AIDC Infrastructure Revolution and Liquid Cooling Application

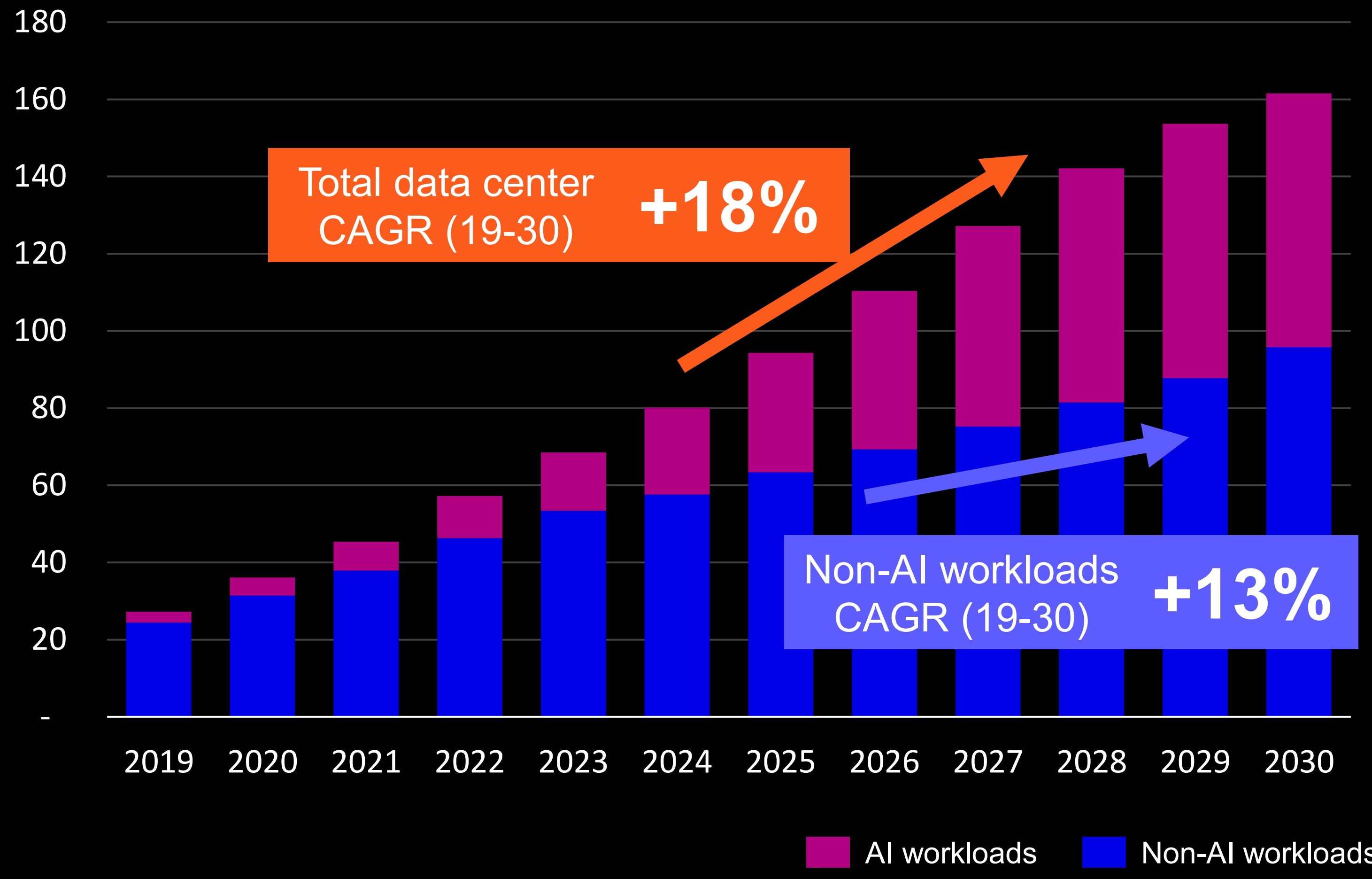
Teddy Zhang

AIDC Chief Researcher of Deep Knowledge Community

In the data center business, AI will accelerate **capacity expansion**.

AI workloads are incremental to conventional IT loads, not substitutive.

Total data center installed capacity (GW)

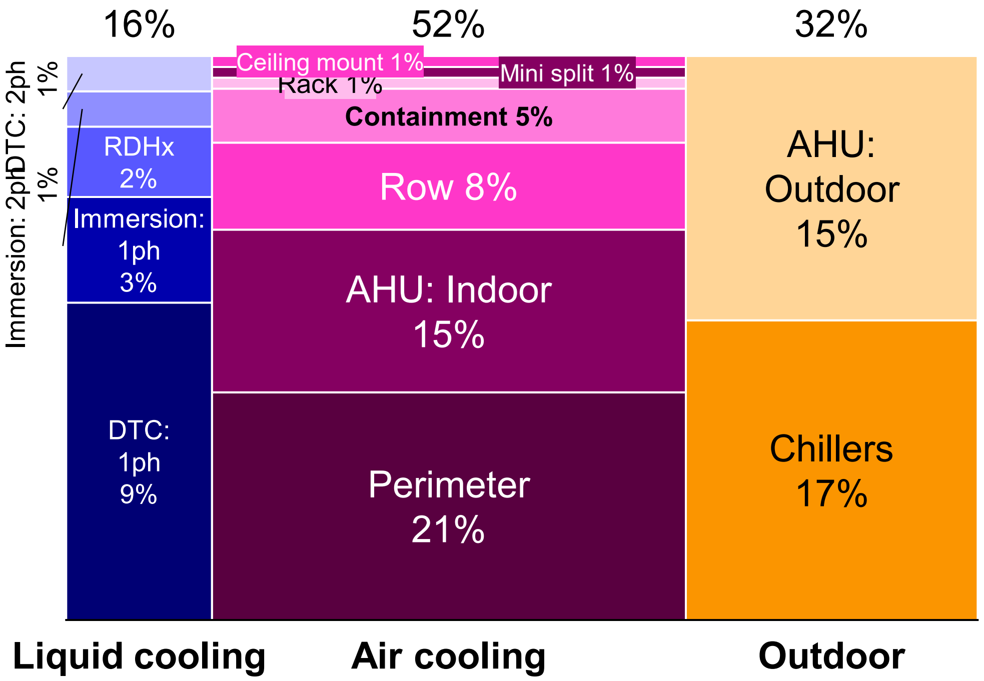


Source: Omdia

Liquid cooling already corresponds to **one seventh of the entire data center cooling market**, and it is forecast to become one third of the total by 2027.

2023 data center cooling market

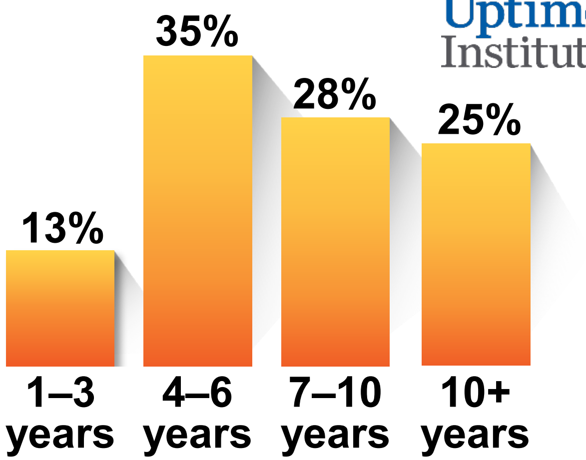
OMDIA



Liquid is expected to be more **important than air** as primary cooling technology for larger data centers. It gains importance without replacing need of air cooling.

How long do you think air-cooling will be the dominant approach for data centers >1MW?

