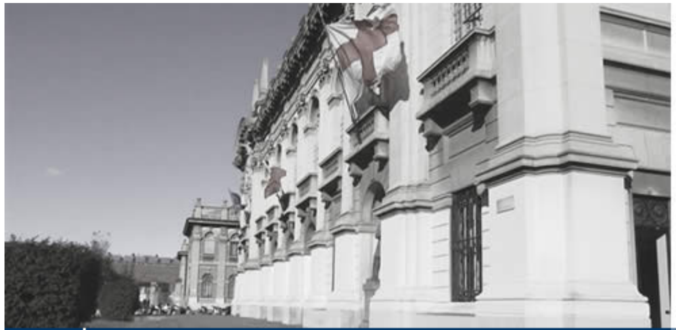


Computing Infrastructures

POLITECNICO DI MILANO

The Datacenter as a Computer: Building and Infrastructures

The topics of the course: what are we going to see today?

HW Infrastructures:

System-level: Computing Infrastructures and Data Center Architectures, Rack/Structure;

Node-level: Server (computation, HW accelerators), Storage (Type, technology), Networking (architecture and technology);

Building-level: Cooling systems, power supply, failure recovery

SW Infrastructures:

Virtualization: Process/System VM, Virtualization Mechanisms (Hypervisor, Para/Full virtualization)

Computing Architectures: Cloud Computing (types, characteristics), Edge/Fog Computing, X-as-a service

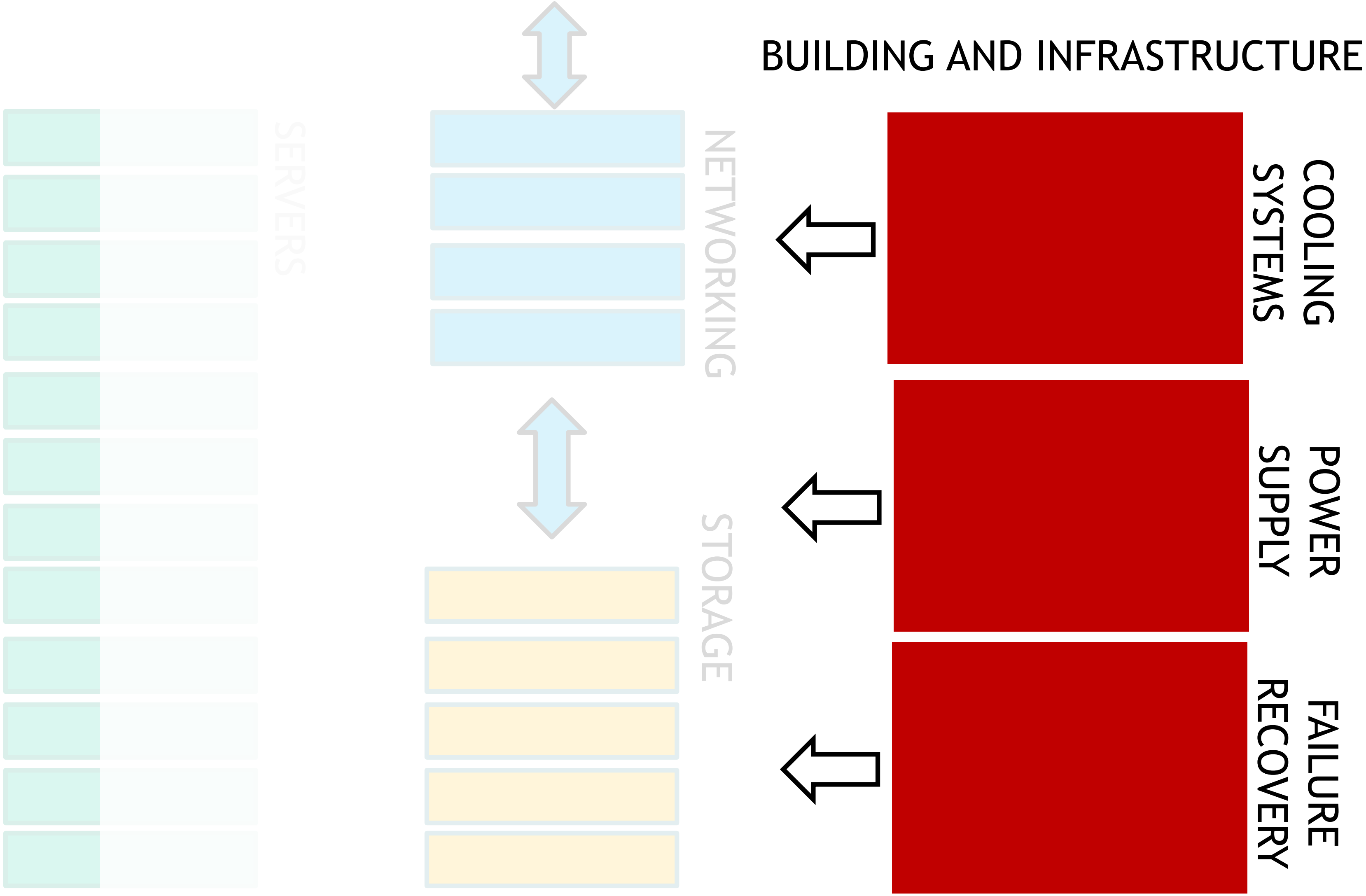
Methods:

Reliability and availability of datacenters (definition, fundamental laws, RBDs)

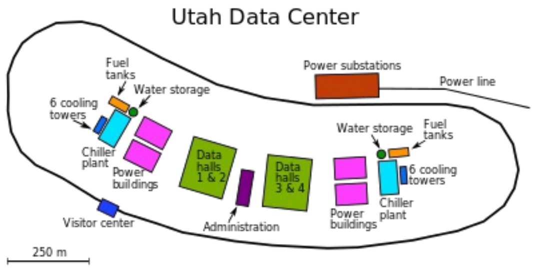
Disk performance (Type, Performance, RAID)

Scalability and performance of datacenters (definitions, fundamental laws, queuing network theory)

BUILDING AND INFRASTRUCTURE

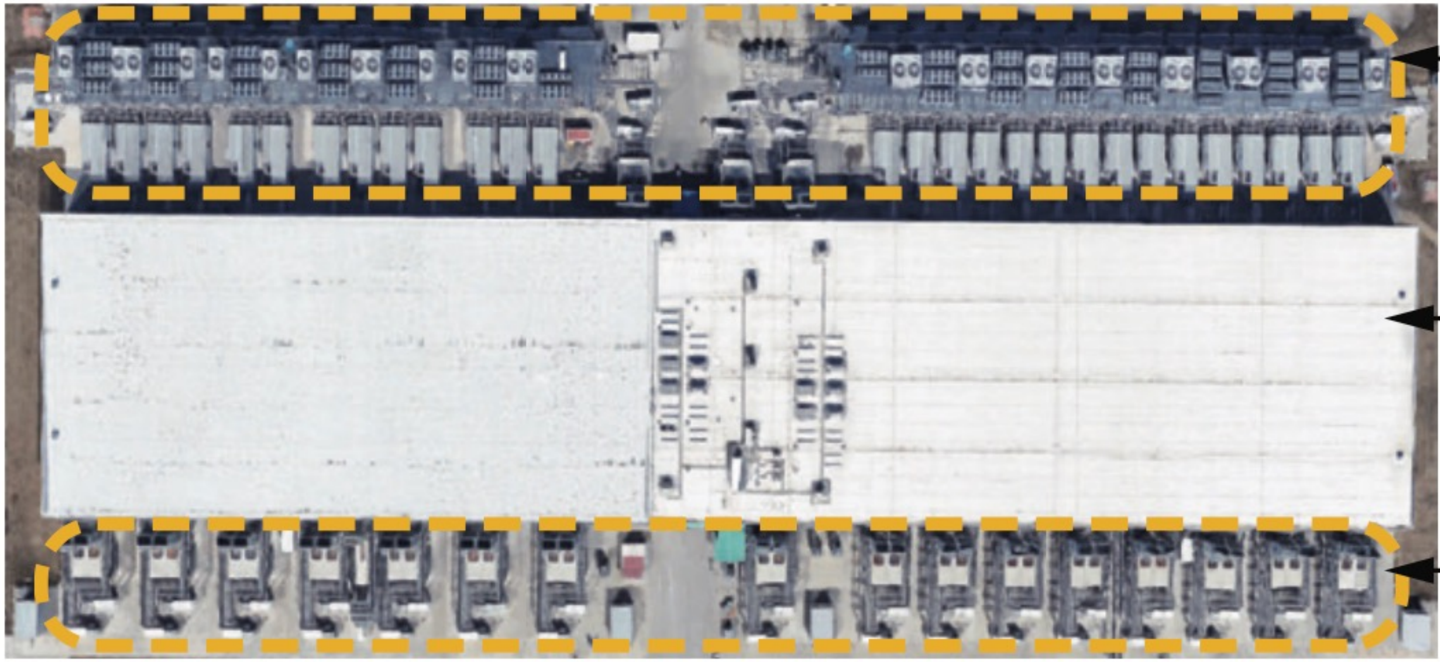


Aerial view of a Google data center campus in Iowa (US)



