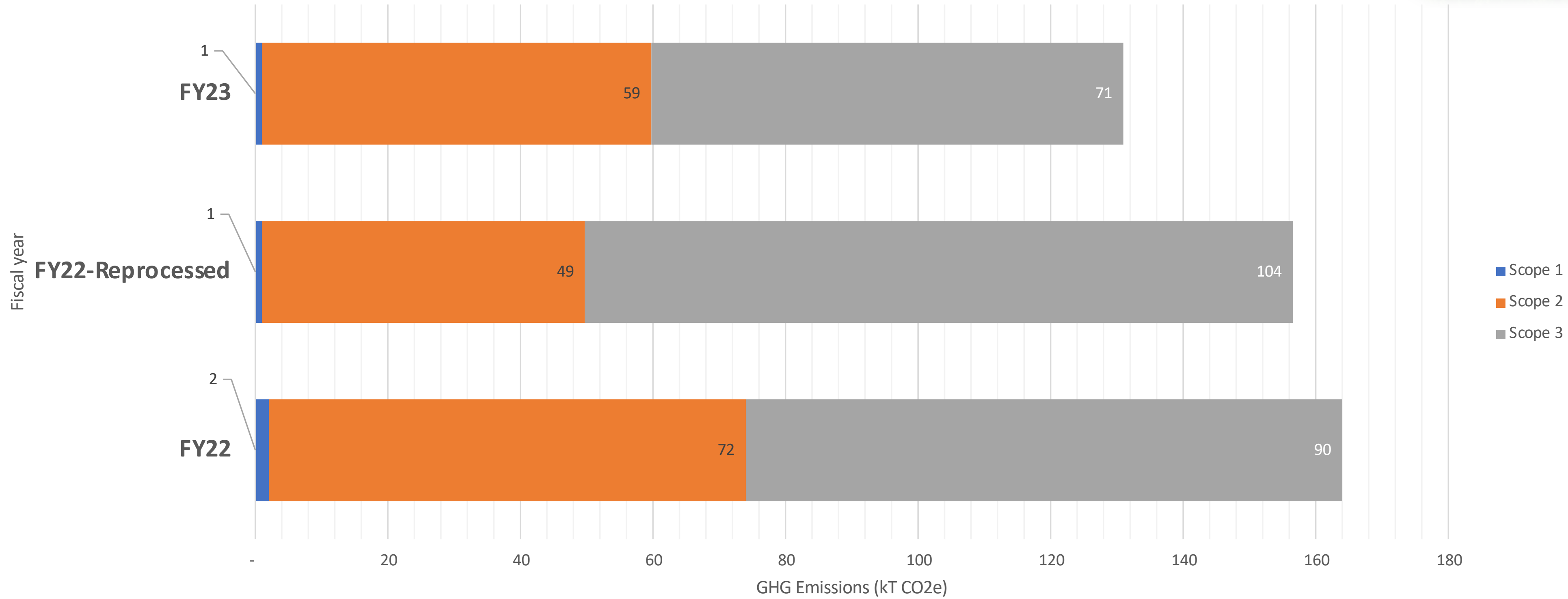
Impact of the Cloud infrastructure

**Source – ADEME-Arcep 2023 assessment of the digital environmental footprint in France*

GHG protocol accounting is divided in 3 Scopes

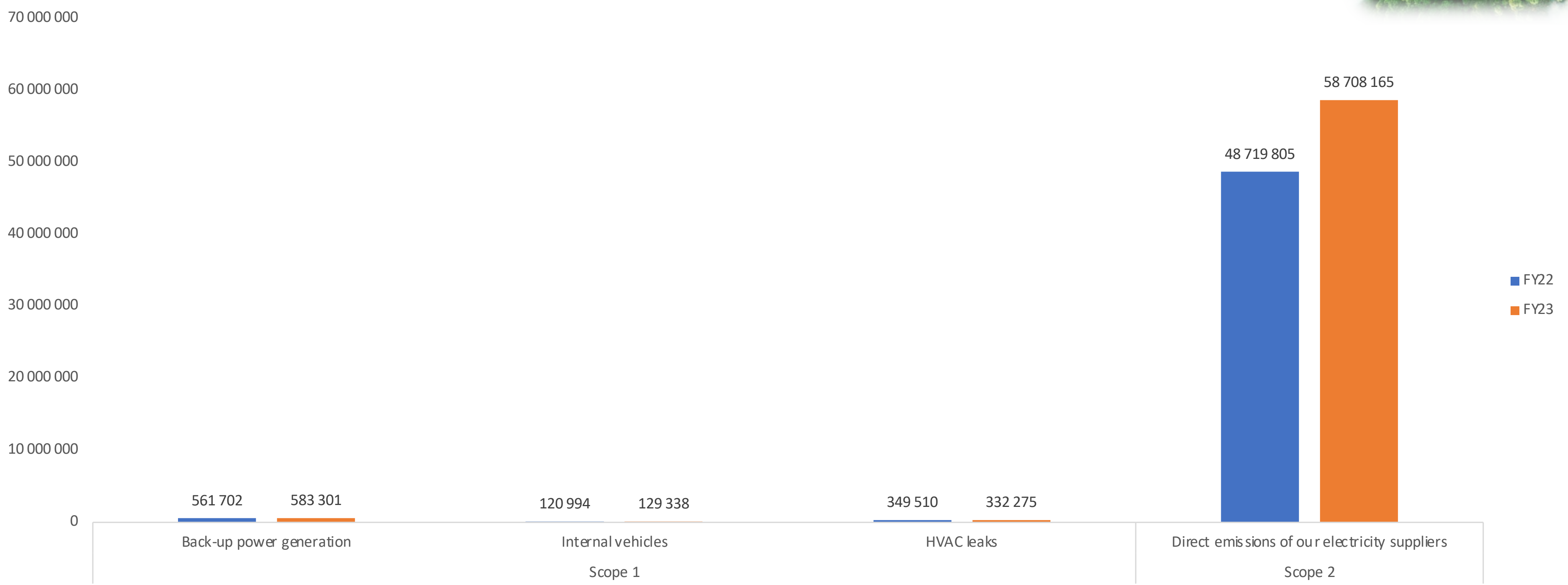


OVHcloud carbon footprint breakdown

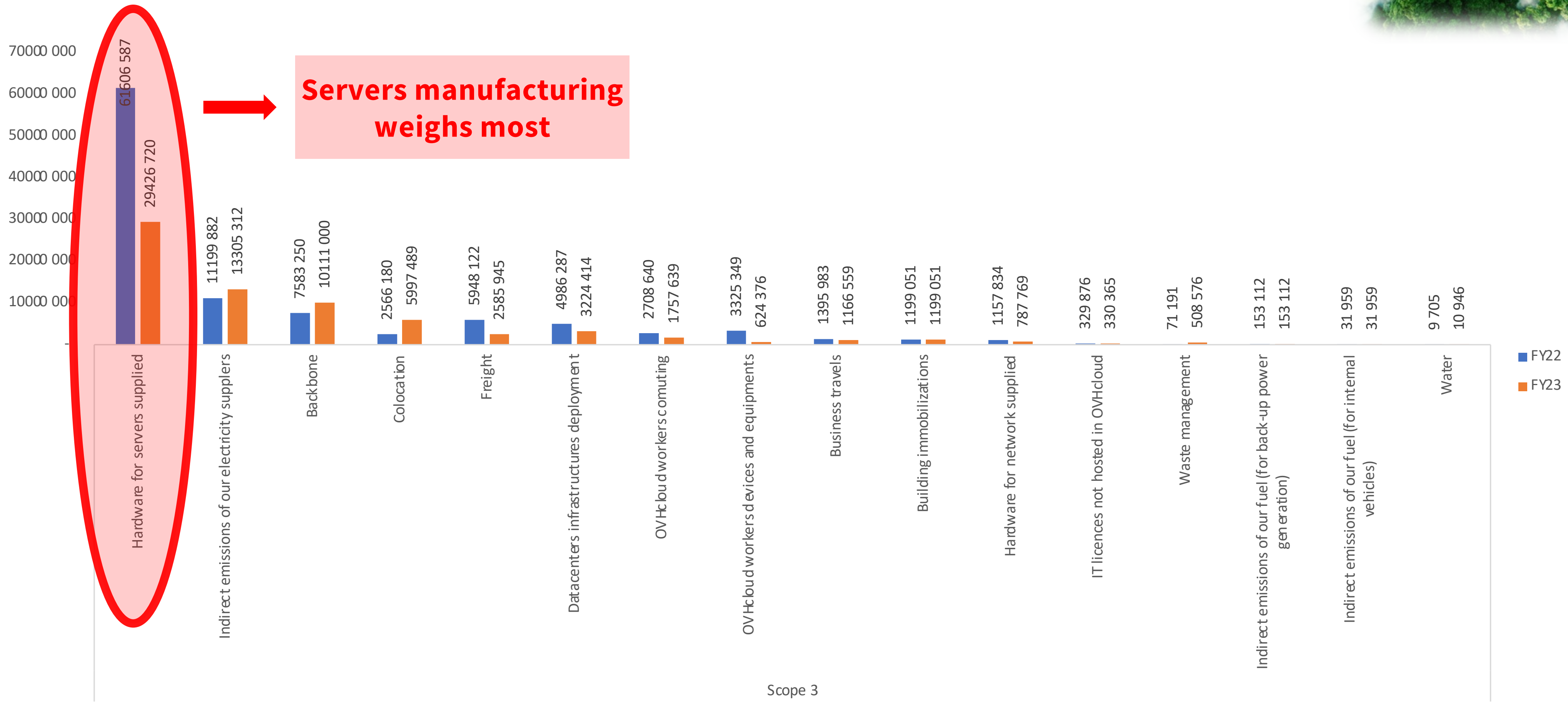


Source – OVHcloud 2023 extra financial report