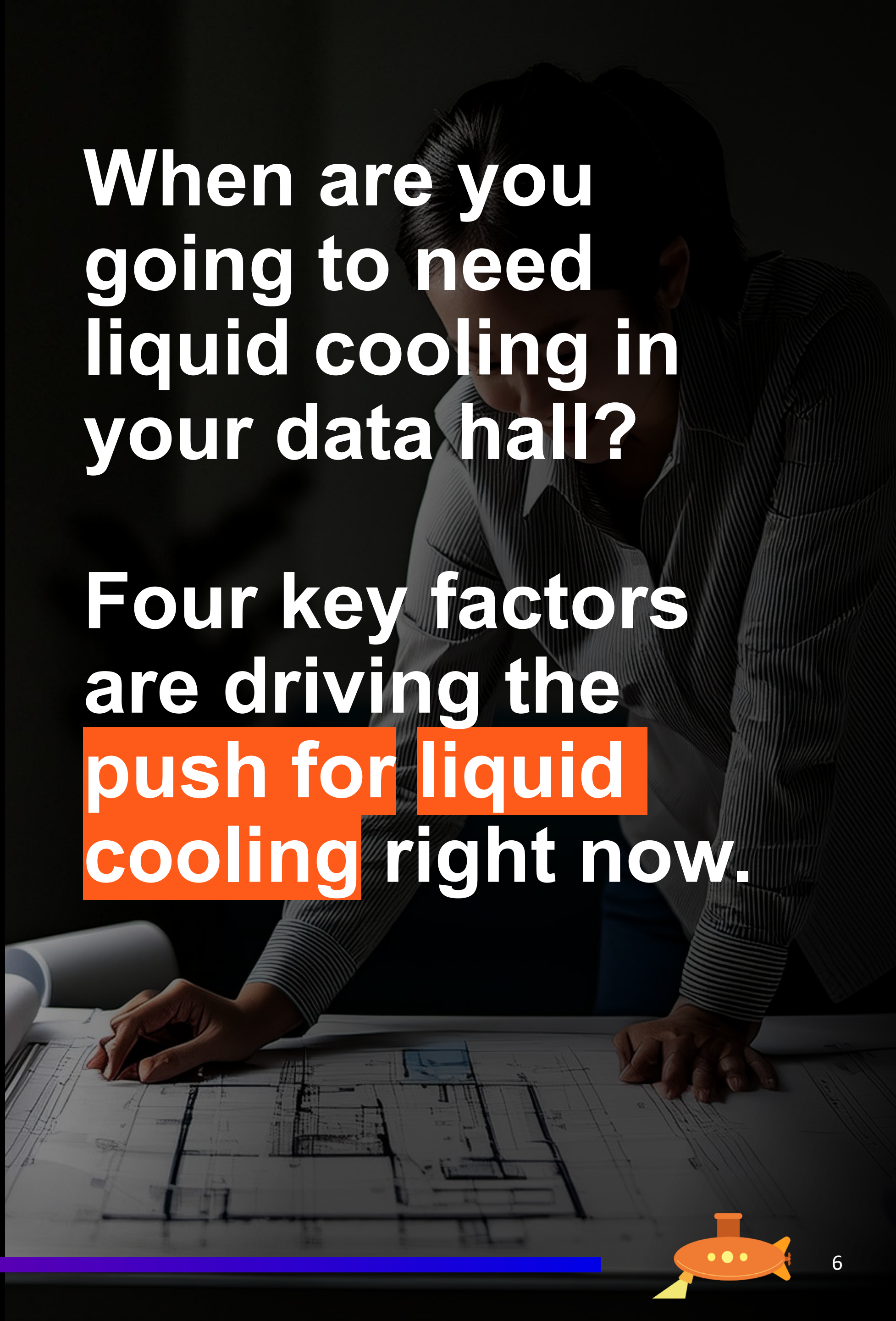
Uptime Institute



Liquid cooling is not a technology of the future.

It is here now, and its adoption is only expected to continue to grow in coming years.

Source: Omdia, Uptime Institute



When are you going to need liquid cooling in your data hall?

Four key factors are driving the push for liquid cooling right now.

TDP

Do chips have a **TDP above 700-800W**?

Air cooling starts being inefficient or unable to collect heat for chips above the 700-800W mark.

Airflow and material cost for heat sinks to extract heat via air become prohibitive.

Rack density

Are **rack densities above 40-60kW** per rack?

Rule of thumb of 40-60kW per rack will need to move to liquid.

Airflow requirements to meet needs with air only would be too costly.

IT compatibility

Was IT **designed or retrofitted to liquid cooling**?

Once cold plates are installed and heat sinks removed, server loses ability to be purely air-cooled.

Readiness

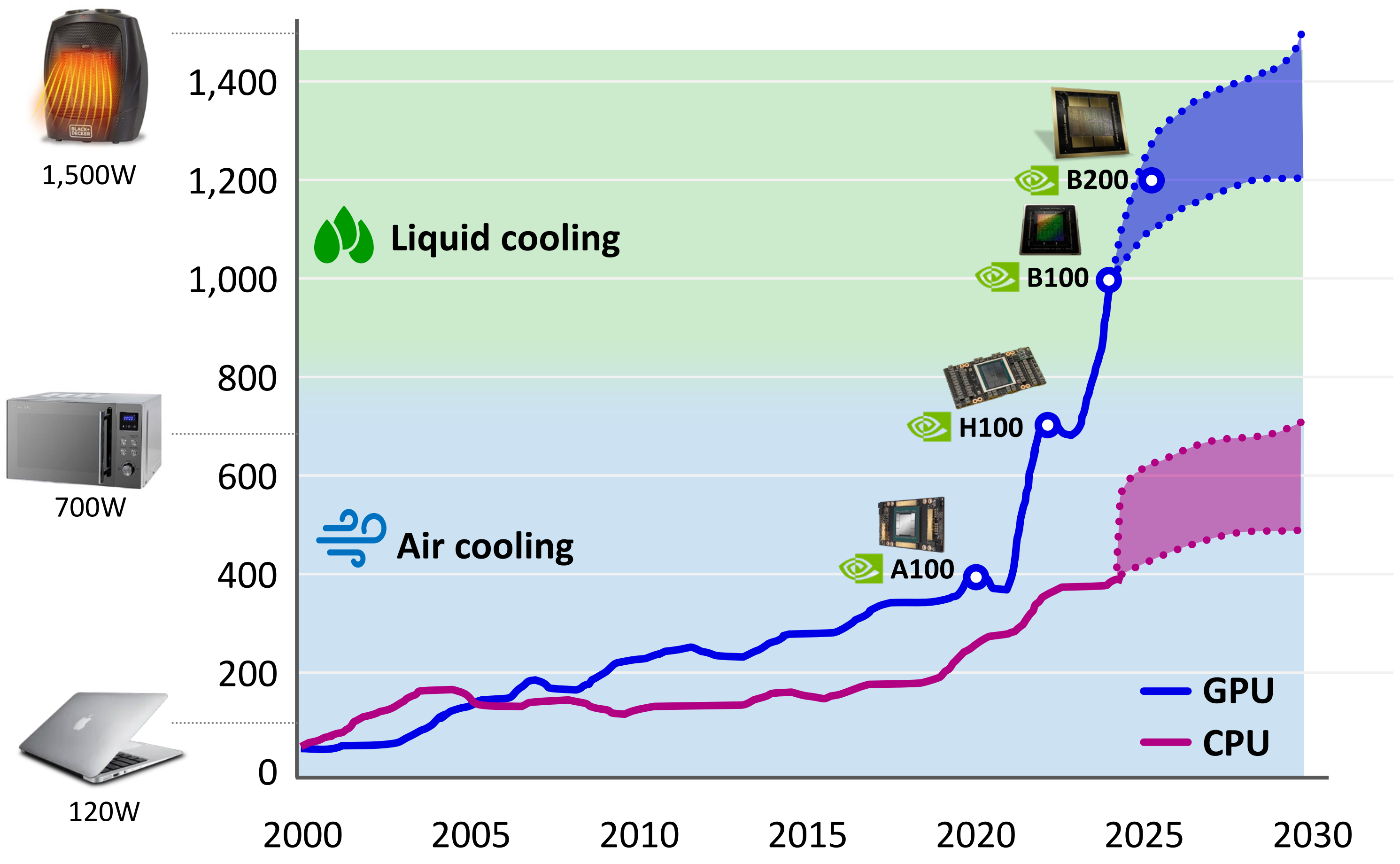
Is the data center expected to go to **liquid cooled soon**?

Data center operators can start experimenting with liquid cooling in small batches to get ready for when the technology is inevitable.

Higher thermal design power (TDP) of AI chips is a key factor driving adoption of liquid cooling technology.

CPU and GPU power consumption forecast

Thermal Density Power - TDP (watts)



Above 700-800W TDP per chip, liquid cooling quickly becomes a necessity.

Source: Omdia, NVIDIA, Data Center Dynamics

More compute packed in the rack is driving rack densities up, making the shift **from air cooling to hybrid air-assisted liquid cooling** a necessity.

Low density
Up to 10kW

Medium density
10 to 25kW

High density
25 to 80kW

Extreme density
> 80kW

Technology to extract heat from IT

Perimeter cooling

Row-based containment

Passive rear door

Active rear door

Direct to chip
1-phase coolant
2-phase

Immersion
1-phase
2-phase

***Hybrid cooling* will be the norm for most high-density designs, with air still responsible for cooling 10-30% of heat**

Around 40-60kW it is time to consider going liquid, but depending on server or chip liquid might start earlier

Technology to extract heat from data room

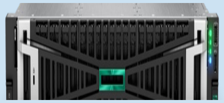
Direct expansion

Chilled water

Even amongst high-performing servers marketed to AI, models will continue being a **mix of air and liquid cooling** for the foreseeable future.

Even among top-of-the-range servers marketed to HPC / AI workloads, not all IT is off-the-shelf compatible with direct-to-chip liquid cooling.