We already saw something similar: NSA UTHA DC

A Google data center building



The main components of a typical data center

COOLING SYSTEM

POWER SUPPLY

COMPUTATION
STORAGE
NETWORKING

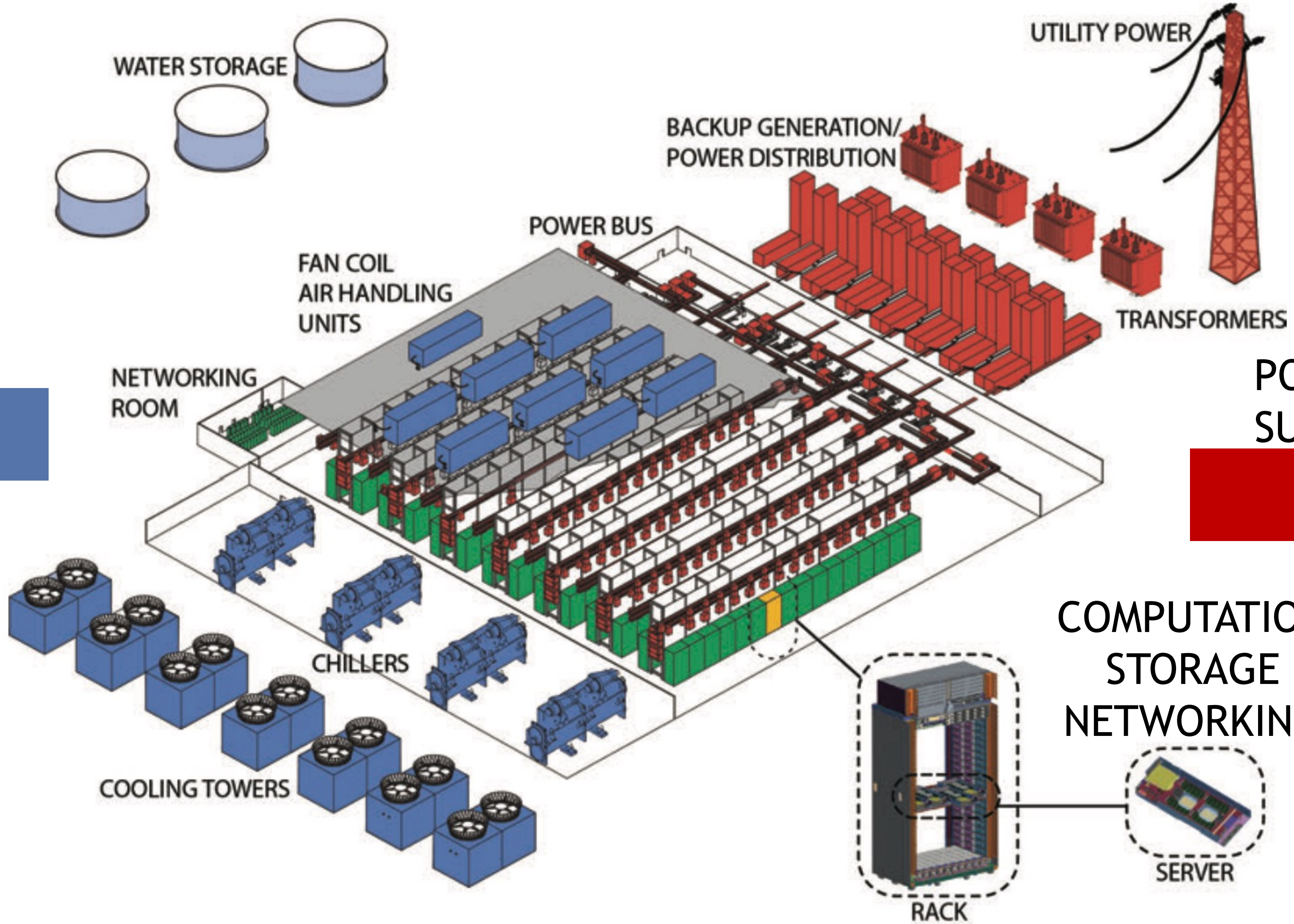
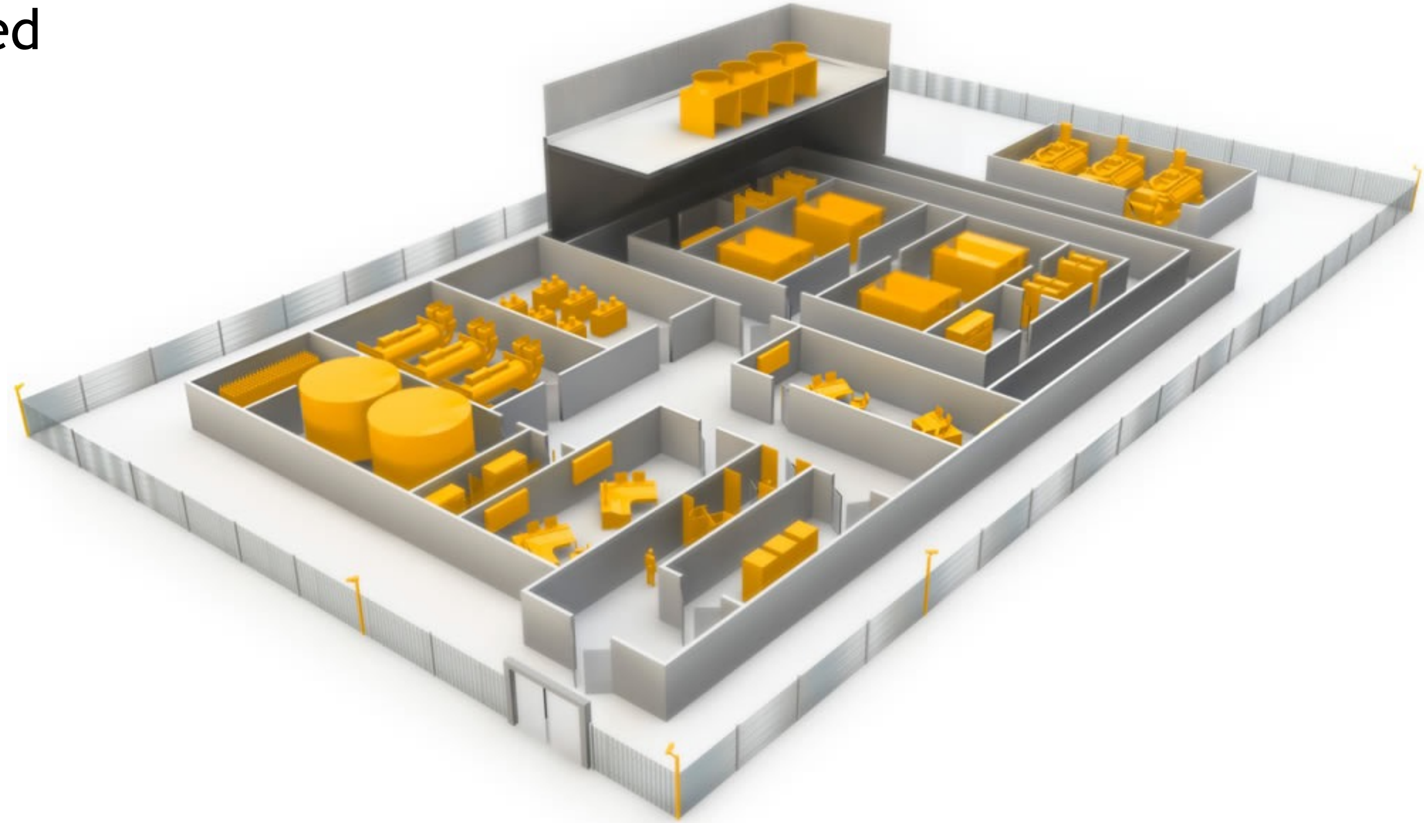