Scopes 1 et 2 break down



Scope 3 breakdown



Data center energy efficiency indicator (PUE)

The **PUE** (Power Usage Effectiveness) is the ratio between the total energy consumed by the data center and the electrical energy consumed by the servers or the network equipment

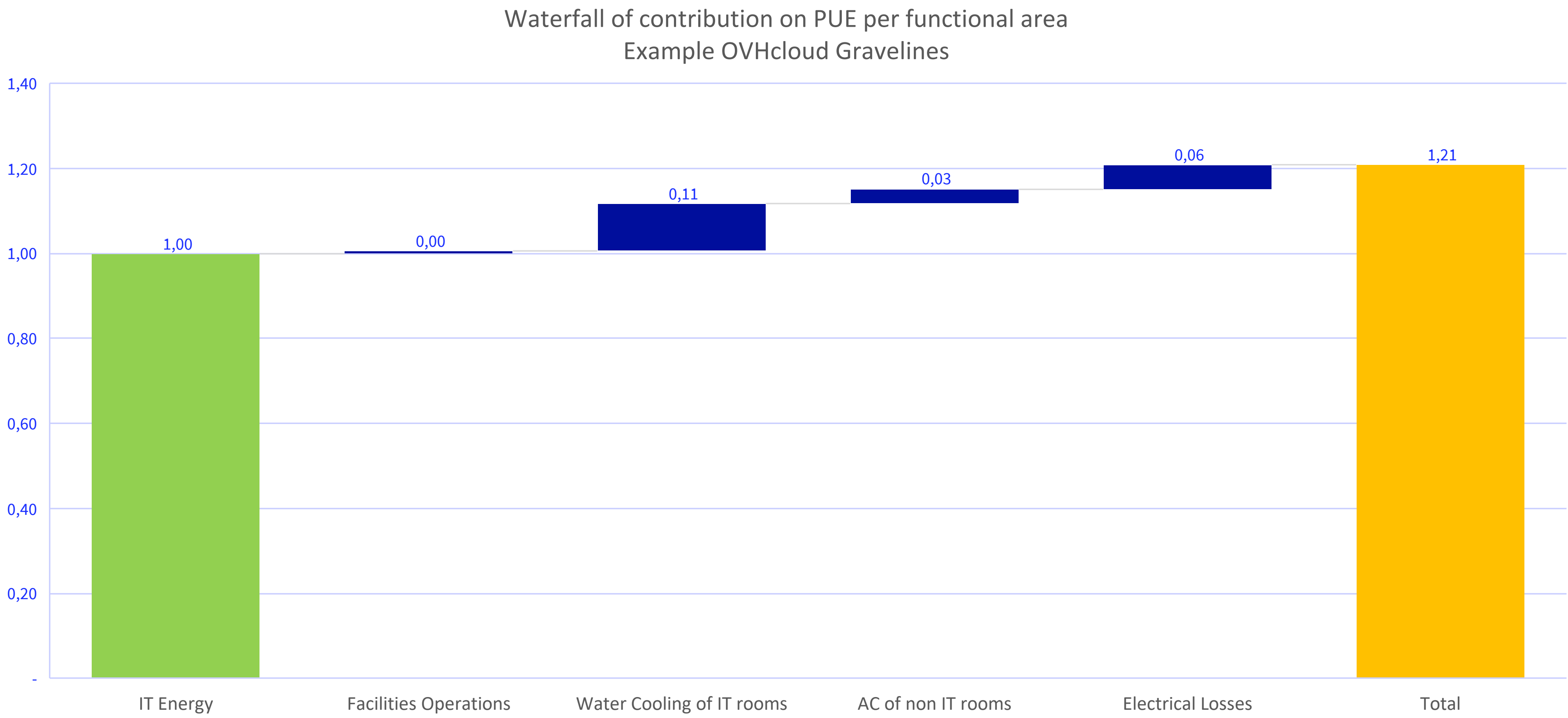
It is measured over 12 rolling months to average seasonality phenomena - it has no unit