A handful of examples from top brands, as of Dec-2023:

	Off-the-shelf air cooling compatible	*Off-the-shelf liquid cooling compatible*
NVIDIA	DGX H100 	A100 for PCIe single slot 
DELL	PowerEdge XE9680 	PowerEdge XE9640 
Lenovo	ThinkSystem SD630 V2 	ThinkSystem SD650-N V2 
Hewlett Packard Enterprise	ProLiant DL380 Gen11 	ProLiant DL325 Gen11 
SUPERMICRO	GPU SuperServer X13 NVIDIA MGX™ 	GPU SuperServer X13 8U Universal 

Source: NVIDIA, Dell, Lenovo, HPE, Supermicro

Cold plate

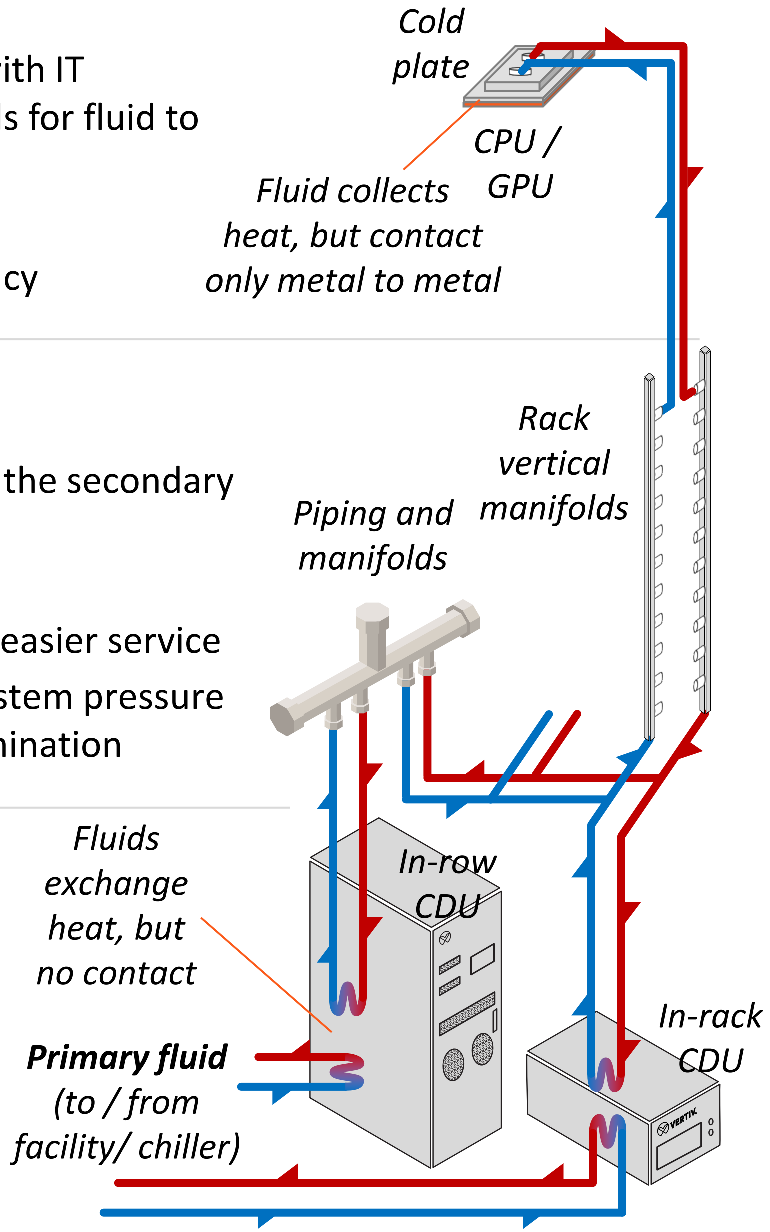
- Highly conductive metal in contact with IT equipment pierced by micro-channels for fluid to go through and collect heat
- Considerable variety in designs
- Critical with little room for redundancy

Secondary fluid network

- Piping, hoses and manifolds carrying the secondary fluid from CDU to cold plate
- Closed loop with minimal fluid load
- Equipped with quick disconnects for easier service
- Requirement to be able to handle system pressure and crucial to avoid leaks and contamination

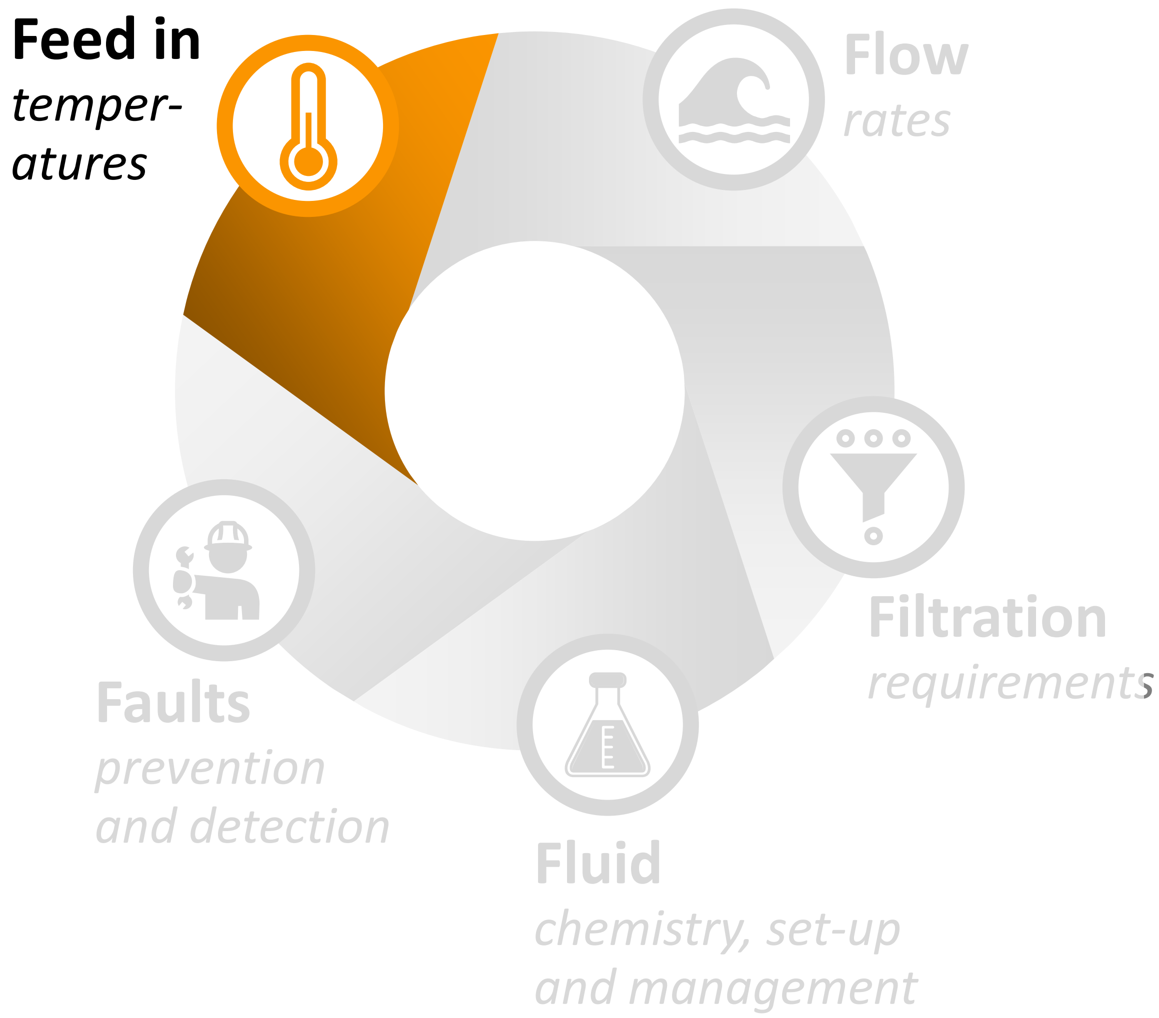
Coolant distribution unit

- Heat coming from cold plates transferred to primary fluid loop
- Crucial role in controlling flow rate, system pressure and filtration
- Redundancy ensured with multiple pumps and connection to UPS



Direct-to-chip liquid cooling introduces three new critical pieces of equipment to the server room.

Designing and operating a liquid cooling system requires mastering the 5Fs of liquid cooling.