Image taken from Barroso

Not just computation, storage and networking

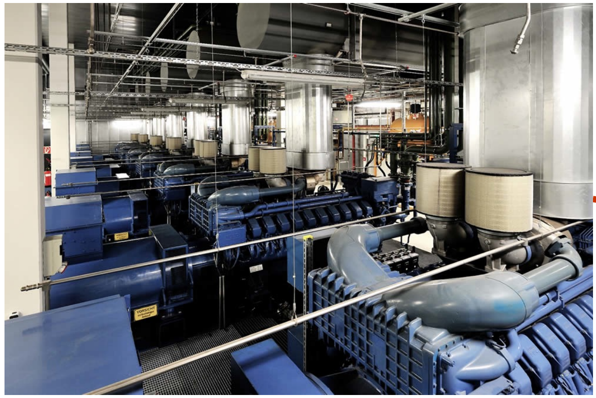
- WSC/DCs has other important components related to **power delivery**, **cooling**, and **building infrastructure** that also need to be considered



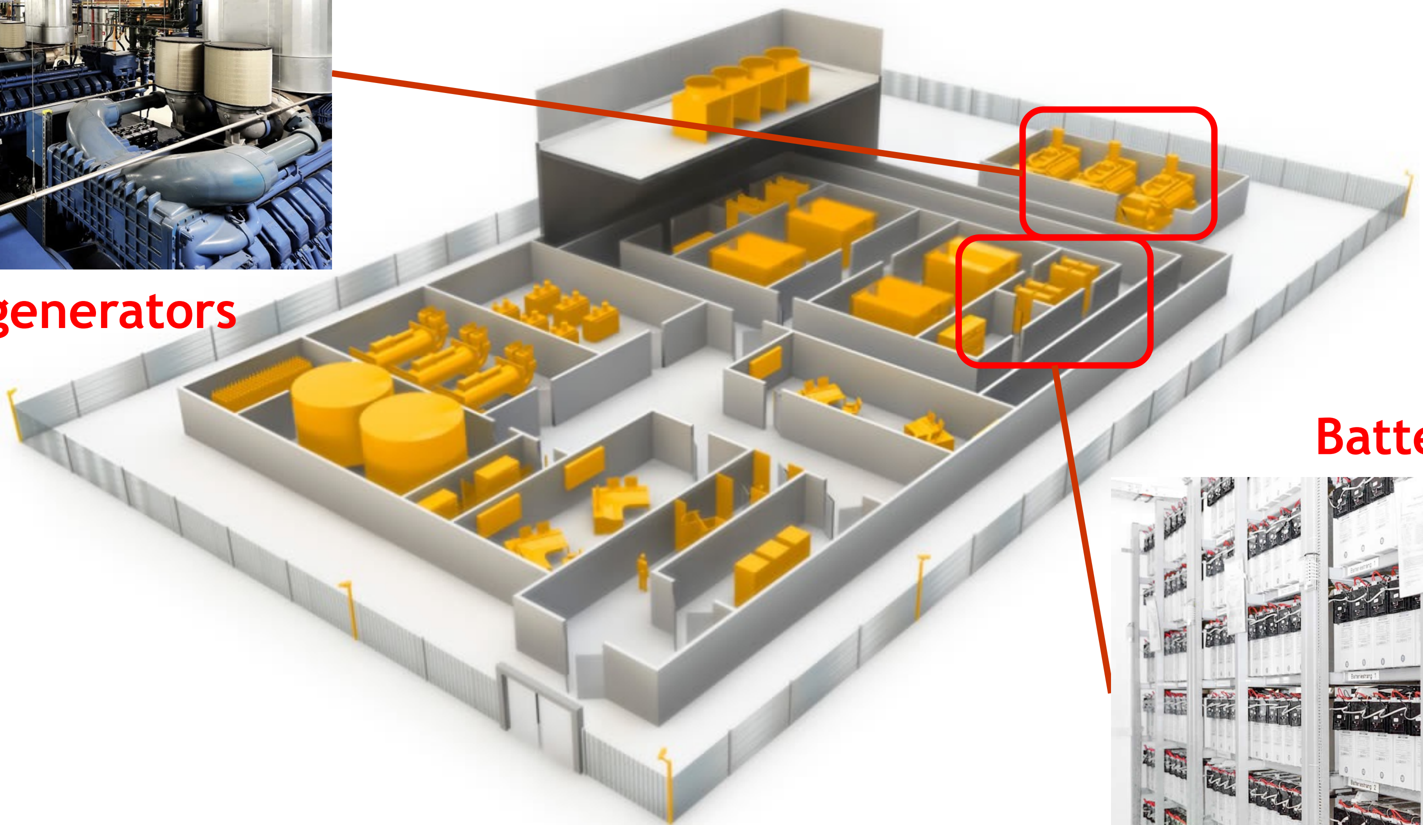
15Years Old HPC center – IT4Innovation Czech Supercomputing Center
https://www.it4i.cz/en/infrastructure/operating-technologies

DATA CENTER POWER SYSTEMS

In order to protect against power failure, battery and diesel generators are used to backup the external supply.