This indicator is defined by the ISO/IEC 30134-2 standard

The global average PUE is 1.56* mainly driven by cooling systems

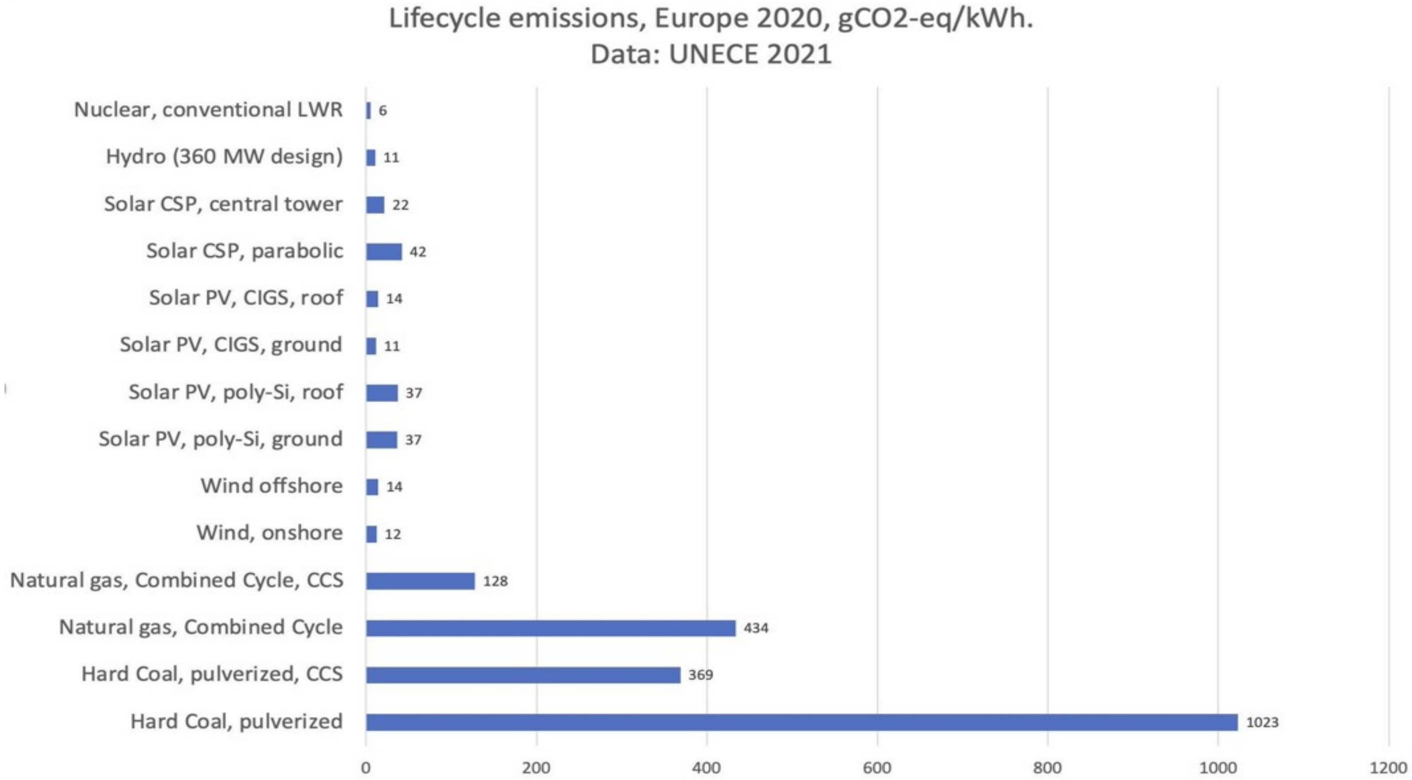
**Source – Uptime institute Global Data Center Survey 2024*

PUE components



Electricity carbon intensity indicator

- The electricity of the data centers is produced from primary energy sources that are heterogenous in terms of carbon intensity



LWR	Light Water Reactor
CSP	Concentrated Solar Power
PV	Photovoltaic
Poly-SI	Polycrystalline Silicon
CCS	Carbon Capture & Storage
CIGS	Copper Indium Gallium Selenide

Renewable energy indicator (REF)

An energy source is deemed to be renewable when the primary energy source from which it is extracted is renewed naturally and continuously

Renewable energy is generally clean, meaning it produces little or no polluting emissions

Clean, renewable energy is called green energy

Note: not all clean energies are renewable (nuclear for example)

The **REF** (Renewable Energy Factor) is the ratio that measures the rate of electricity from renewable sources relative to the total energy consumed by a data center - it is expressed in %

This indicator is defined by the ISO/IEC 30134-3 standard

Data center carbon efficiency indicator (CUE)

CUE (Carbon Usage Effectiveness) is the ratio of CO2 emissions linked to the data center energy consumption to the electrical energy consumed by the servers or the network equipment

It is measured over 12 rolling months - expressed in kgCO2e/kWh

This indicator is defined by the ISO/IEC 30134-8 standard

The CUE might be a misleading indicator
As the energy mix varies greatly from country to country*, the Cloud Service Provider footprint has a huge impact on it

**Information available at https://app.electricitymaps.com/map*

Data center water efficiency indicator (WUE)

Reminder: most of the water used by the data centres is the evaporative water used in outdoor cooling systems

Water Usage Effectiveness (**WUE**) is the ratio of the amount of water used directly by a data center to the electrical energy consumed by the servers or the network equipment

It is measured over a rolling 12-month period to average seasonality phenomena - expressed in l/kWh