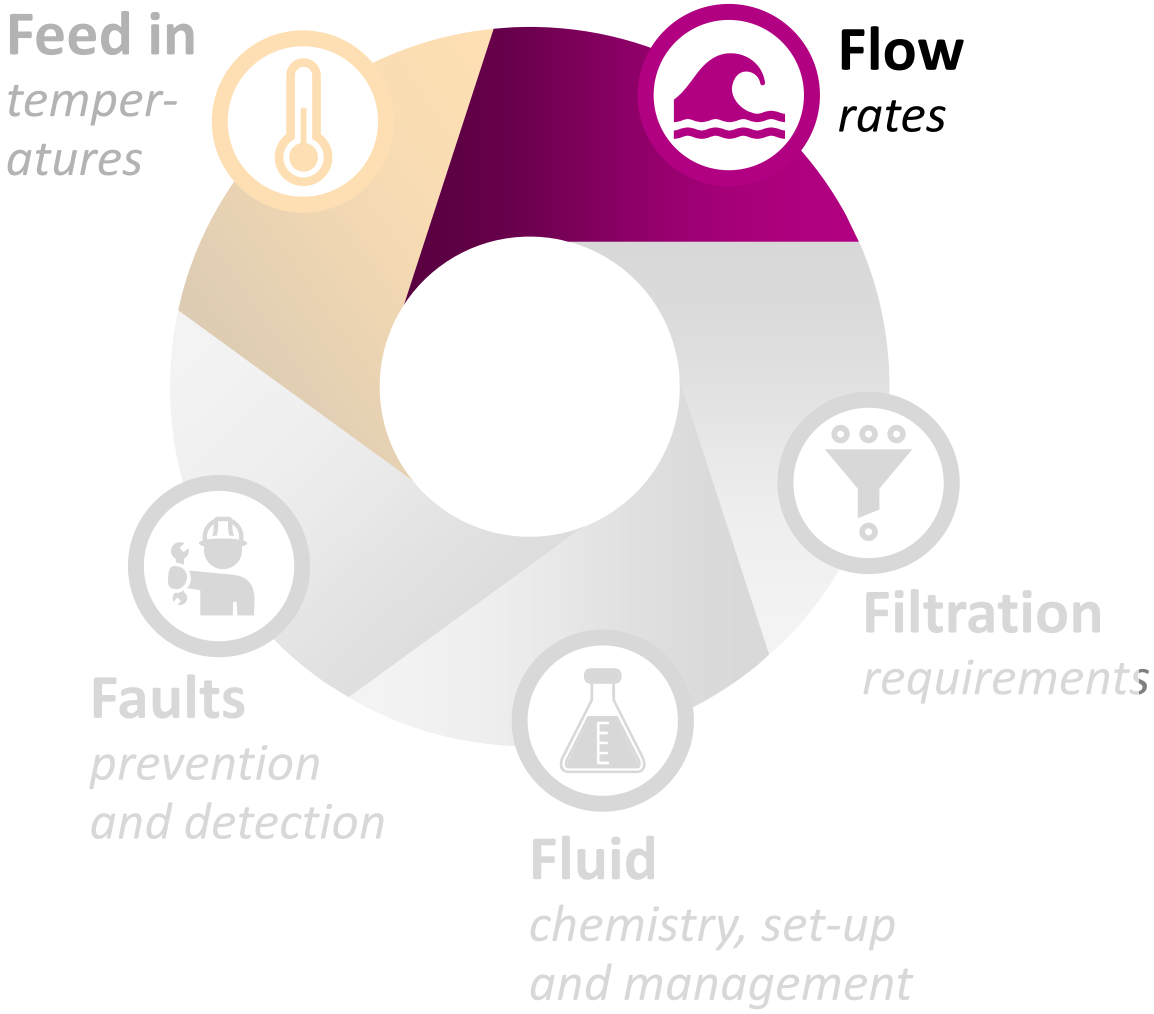
- Feed in **temperatures can differ considerably** by chip or server manufacturer:

  - DELL: *32°C feed in temperature (or maybe higher)* [1]
  - NVIDIA: *25-45°C feed in temperature* [2]

- Very low temperatures must be avoided otherwise risk of **water vapor condensation** and damaging IT equipment increases.

- CDUs **control heat exchange** between primary and secondary fluid networks and ensure **consistent temperature** feeding cold plates, even if ΔTs vary in time to accommodate different load levels.

Designing and operating a liquid cooling system requires mastering the 5Fs of liquid cooling.



- CDUs are **instrumental to control flows** of primary and secondary fluids.
  - **Secondary fluid** must be keep at **constant flow rate** at steady IT inlet temperature – designed to extract heat from cold plates at maximum load.
  - **Primary fluid** at **variable flow** according to the heat amount that needs to be exchanged, adjusted with approaching temperatures at CDU.
- Control of flow applied with **secondary fluid ΔP** and monitored to ensure pressure drops not being caused by leaks in the system.
- Mission critical flow with **pump redundancy** within the CDU and power supply redundancy.

Primary fluid flow rates variations are controlled by CDU in order to keep secondary flow rate and temperature feeding IT steady.

Primary fluid flow rate and ΔT controls are instrumental to maintain **secondary fluid flow rate and CDU T_{out} steady**.

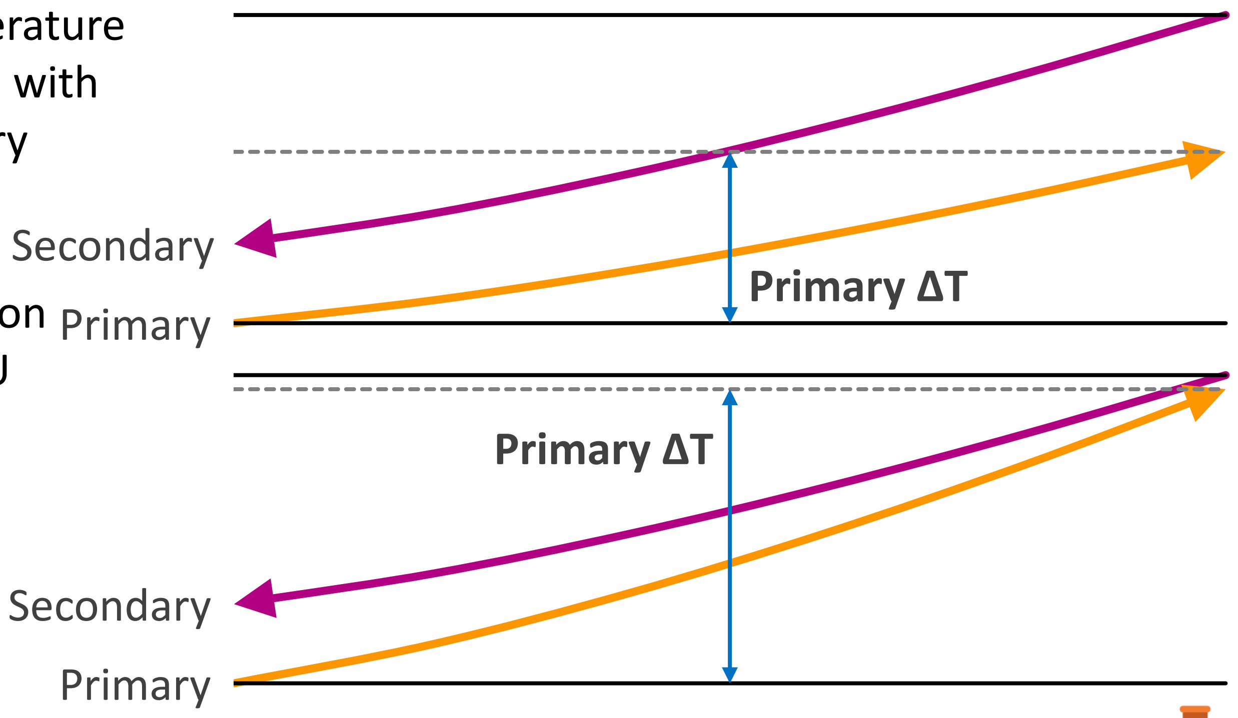
Heat balance equation

$$\underbrace{\dot{m}_1 \cdot cp_1 \cdot \overbrace{(T_{1\,in} - T_{1\,out})}^{\text{primary } \Delta T}}_{\text{heat collected by primary fluid}} = Q = \underbrace{\dot{m}_2 \cdot cp_2 \cdot \overbrace{(T_{2\,out} - T_{2\,in})}^{\text{secondary } \Delta T}}_{\text{heat rejected from secondary fluid}}$$