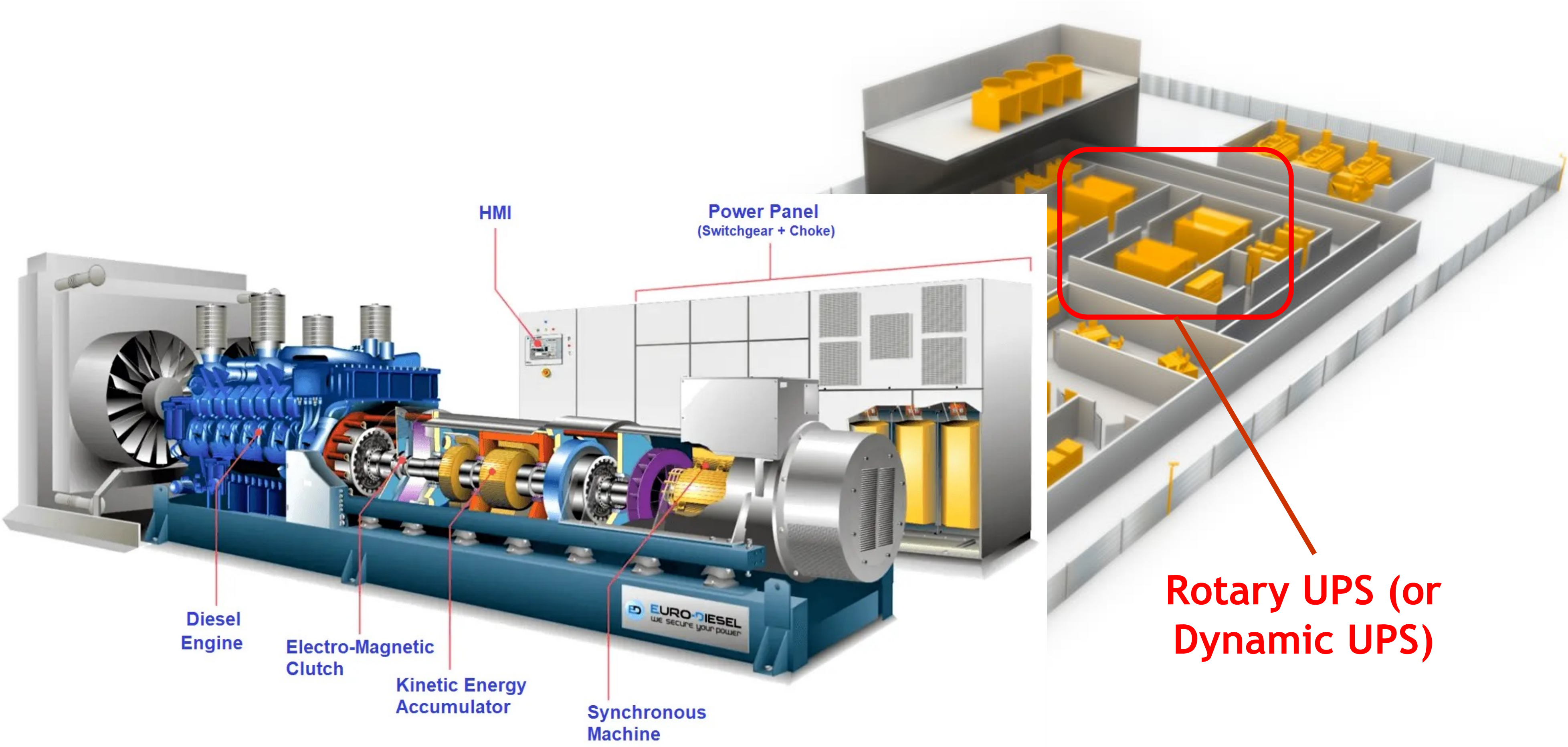
Diesel generators



Batteries

DATA CENTER POWER SYSTEMS

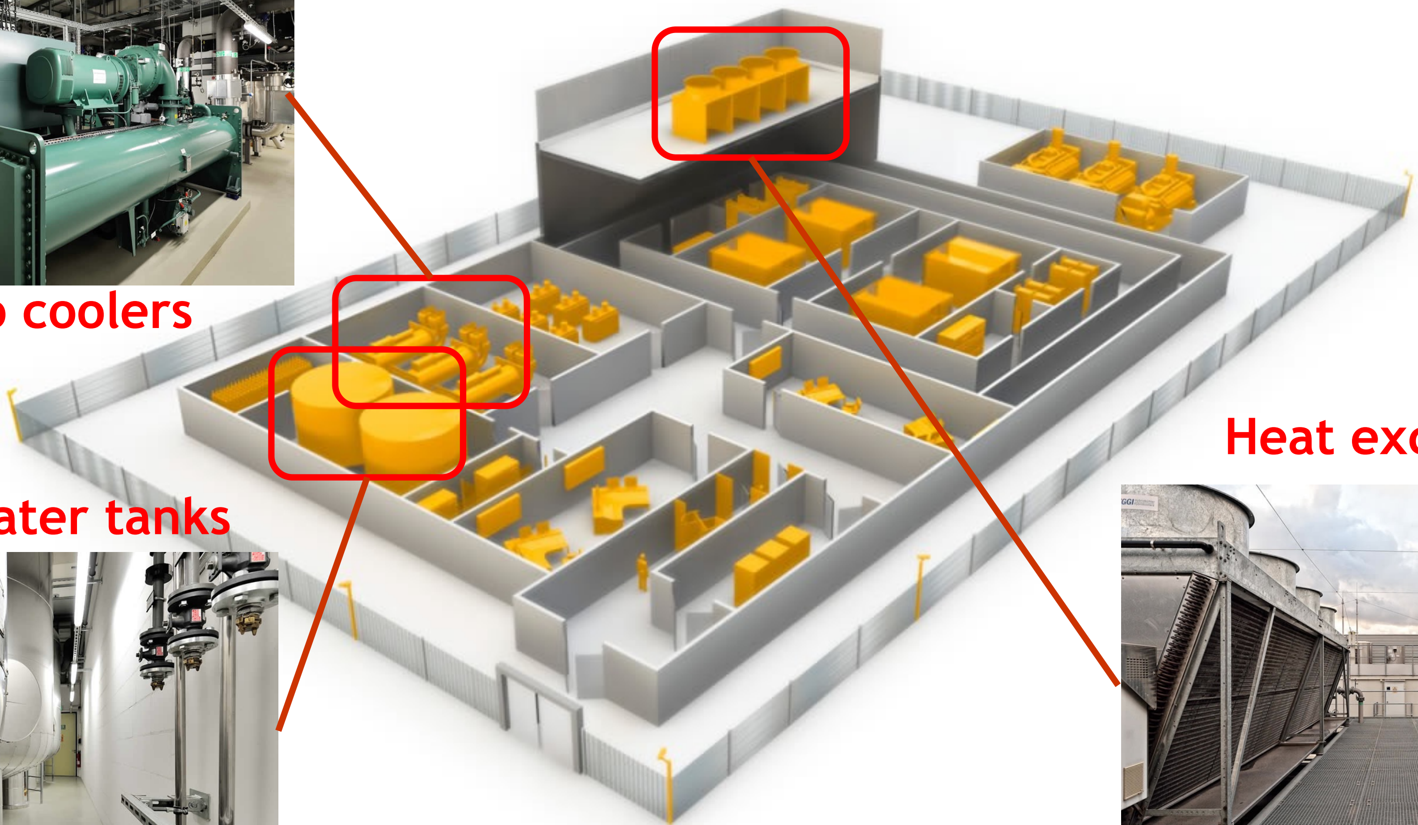
- **Rotary UPS Systems.**
- They store electrical energy in the form of kinetic energy.



DATA CENTER COOLING SYSTEMS

IT equipment generates **a lot of heat**: the **cooling system** is usually a very expensive component of the datacenter, and it is composed by coolers, heat-exchangers and cold water tanks.

Turbo coolers

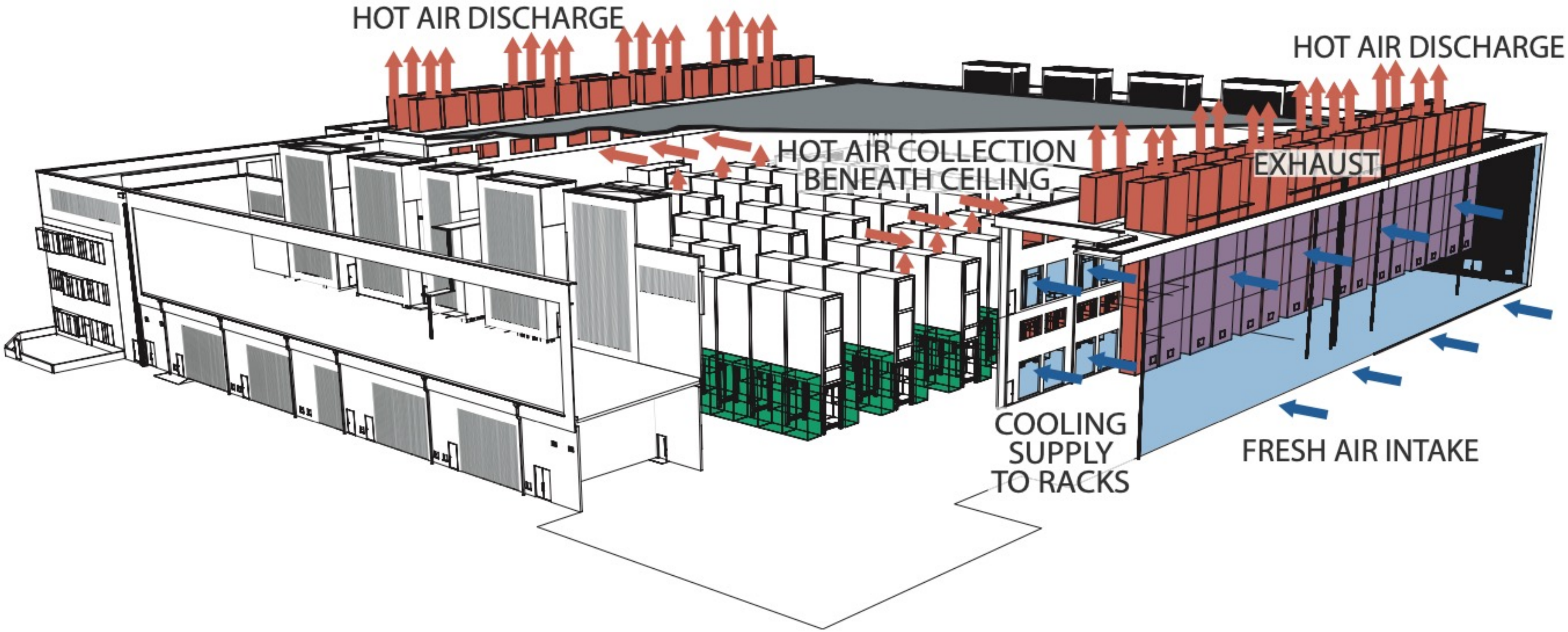


Heat exchangers

Cold water tanks

Open-Loop

- The simplest topology is fresh air cooling (or air economization)—essentially, opening the windows.
- This is a single «open-loop» system