This indicator is defined by the ISO/IEC 30134-9 standard

The global average WUE is 1.8 l/kWh*

**Source – US Department of Energy 2019*

04 Technological mitigation levers

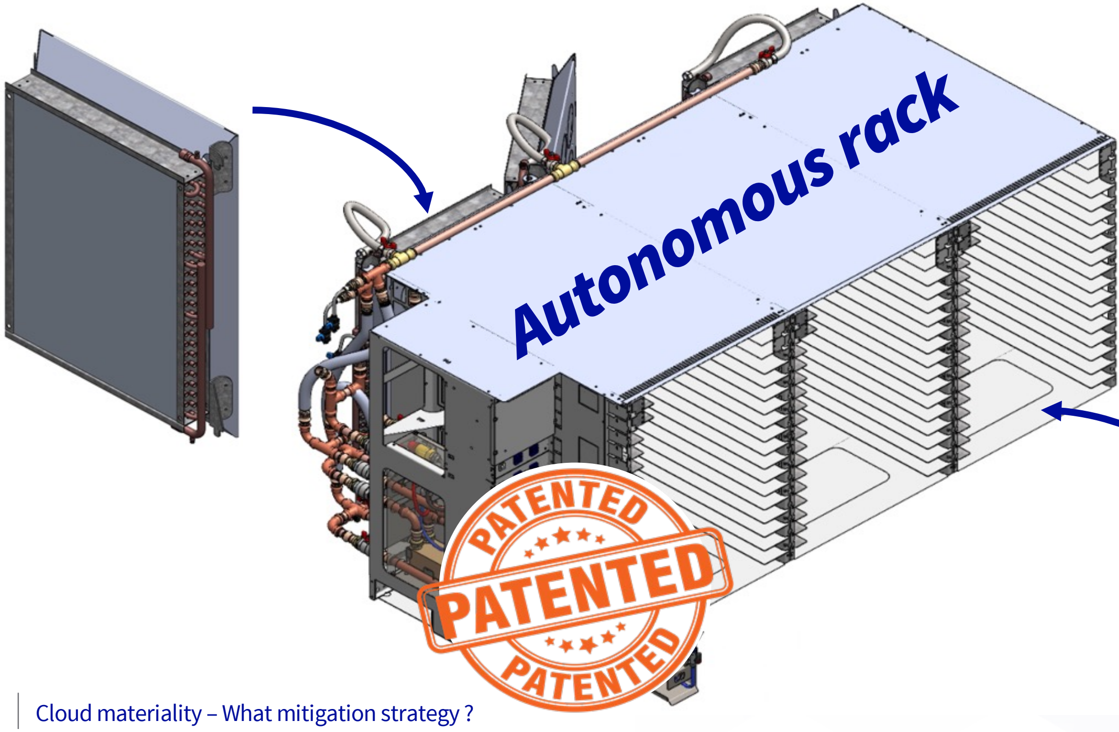
1- Scope 1 reduction

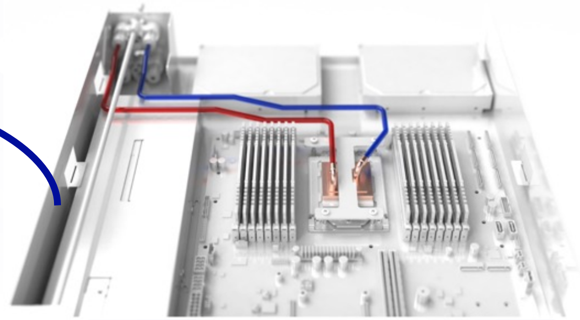
- Introduction of bio-sourced fuel (HVO100) for a 70% to 80% reduction in generator emissions
- Removal of refrigerants from servers/network/power (replaced by closed loop water cooling systems)
- Waste heat recovery for heating (removal of gas heater in Limburg)

2- Scope 2 reduction via water cooling

Power efficiency: OVHcloud PUE = 1.26 (1.29 previous year)
(vs 1.56 industry average which is no longer decreasing in the last 4 years)

New design allows a PUE as low as 1.15



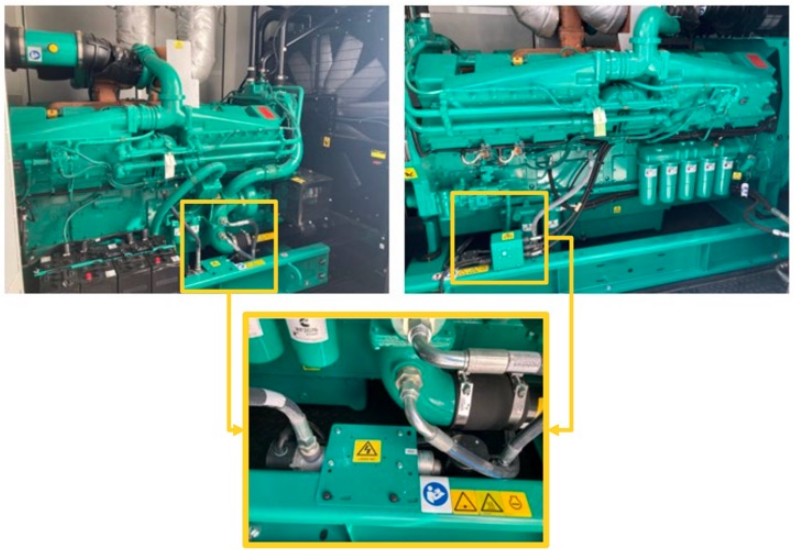


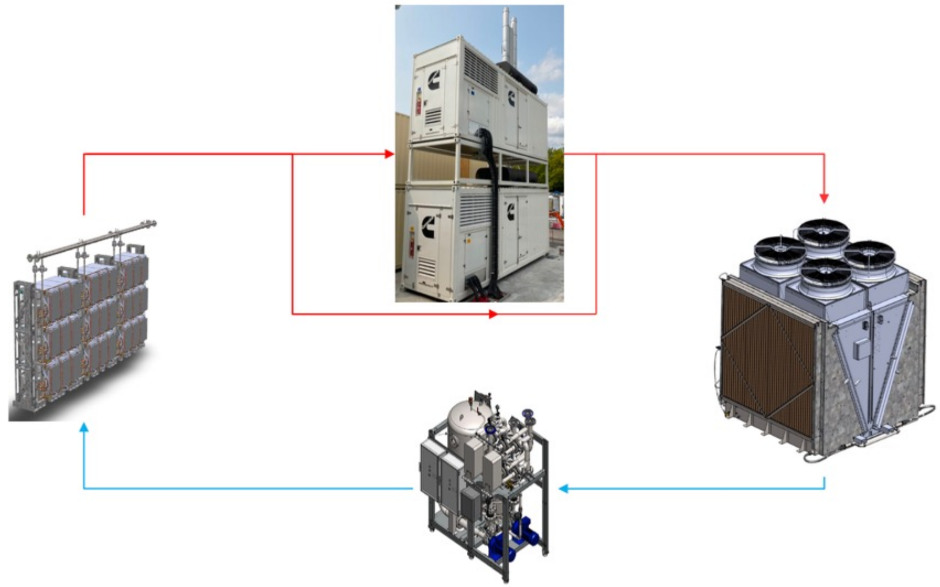
Water cooled CPU/GPU

2bis- Scope 2 reduction via waste heat recovery

Back-up generators have pre-heating systems which power ranges from 4kW up to 18kW with a temperature set-up at 40°C (run time close to 50%)