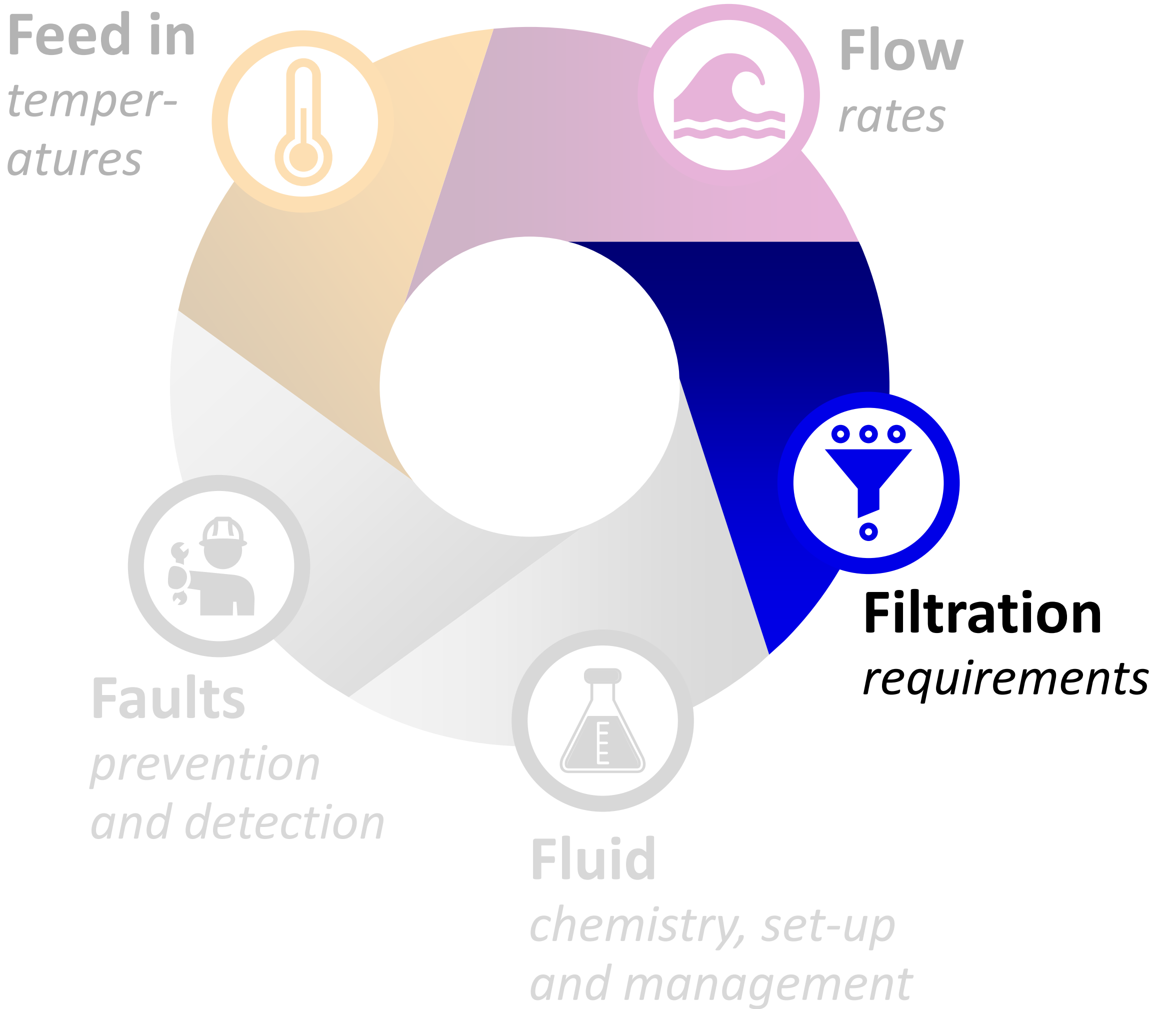
*where $\dot{m}$ is the flow rate and **cp** the specific heat of each fluid*

- **Constant** dependent on fluid chemistry
- **CDU must keep constant**
- **Controlled by CDU**

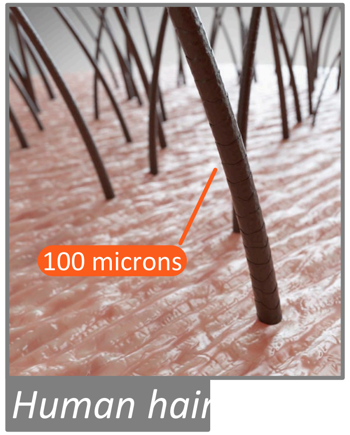
Temperature curves with primary flow rate variation at CDU



Designing and operating a liquid cooling system requires mastering the 5Fs of liquid cooling.

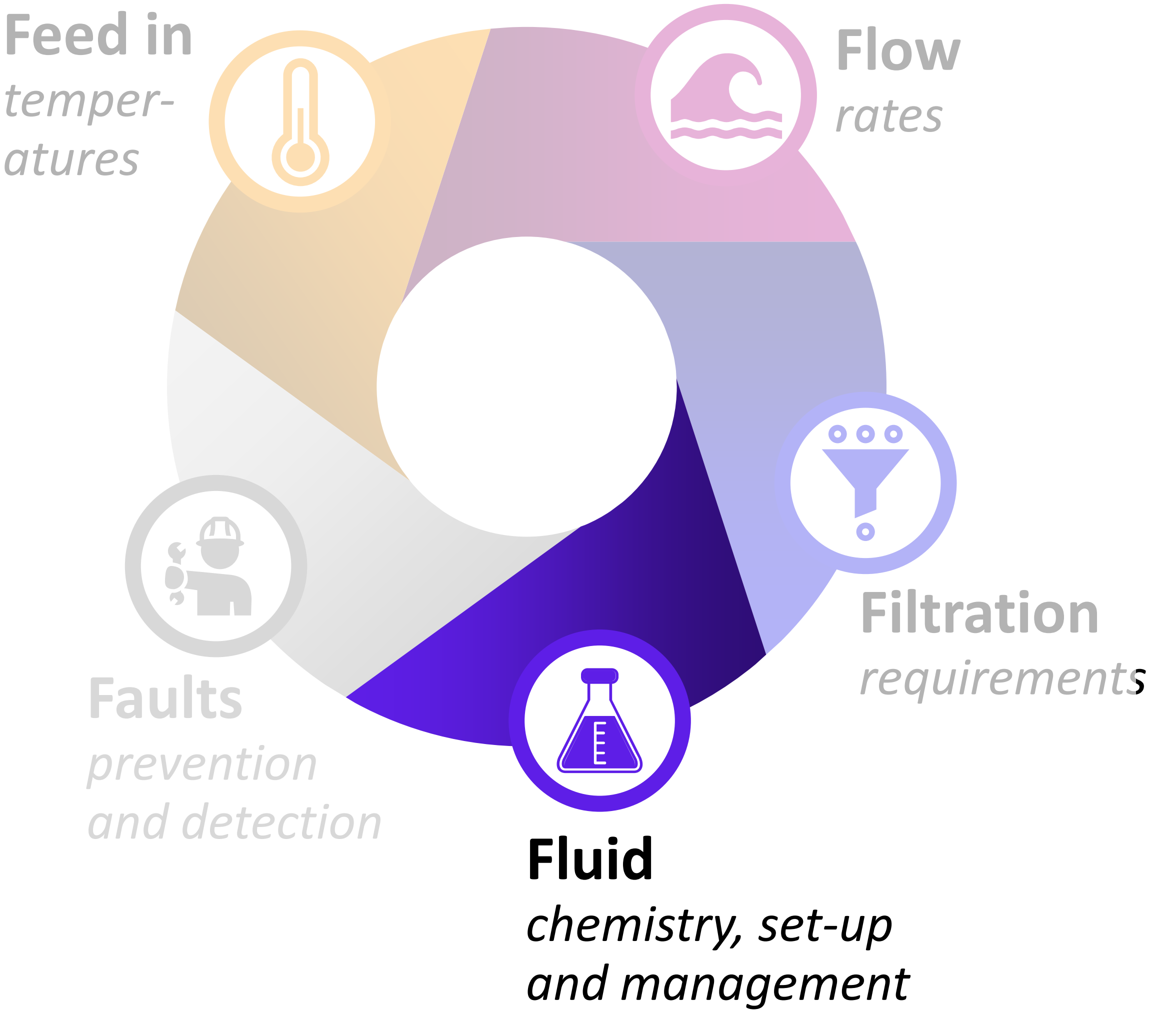


- Coolant flows through micro-channels in cold plates that can be as narrow as 27 microns.



- Fouled cold plates can obstruct flow throttling or shutting down IT gear, adding to maintenance costs.
- Filtration must be lower than cold plate channel size (rule-of-thumb design converging towards 25 microns).
- Always-on filtration at CDU is fundamental to keep system clean of impurities.

Designing and operating a liquid cooling system requires mastering the 5Fs of liquid cooling.



- Right fluid chemistry and supplier are fundamental decisions early on design – changing fluid strategy is costly requiring purging and decontamination.
- Fluid creates considerable complexity in commissioning entire solution and each server, including test fluid loop, flushing to remove impurities and cycling air bubbles out of system.
- Fluid requires considerable attention through its lifetime to ensure good condition: recurrent pH, visual appearance, inhibitor concentration and contaminant levels testing.
- All coolant fluid needs specialized storage and disposal, and handling requires adequate PPE.