Open vs Closed Loop

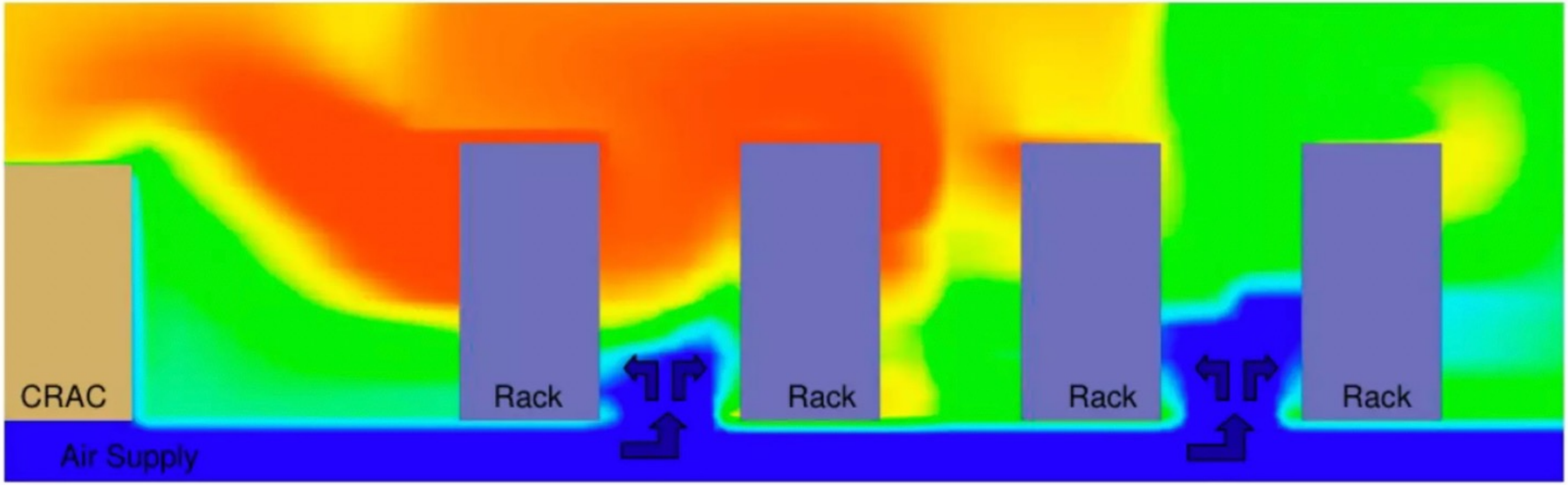
- Free cooling, i.e., **open-loop**, refers to the use of cold outside air to either help the production of chilled water or directly cool servers. It is not completely free in the sense of zero cost, but it involves very low-energy costs compared to chillers.

- **Closed-loop** systems come in many forms, the most common being the air circuit on the data center floor.
  - The goal is to isolate and remove heat from the servers and transport it to a heat exchanger.
  - Cold air flows to the servers, heats up, and eventually reaches a heat exchanger to cool it down again for the next cycle through the servers.

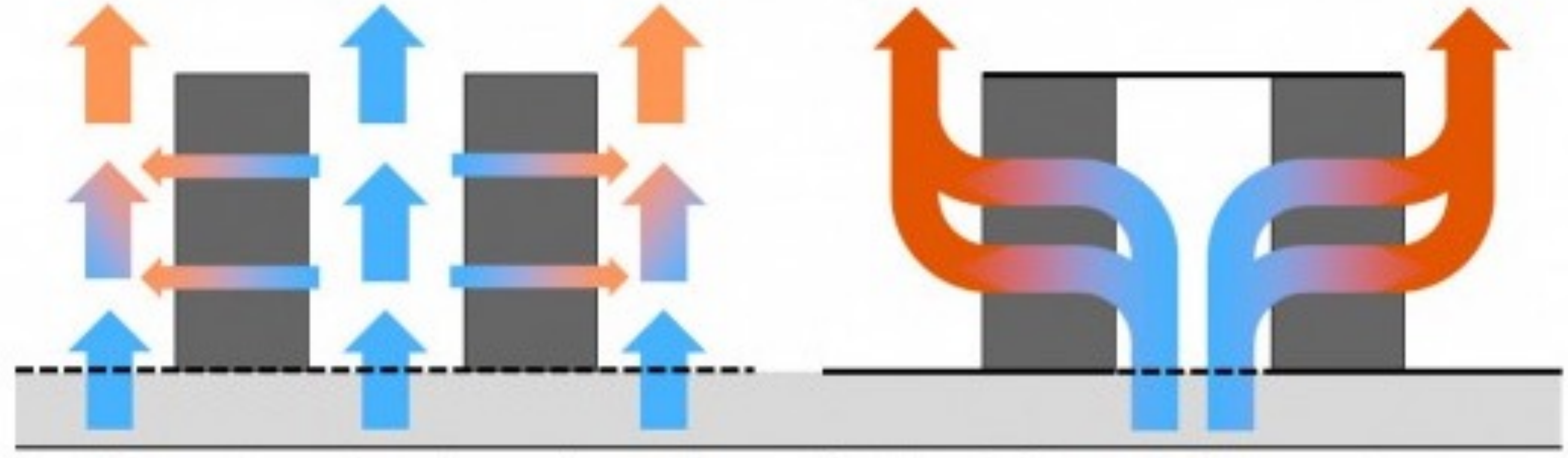
Data-center corridors

- Server Racks are **NEVER BACK-to-BACK**
- Corridors where servers are located are split into *cold aisle*, where the front panels of the equipment is reachable, and *warm aisle*, where the back connections are located
- Cold air flows from the front (cool aisle), cools down the equipment, and leave the room from the back (warm aisle)