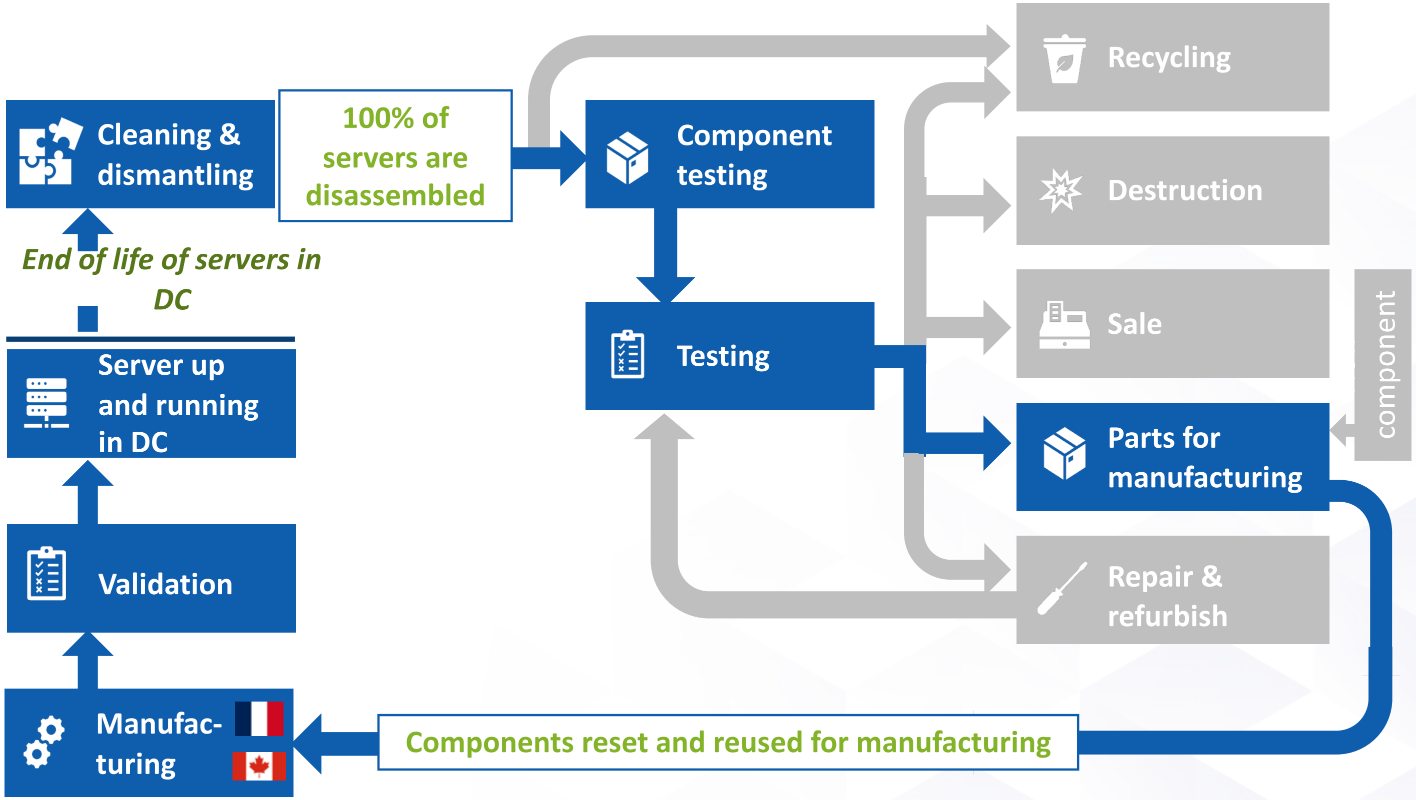
Waste heat recovery feasible with OVHcloud water profile (inlet 27°C – outlet 47°C)





3- Scope 3 reduction and resource savings by extending the lifespan of the components

Reverse supply chain - OVHcloud components reuse ratio between 25% to 36%

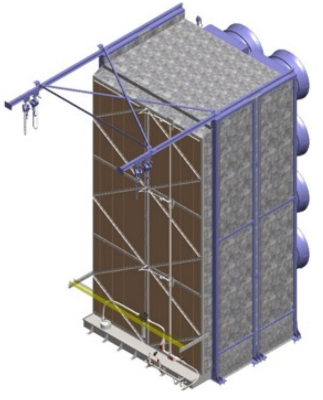
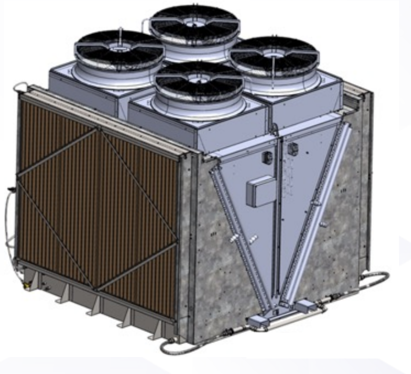


4- Water frugality on the heat dissipation front

Due to OVHcloud water profile, evaporative cooling is working only when outside temp is above 25°C

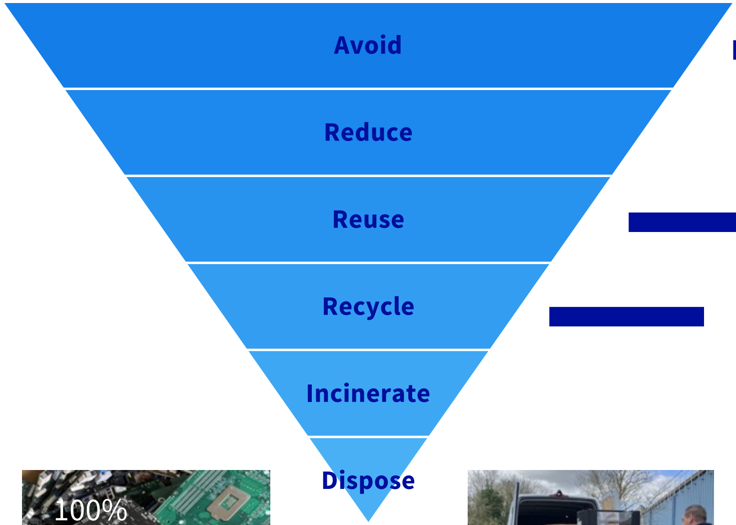
Water efficiency: OVHcloud WUE = 0.30-0.37 l/kWh (vs 1.8 l/kWh industry average)

New design allows a WUE as low as 0.1 l/kWh

SlideIN dry cooler	V shaped dry cooler (VDC)
	