Most common fluid options for 1-phase direct-to-chip cooling

	PG-25	PG-55	Treated water	Dielectric fluids
Specific heat	~ 3.9 J/g·K	~ 3.4 J/g·K	~ 4.2 J/g·K	*Vary*
Thermal conductivity	~ 0.49 W/m·K	~ 0.34 W/m·K	~ 0.61 W/m·K	*Vary*
Additives	Inhibitors, anti-foam	Inhibitors, anti-foam	Inhibitors, anti-foam, biocides	*Vary*
Pros / Cons	▲ Easier to maintain ▲ Packaged solution ▼ Higher ΔViscosity with ΔT	▲ Easier to maintain ▲ Packaged solution ▼ Higher ΔViscosity with ΔT ▼ Lower heat transfer properties	▲ Good heat transfer properties ▼ Harder to maintain with more frequent checks needed	▲ No short circuit risk posed by leaks ▼ Higher weight ▼ Higher cost ▼ Higher GWP ▼ Limited suppliers

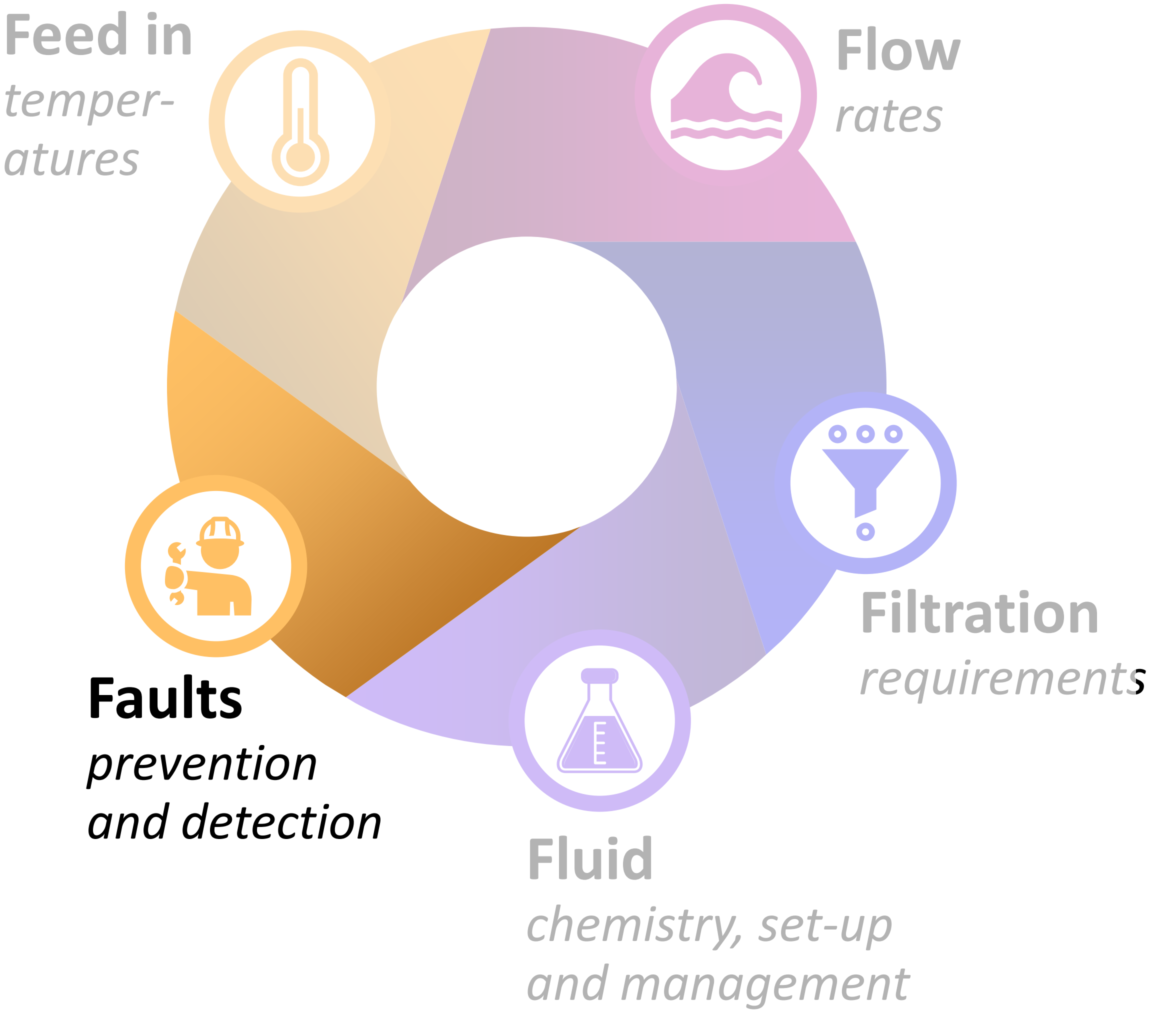
Emerging as industry consensus

Note: fluid properties can vary with temperature, approximate values given at 25°C or 50°C for comparison only.

Industry is starting to converge towards PG25 as fluid of choice for 1-phase direct-to-chip, but other options are available.

Source: OCP, Dow Chemical

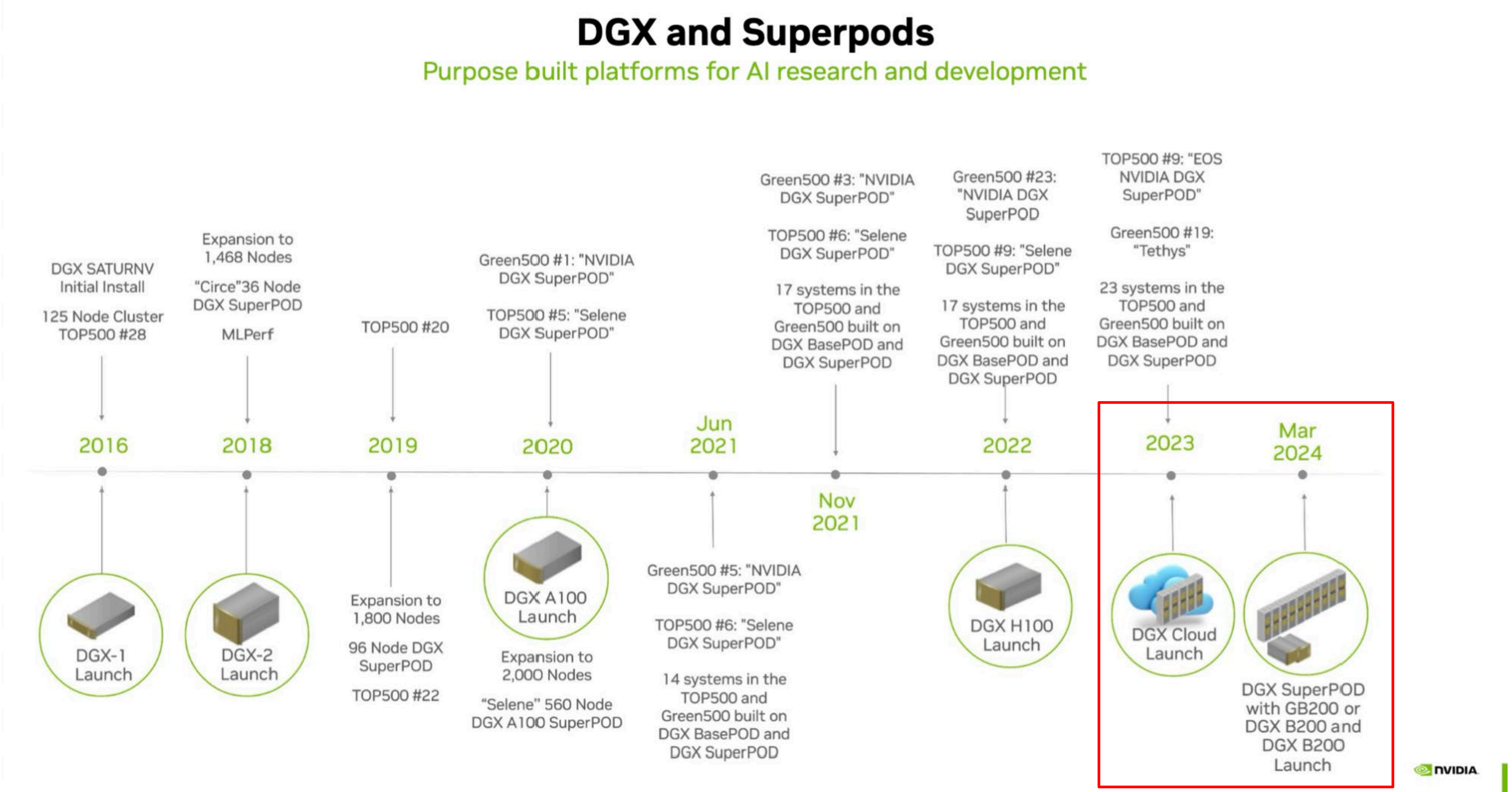
Designing and operating a liquid cooling system requires mastering the 5Fs of liquid cooling.



- **Monitoring and management of CDUs**, in addition to other sensors in the secondary fluid network, are crucial to ensure faults are identified early.
- Top of worry of data center managers is **leaking**:
  - Most leaks currently observed in liquid cooling are near quick-disconnect attachments between manifold and server hose, and they bring small risk to IT
  - Leaks within server chassis (between internal manifolds, hoses and cold plates) pose the greatest danger to IT equipment.
- Fool-proof system with **extra filtration and sensors** to curb risk of human faults adding contaminants or missing fluid quality checks while thermal exchange properties degrade.