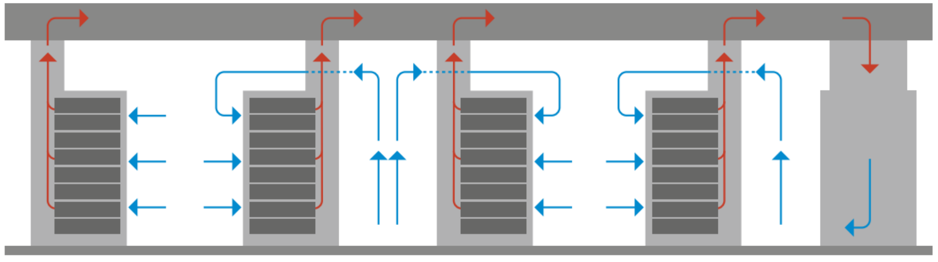
Hot/Cold aisle are not enough

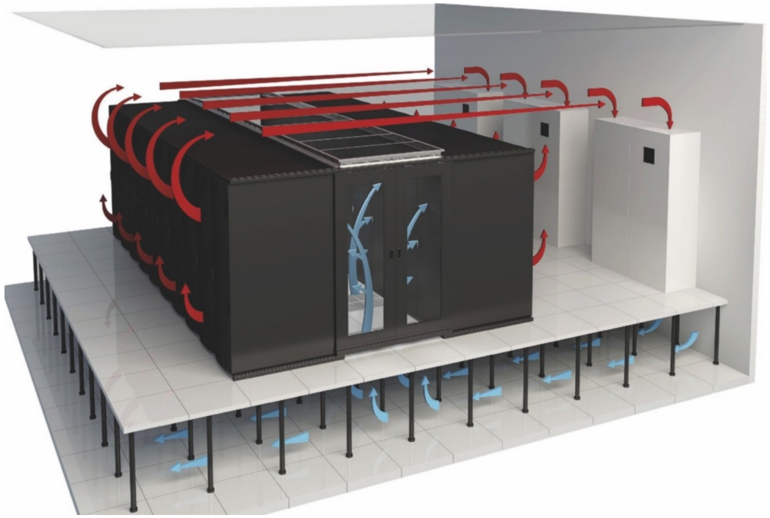


Hot/Cold aisle are not enough



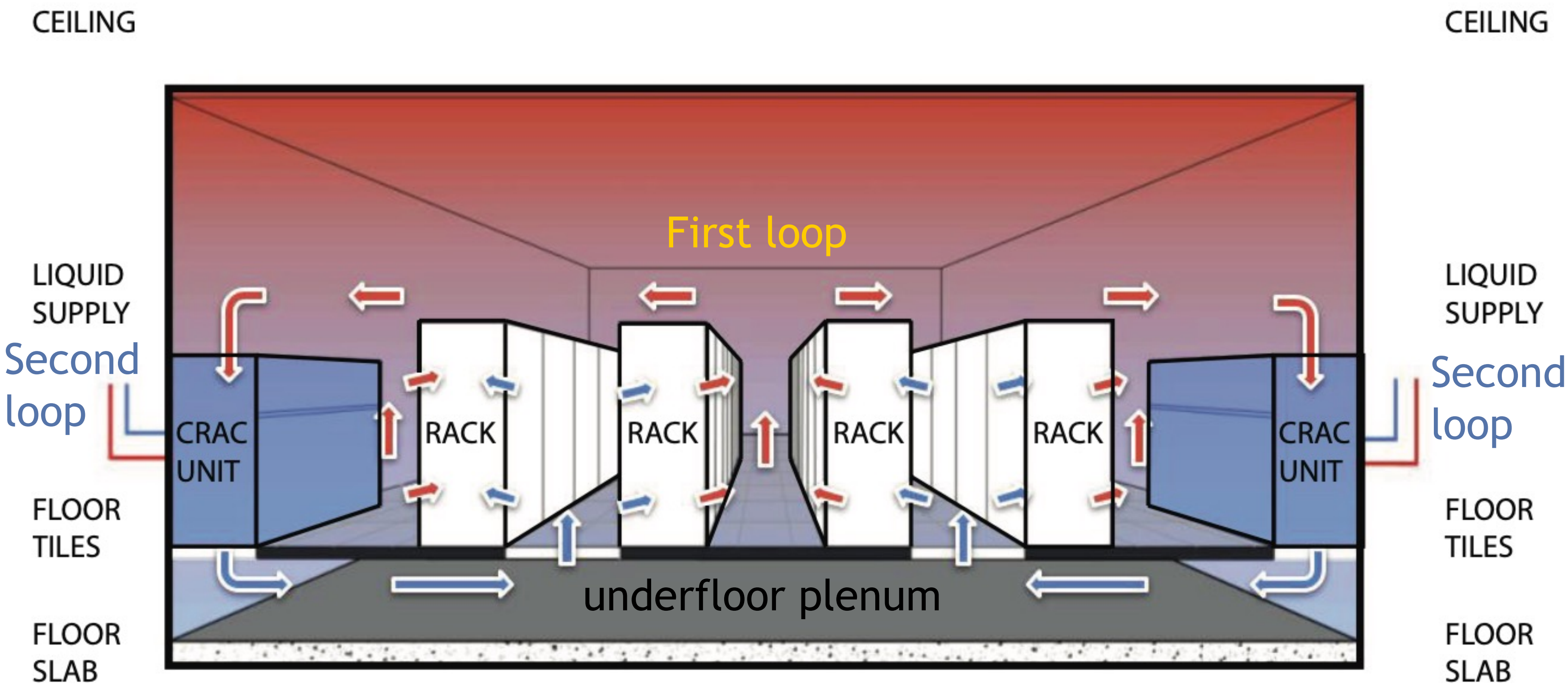
Not Unique Solution



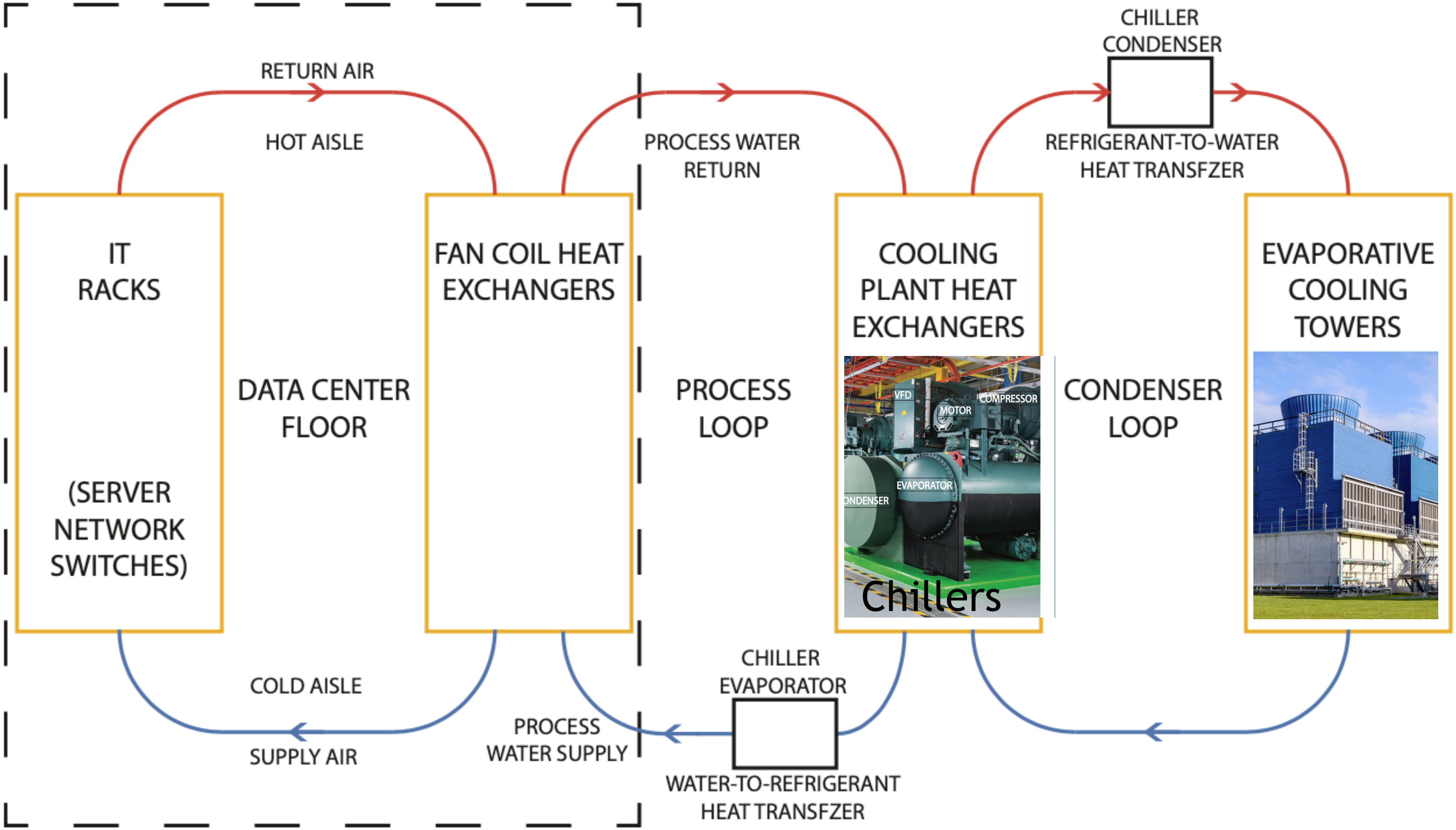


Closed-loop with two loops

- ✓ The airflow through the underfloor plenum, the racks, and back to the CRAC (a 1960s term for *computer room air conditioning*) defines the primary air circuit, i.e., the first loop.
- ✓ The second loop (the liquid supply inside the CRACs units) leads directly from the CRAC to external heat exchangers (typically placed on the building roof) that discharge the heat to the environment.



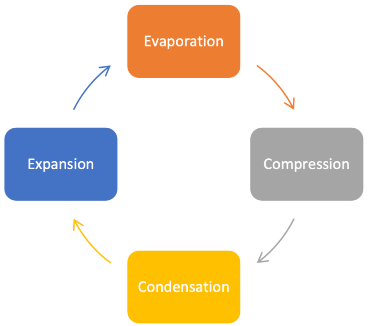
A three-loop system commonly used in large-scale data center



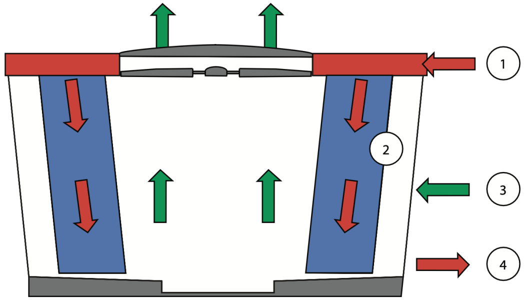
First loop Second loop Third loop

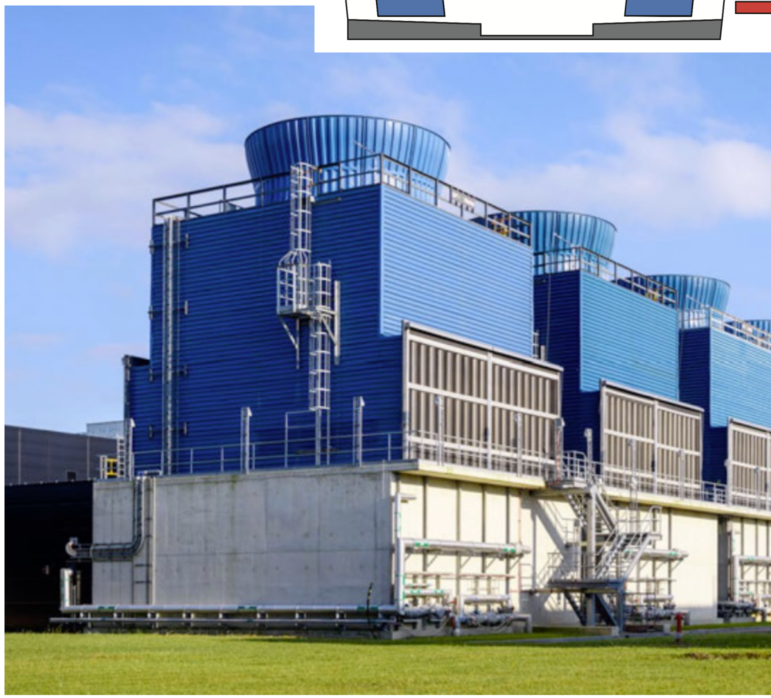
Chillers and Cooling Towers

A water-cooled chiller can be thought of as a water-cooled air conditioner



Refrigeration Cycle





Cooling towers cool a water stream by evaporating a portion of it into the atmosphere. They do not work as well in very cold climates because they need additional mechanisms to prevent ice formation