5- Resource savings through minerals recovery

Pyrolysis + electrolysis processes to recover Copper and aggregated Gold/Silver/Nickel/Palladium/Platinum from WEEE



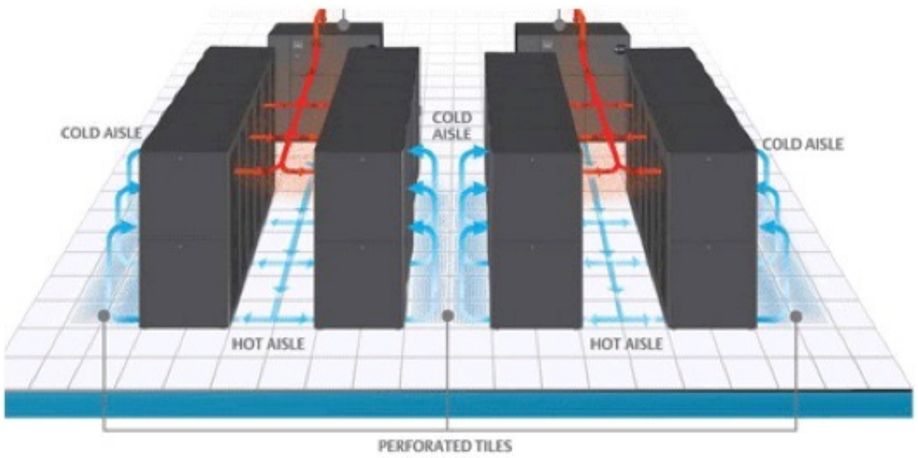
	CPU (kg)	DD (kg)	MB (kg)	RAM (kg)	TOTAL (kg)
Reused components (reverse supply chain)	935	9 637	7 742	12 781	**31 094**
Resold components (reverse supply chain)	2 190	8 665	10 912	17 755	**39 841**
Minerals recovery (recycling chain)	194	5 945	6 392	480	**13 057**

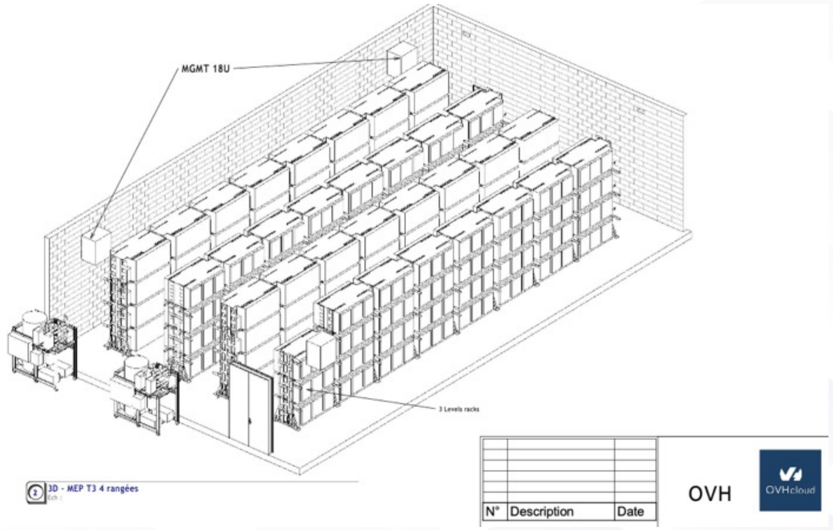
6- Land artificialisation reduction via building revamp

Brownfield vs greenfield rate: 27 out of 30 OVHcloud owned DCs are old industrial buildings

6bis- Land artificialisation reduction via densification

OVHcloud servers room urbanization allows a greater U/m²





Typical colocation layout
Cold aisle / hot aisle containment
48 U **vertical** racks
104 racks hosted in a 207 m² server room

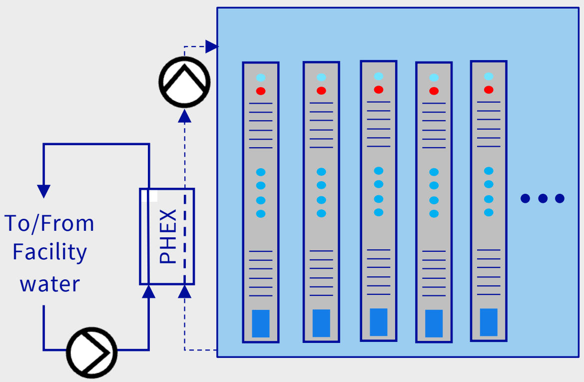
24 U/m²

+ 25% →

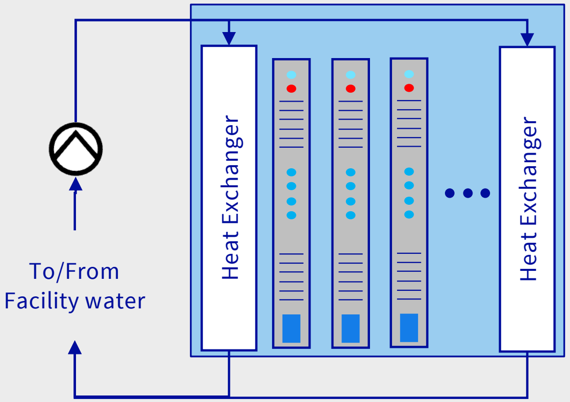
OVHcloud layout
No cold aisle / hot aisle containment
48 U **horizontal** racks (stackable up to 4 levels)
128 racks hosted in a 207 m² server room

30 U/m²

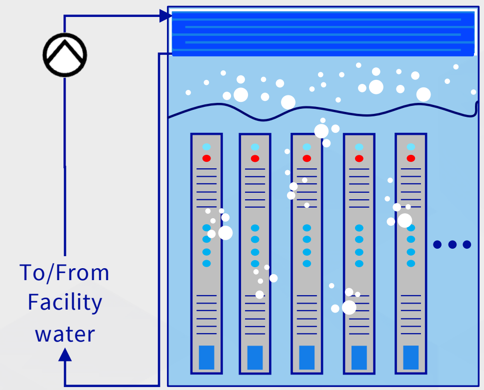
Immersive cooling : the buzz word ? Overview of current technologies