Source: [1] Dell XE9640 Technical Guide; [2] NVIDIA GB200 recommendation

DGX Standard Architecture Roadmap



B200 Products Series

Building at scale with DGX B200 and GB200NVL

Reference systems

DGX B200

Enterprise form factor system that can be used as "drop-in" for DGX scale builds w/ H100 today :

- 10U chassis
- 2x Intel EMR
- 2 to 4TB Memory
- 2x 1.92TB M.2 System Drives
- 8x 3.84TB U.2 Data drives
- 2x BF3 for Storage N/S Dual Port NIC QSFP112 (FHHL)
- 2x module w/ 4x CX-7 each for Compute E/W NIC

DGX GB200 NVL

OCP form factor *Liquid Cooled* DGX system:

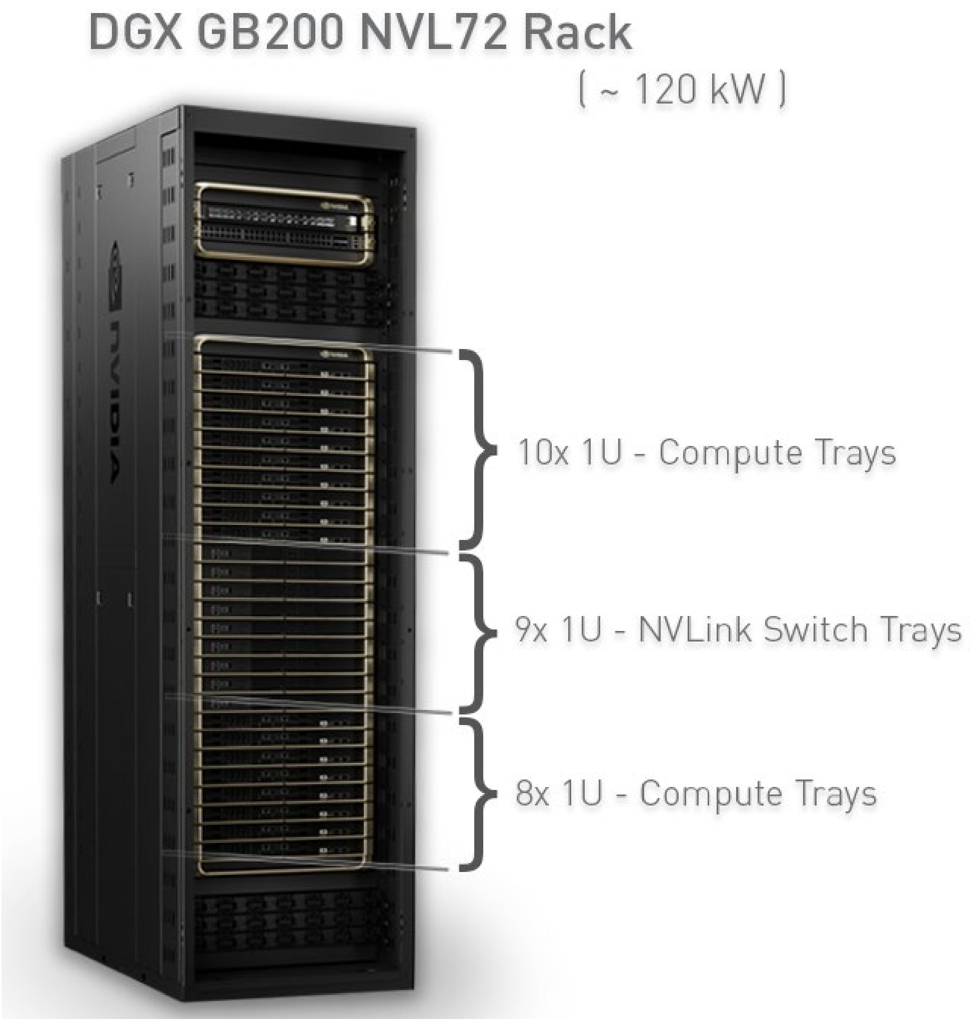
- Single OCP rack, 72-GPU NVL domain
- 18x compute trays:
  - 2x Grace CPU + 4x B200 GPU (as GB modules)
- 9x NVLink switch trays with shared backplane
- 4x InfiniBand per compute tray
- 2x BF-3 NIC per compute tray for Storage N/S

NVIDIA

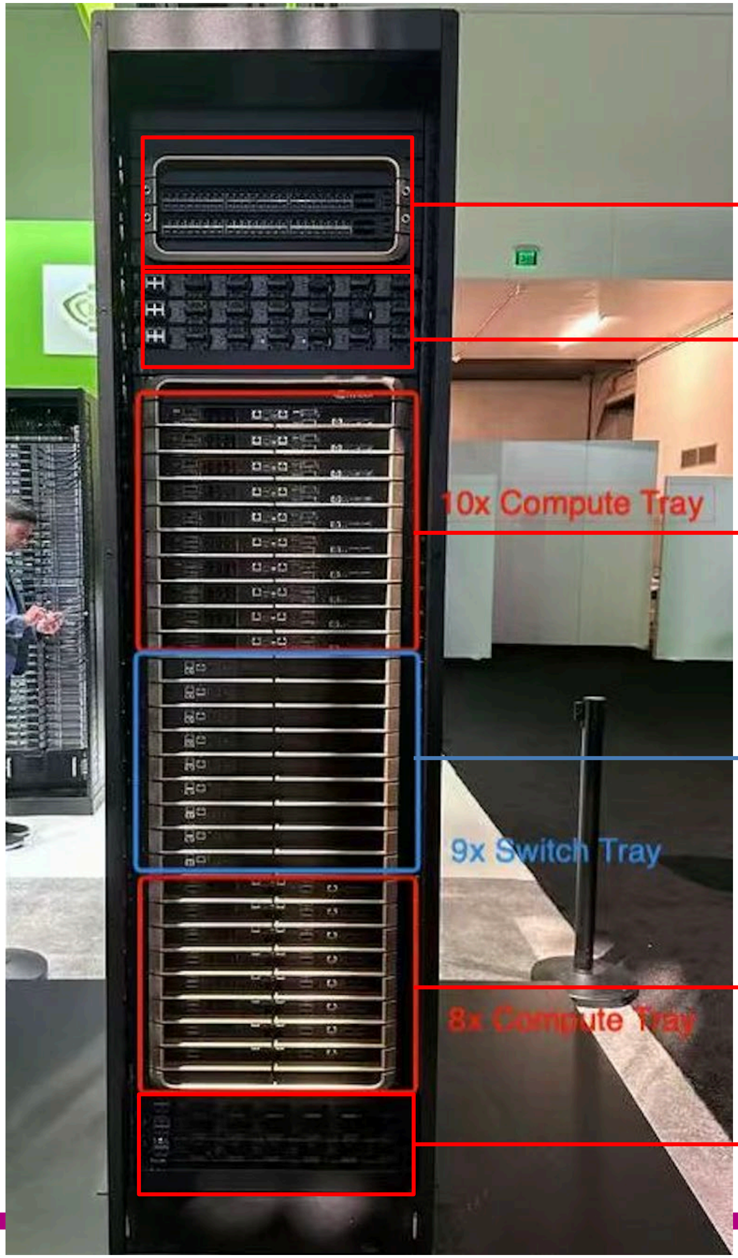
NVL72 Rack Architecture

➢ **High level summary-72 GPU Single Rack Configuration**

- Single Rack with 72 GPUs
- Oberon GB200 building blocks
  - Compute Trays with GB200 compute boards
  - Non-Scalable NVLink Switch Trays
  - NVLink passive copper cable backplane
  - Power Shelves, Bus Bar
  - Liquid Cooling Manifolds
  - Rack infrastructure
- Reference configuration
  - **ORV3 Rack**
  - 72GPU Rack
    - 18×1RU Compute trays
    - 9×1RU Non-scalable NVLink Switch Trays
  - **Hybrid Cooling trays**
    - Grace CPU, Blackwell GPU, CX7, and NVLink Switches ASICs are liquid cooled
    - Rest of the components are air cooled
- ORV3 compatible. Houses EIA (19” and RU pitch) compatible trays, cable cartridges and manifolds.
- Customized to interface to custom racks.