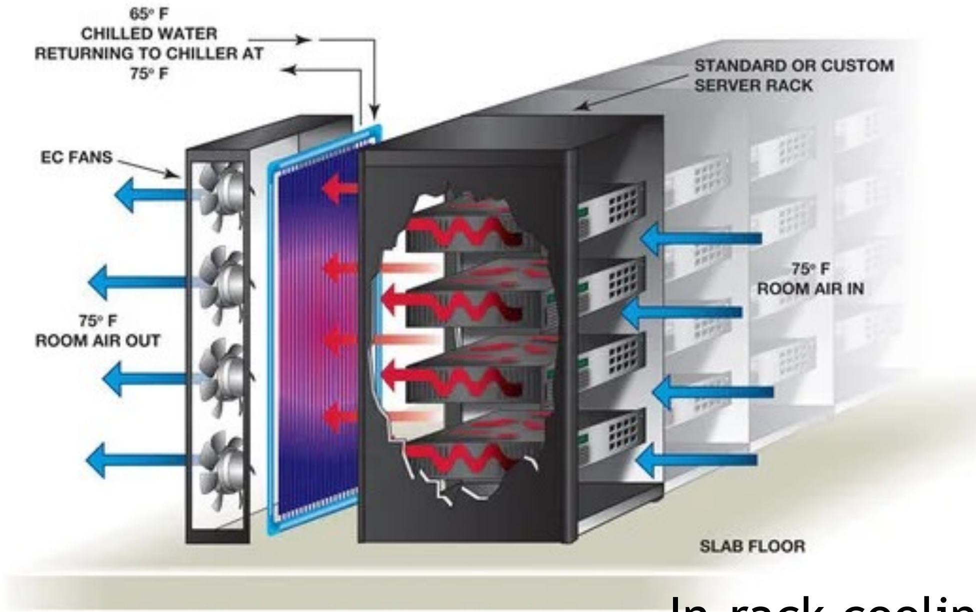
A critical comparison

Each topology presents tradeoffs in complexity, efficiency, and cost:

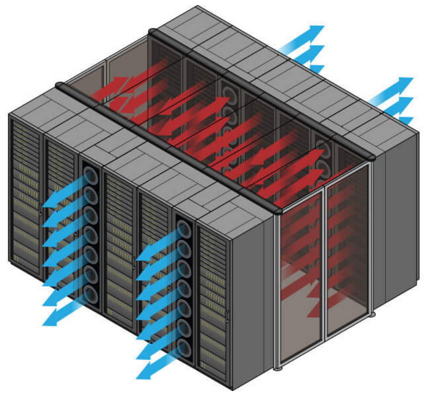
- ✓ **Fresh air cooling** can be very efficient but does not work in all climates, requires filtering of airborne particulates, and can introduce complex control problems.
- ✓ **Two-loop systems** are easy to implement, relatively inexpensive to construct, and offer isolation from external contamination, but typically have lower operational efficiency.
- ✓ **A three-loop system** is the most expensive to construct and has moderately complex controls, but offers contaminant protection and good efficiency.

What's next? In-rack, In-row Cooling

- **In-rack cooler** adds an air-to-water heat exchanger at the back of a rack so the hot air exiting the servers immediately flows over coils cooled by water, essentially reducing the path between server exhaust and CRAC input
- **In-row cooling** works like in-rack cooling except the cooling coils aren't in the rack, but adjacent to the rack.



In-rack cooling

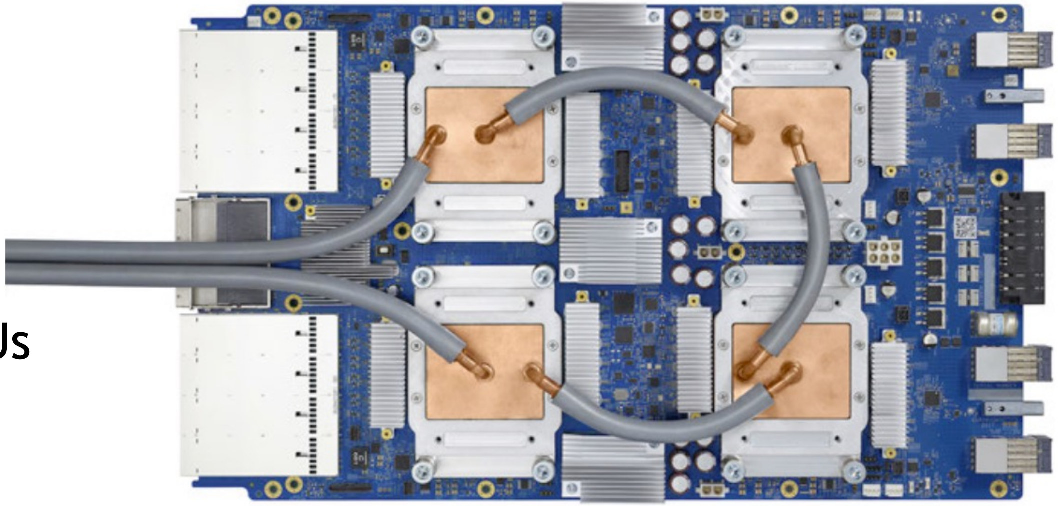


In-row cooling

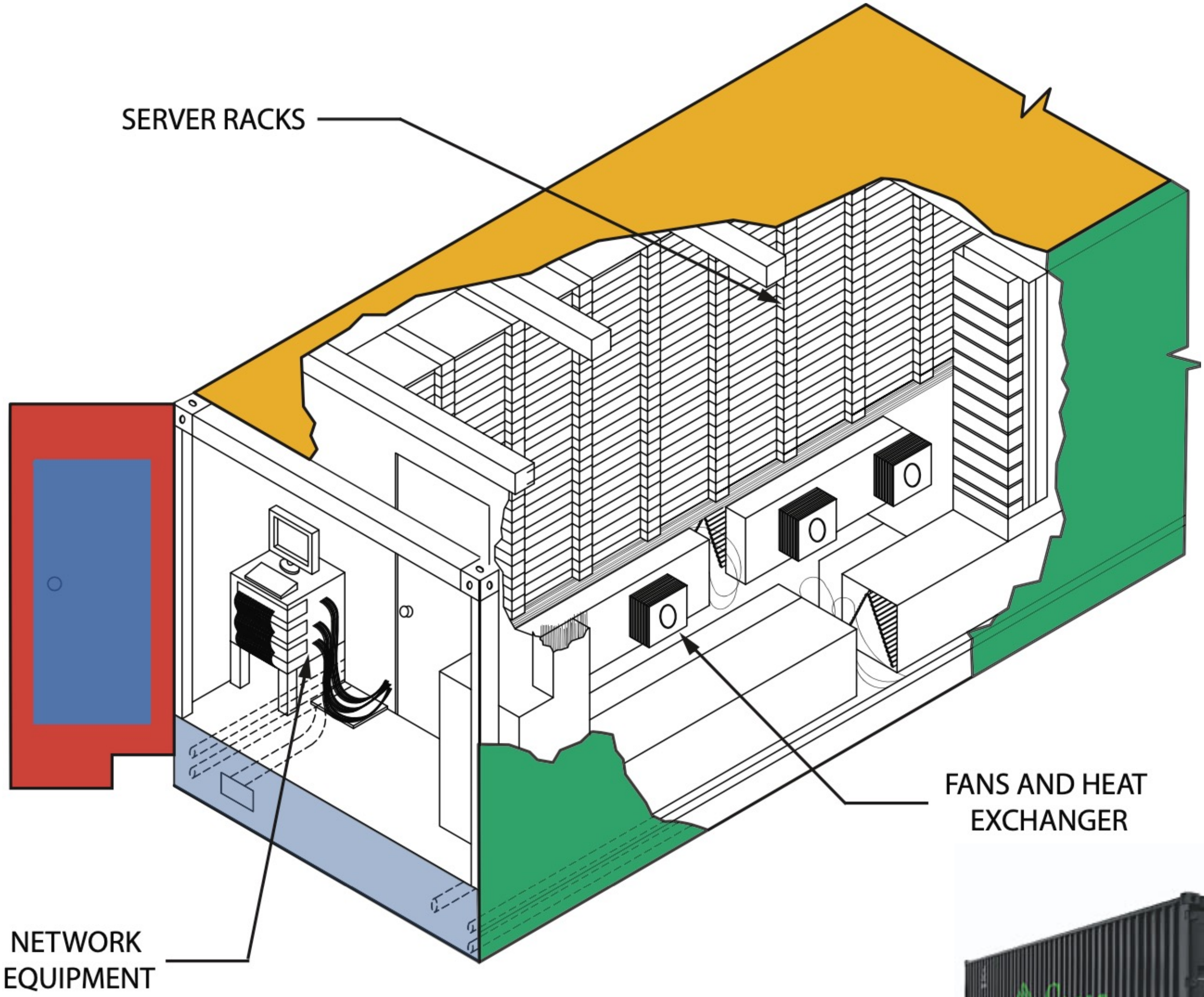
Liquid cooling

- We can directly cool server components using cold plates, i.e., local liquid-cooled heat sinks:
  - Impractical to cool all compute components with cold plates.
  - Components with the highest power dissipation are targeted for liquid cooling while other components are air-cooled.
- The liquid circulating through the heat sinks transports the heat to a liquid-to-air or liquid-to-liquid heat exchanger that can be placed close to the tray or rack, or be part of the data center building (such as a cooling tower).