Forced convection

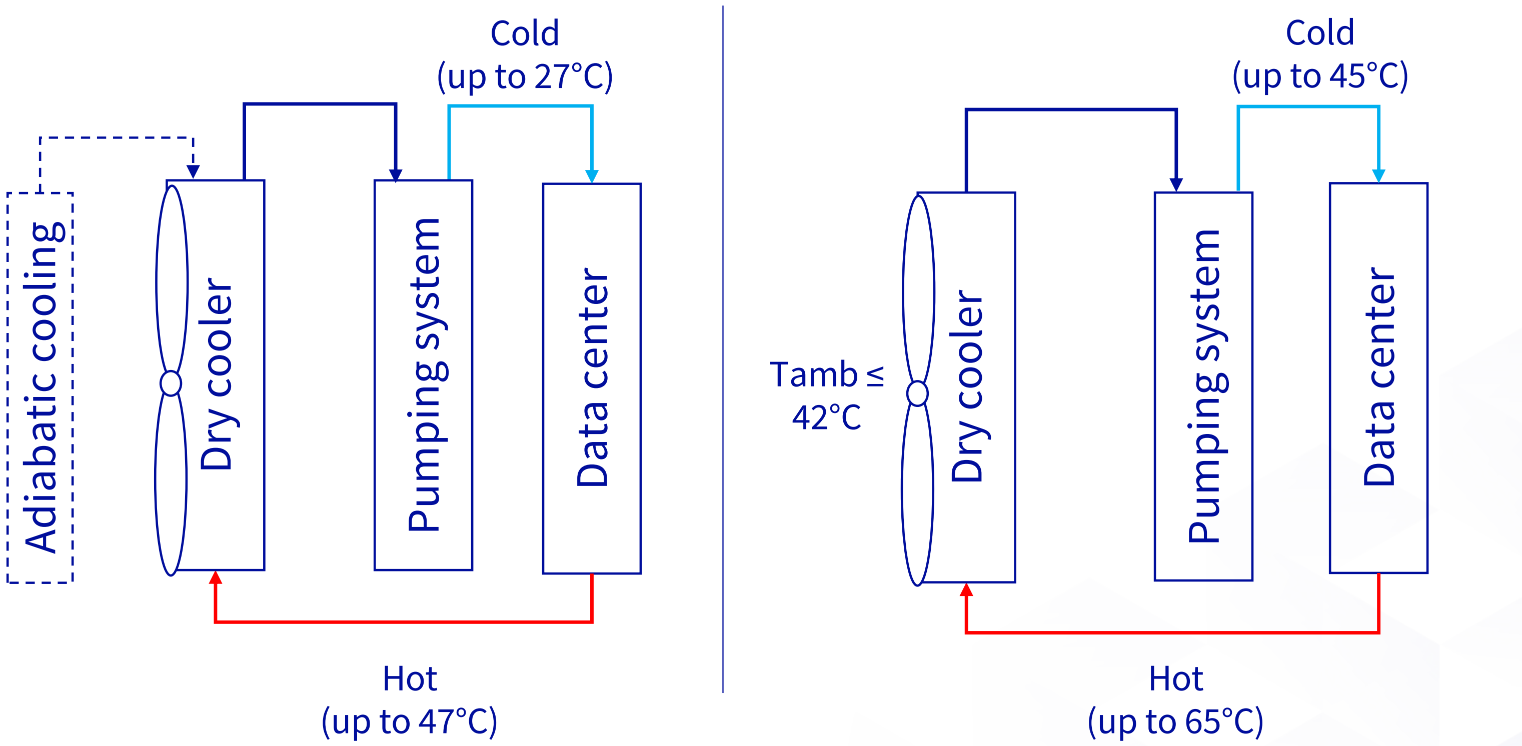


Natural convection



Single-phase	Single-phase	Two-phase
Environmentally friendly nonvolatile fluids	Environmentally friendly nonvolatile fluids	Volatile fluids with high environmental impact
Sealed cover not required	Sealed cover not required	Sophisticated sealed cover is required
Local pumping system	High electrical efficiency	High electrical efficiency
Typical 48U rack	Only 24U per rack	Typical 48U rack
Inflammable (@ T>159 °C)	Inflammable (@ T>200°C)	Not inflammable

OVHcloud hybrid cooling concept

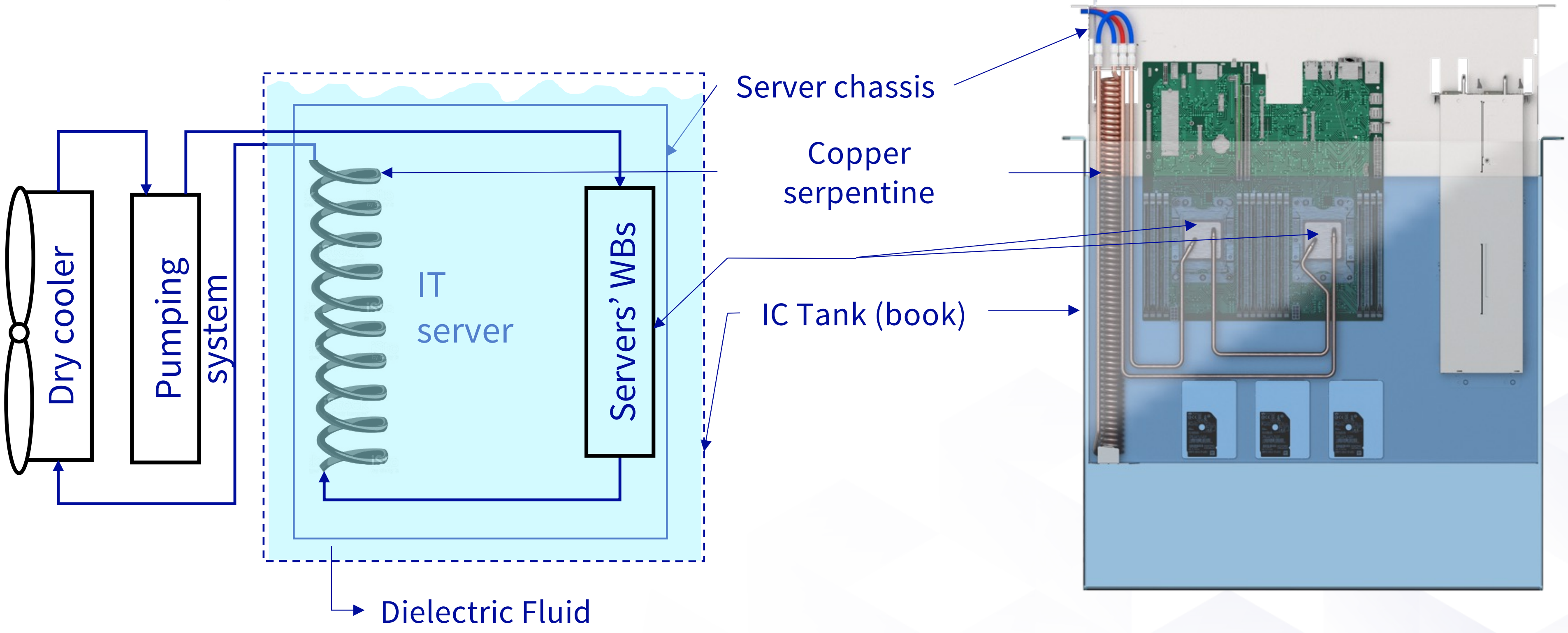


OVHcloud liquid cooling

OVHcloud Hybrid Cooling

- Sealed tank not required
- High efficiency
- Environmentally friendly
- Adiabatic cooling elimination
- Better energy recovery
- ↙ PUE and WUE
- Silent DC

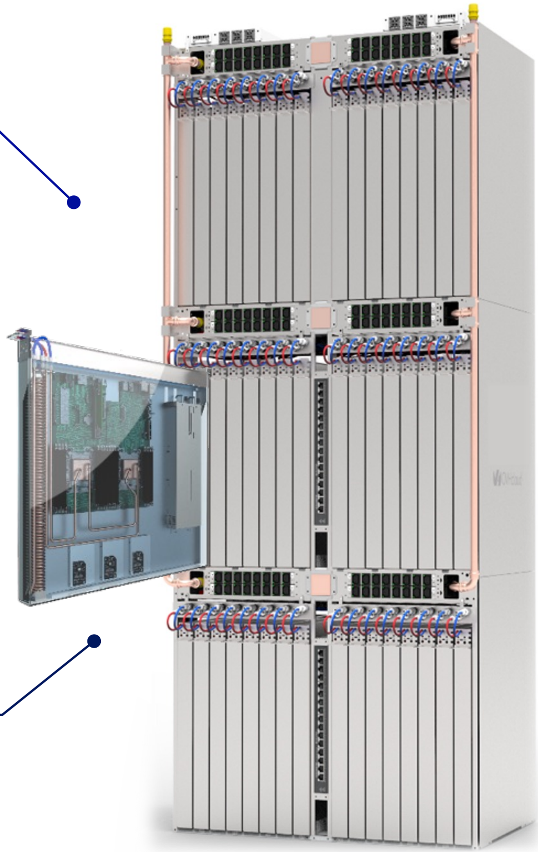
OVHcloud hybrid server concept



OVHcloud hybrid cooling kills 3 birds with 1 stone

PPUE 1.004-1.006
WUE tends to zero

High footprint density (U/m2)



Energy recovery
Feasible at high temperatures

05 What about the usages ?

AI : another buzz word ? Numbers are stunning

1283 MWh	ChatGPT3 training (⇔ 600 French households)*
206 GWh	ChatGPT inference (⇔ 90 000 French households)*
2.9 vs 0.3 Wh	ChatGPT vs Google request*
500 000	# of H100 produced by NVIDIA in 2023 (1 500 000 in 2024)**
85-134 TWh	Projected consumption (⇔ Belgium-Norway)*
½ litre of water	ChatGPT typical prompt***

**Source – Dr Alex de Vries VU Amsterdam School of Business and Economics*
***Source – Financial Times*
****Source – OECD report / Pr Shaolei Ren University of California Riverside*

OVHcloud GPU installed base and findings

Close to **30 000** GPU installed already (NVIDIA, AMD)

Introduced massively in 2021 for the gaming industry

In average severs with GPU dedicated to AI consume **5x** as much electricity as the average one

Typical LCA (from cradle to gate without uncertainty)* :

• Intel CPU range	5 – 25 kgCO2e
• NVIDIA GPU L4	50 kgCO2e
• NVIDIA GPU L40s	100 kgCO2e
• NVIDIA GPU A100	150 kgCO2e
• NVIDIA GPU H100	150 kgCO2e

**Source – Intel PCF / OVHcloud LCA*

Example of arbitration for Public Cloud platforms

Extending lifespan of GPU through PCI compute platforms

NVIDIA® Tesla V100
(Parallel computing)
June 21st 2017

Public Cloud Compute

Nom	Mémoire	vCore	GPU	Stockage	Réseau public	Réseau privé
t1-45	45 Go	8	Tesla V100 16 Go	400 Go NVMe	2 Gbit/s garantis	4 Gbit/s max.

NVIDIA® Quadro RTX5000
(Gaming)
August 13th 2018

Example of arbitration for AI platforms

Choosing the right location to train the models

AI PaaS (NVIDIA® H100)
Poland
24.3 tCO_2e (over 3 years)

- 85%

AI PaaS (NVIDIA® H100)
Canada
3.6 tCO_2e (over 3 years)

Poland

kCO_2e/month
(Location based)

- Manufacturing 73
- Operations 4
- Electricity 598

Canada

kCO_2e/month
(Location based)

- Manufacturing 73
- Operations 4
- Electricity 23

Source – OVHcloud carbon calculator

Example of arbitration for Private Cloud platforms

Choosing the right performance even for the latest range of dedicated servers

ADV4 – GEN3
France
2.6 tCO2e (over 3 years)

AMD EPYC™ 4584PX
128GB
2x SSD NVMe 960GB

- 47% →

ADV1 - GEN3
France
1.4 tCO2e (over 3 years)

AMD EPYC™ 4244P
32GB
2x SSD NVMe 960GB

kCO2e/month
(Location based)

- Manufacturing 41
- Operations 4
- Electricity 26

kCO2e/month
(Location based)

- Manufacturing 8
- Operations 4
- Electricity 26

Source – OVHcloud carbon calculator

Example of arbitration for Private Cloud platforms

Choosing second/third life servers for non-demanding workloads

ADV1 – GEN3
France
1.4 tCO_2e (over 3 years)

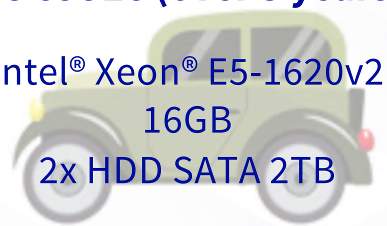
AMD EPYC™ 4244P
32GB
2x SSD NVMe 960GB

kCO2e/month
(Location based)

- Manufacturing 8
- Operations 4
- Electricity 26

- 65% →

KIMSUFI
France
0.5 tCO_2e (over 3 years)



Intel® Xeon® E5-1620v2
16GB
2x HDD SATA 2TB

kCO2e/month
(Location based)

- Manufacturing 4
- Operations 4
- Electricity 4

Source – OVHcloud carbon calculator

Example of arbitration for Cold Storage platforms

For cold storage instances, archiving on tapes is a sustainable option

OCP JBOD
3576 tCO2e (over 8 years)

- 90%

IBM TS4500
369 tCO2e (over 8 years)

110PB
5-year life
16TB Archive drives
3 JBOD per Controller
1.26 Erasure Coding
Assumes 20TB drives

110 PB
15-year life
75TB/HR Data Rate
60 Tape drives
6112 Cartridges

Source – IBM LCA published by Shawn O. Brume Sc.D.