Google TPUs



Modular Container-based Data Centers



Container-based data centers place the server racks inside a container (typically 6 to 12 mt long) and integrating heat exchange and power distribution into the container as well.

Data-center power consumption

- Data-center power consumption is an issue, since it can reach several MWs.
- Cooling usually requires about half the energy required by the IT equipment (servers + network + disks).
- Energy transformation creates also a large amount of energy wasted for running a datacenter.

Example numbers from the Book

Amortized Cost	Component	Sub-Components
~45%	Servers	CPU, memory, disk
~25%	Infrastructure	UPS, cooling, power distribution
~15%	Power draw	Electrical utility costs
~15%	Network	Switches, links, transit

PUE Metric

- **Power usage effectiveness** (PUE) is the ratio of the total amount of energy used by a DC facility to the energy delivered to the computing equipment

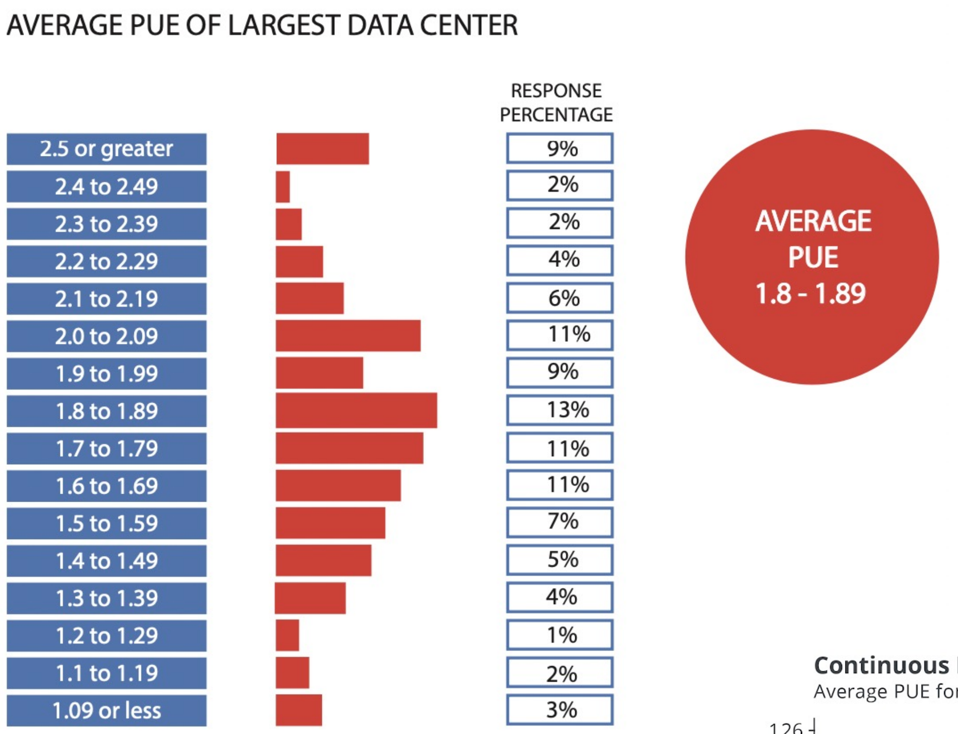
$$PUE = \frac{Total\ Facility\ Power}{IT\ Equipment\ power}$$

- Total facility power = covers IT systems (servers, network, storage) + other equipment (cooling, UPS, switch gear, generators, lights, fans, etc.)

- Data Center infrastructure Efficiency (DCiE): PUE inverse

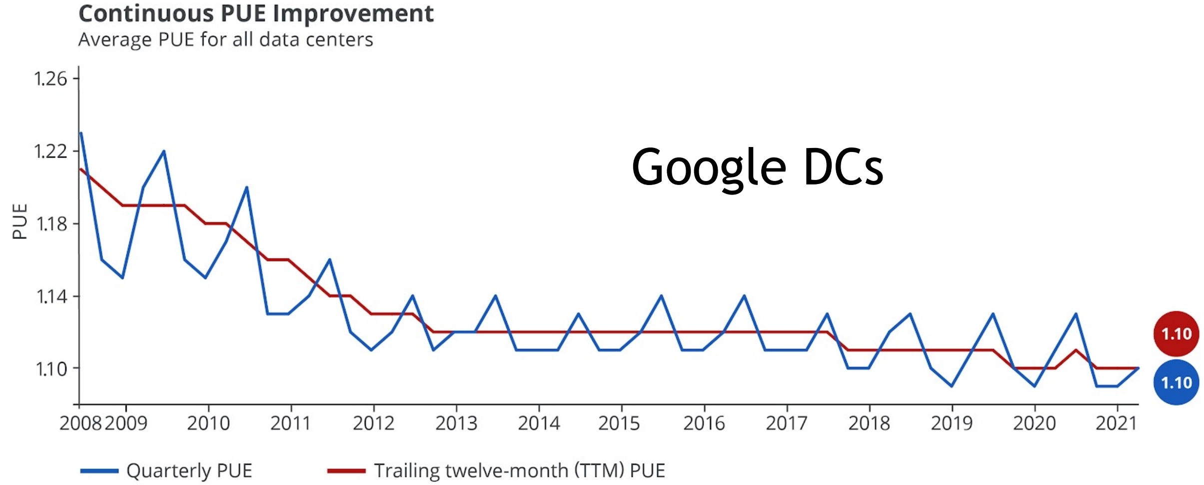
- ***Not a ONE FITS ALL metric for ENERGY EFFICIENCY***

PUE Metric



2012 Study

PUE	DCiE	Level of Efficieny
3.0	33%	Very Inefficient
2.5	40%	Inefficient
2.0	50%	Average
1.5	67%	Efficient
1.2	83%	Very Efficient



Google DCs

Interesting Video on Youtube by Ana Radovanovic: «Carbon-Aware Computing at Google»

Data-center tiers

- Data-center availability is defined by in four different tier level. Each one has its own requirements.

Tier Level	Requirements	
1	• Single non-redundant distribution path serving the IT equipment • Non-redundant capacity components • Basic site infrastructure with expected availability of 99.671%	29h
2	• Meets or exceeds all Tier 1 requirements • Redundant site infrastructure capacity components with expected availability of 99.741%	22,5h
3	• Meets or exceeds all Tier 2 requirements • Multiple independent distribution paths serving the IT equipment • All IT equipment must be dual-powered and fully compatible with the topology of a site's architecture • Concurrently maintainable site infrastructure with expected availability of 99.982%	1,5h
4	• Meets or exceeds all Tier 3 requirements • All cooling equipment is independently dual-powered, including chillers and heating, ventilating and air-conditioning (HVAC) systems • Fault-tolerant site infrastructure with electrical power storage and distribution facilities with expected availability of 99.995%	0,5h

Data-center tiers

- Data-center availability is defined by in four different tier level. Each one has its own requirements.

Tier Level	Requirements	
1	• Single non-redundant distribution path serving the IT equipment • Non-redundant capacity components	
3	• Multiple independen • All IT equipment mu ... ogy of a site's architecture • Concurrently mainta ... 9.982%	1,5h
4	• Meets or exceeds a • All cooling equipme ... d heating, ventilating and air-conditioning (HVAC) • Fault-tolerant site infrastructure with electrical power storage and distribution facilities with expected availability of 99.995%	0,5h

Different from tiers definition in European HPC facilities Supercomputers:
- Tier-0 are European Centres (Highly ranked in Top500)
- Tier-1 are national centres, (Ranked in Top500)
- Tier-2 are regional centres