NVL72 Rack Architecture



120kW / Rack; 72GPU; 1440P@FP4

Switch:
2*NV Switch Chips; 72*2 Ports

Power Shelf:
PSUs Update to 5.5kW*6*2=66kW, 54V Output

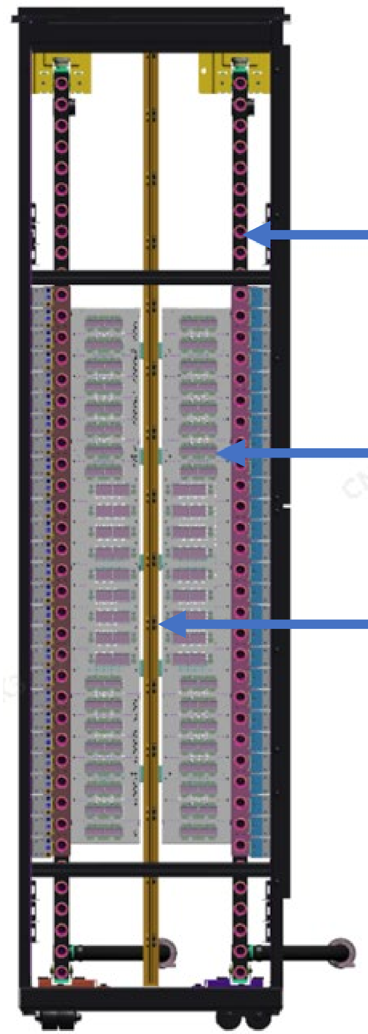
Compute Tray(1U):
10*2*(Grace CPU+2*B200)

NV Switch:
9*(2*NV Switch Chips); 72Ports

Compute Tray(1U):
8*2*(Grace CPU+2*B200)

Power Shelf:
PSUs Update to 5.5kW*6*2=66kW, 54V Output

NVL72 Rack Architecture



Liquid Cooling Manifold

NVLink Interconnect Cable Cartridges

OCP Power Self 48V Busbar

NVL72 SuperPod Architecture

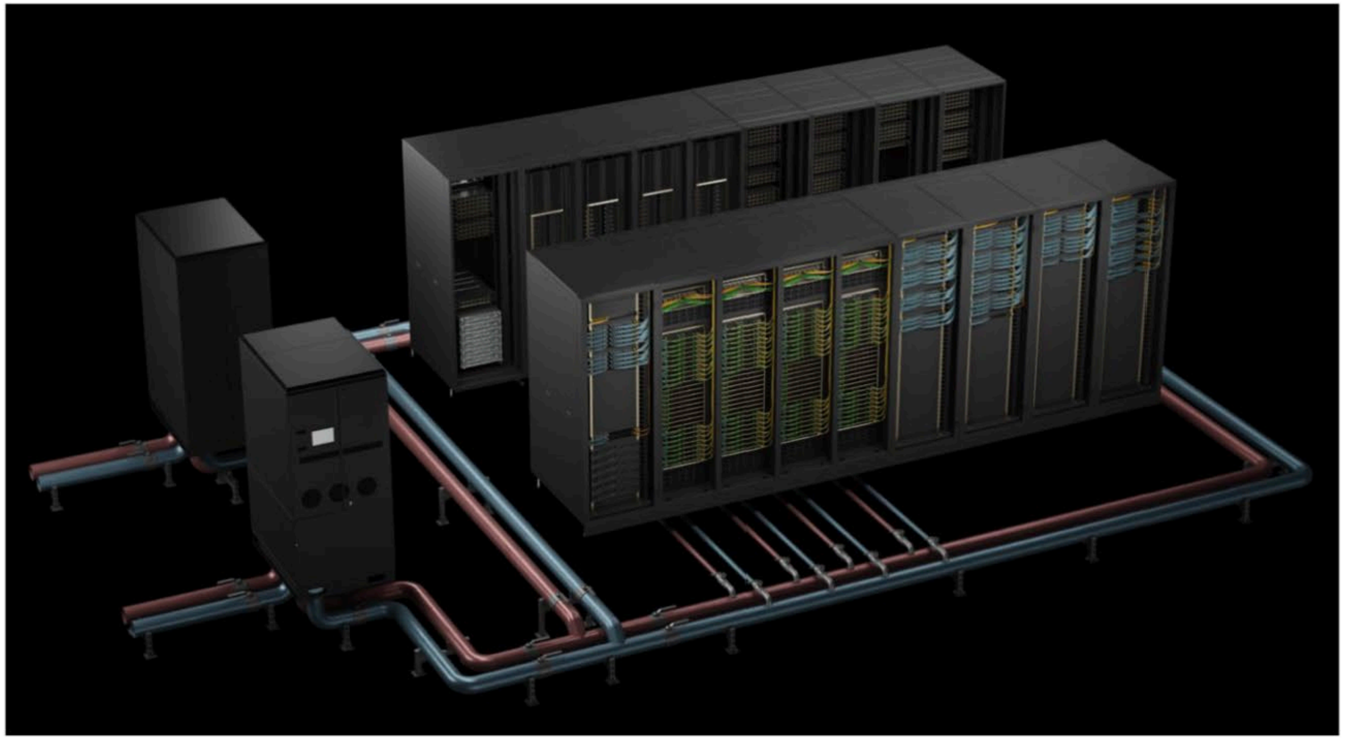
Liquid Cooling

GB200 NVL72

Using the liquid cooling options presented before

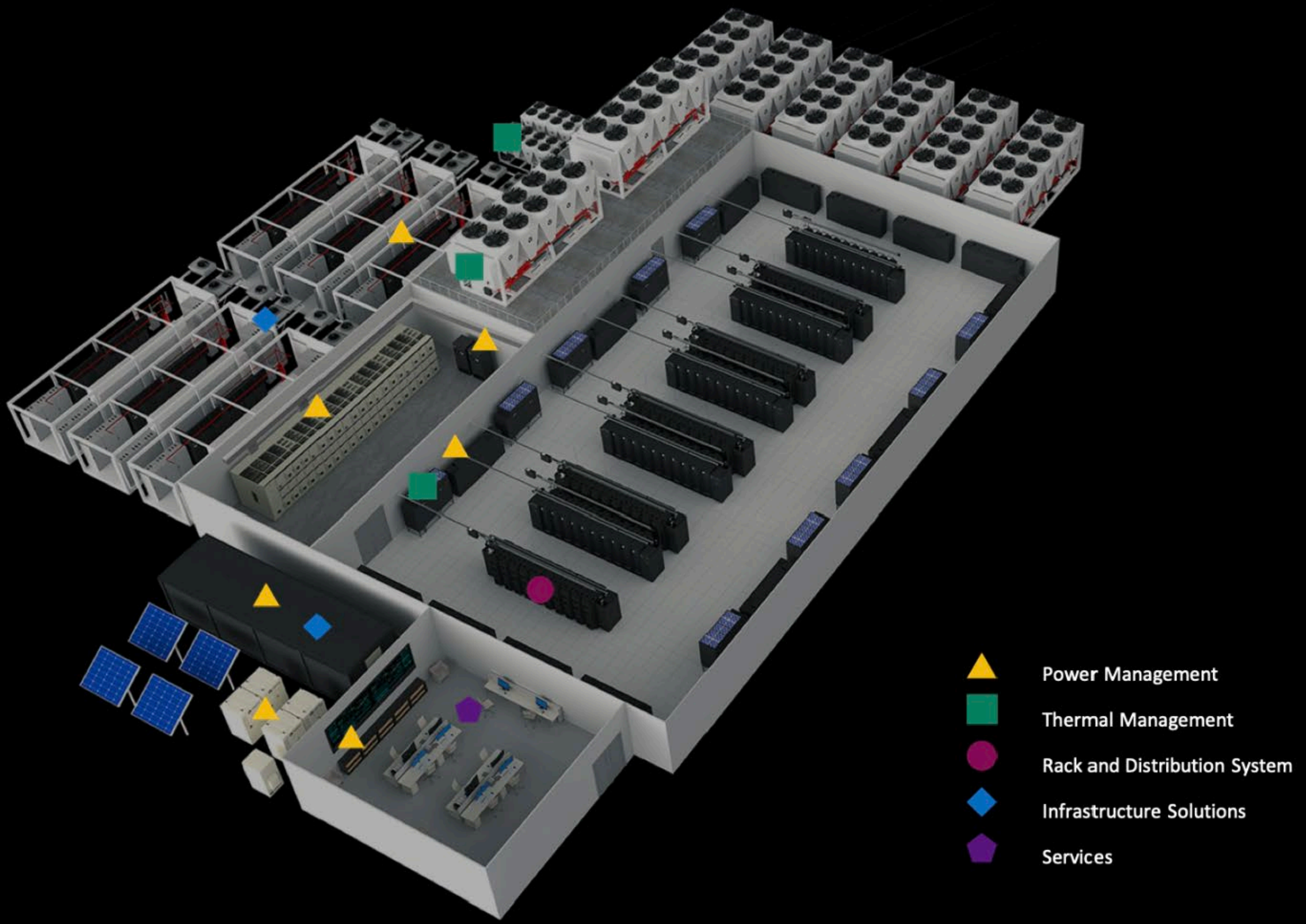
- In row or in rack CDUs requiring primary/secondary loops
- In-row heat exchangers

With telemetry and systems management



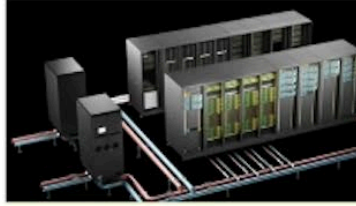
Vertiv Announces GB200 NVL72 Energy Efficient Reference Architecture

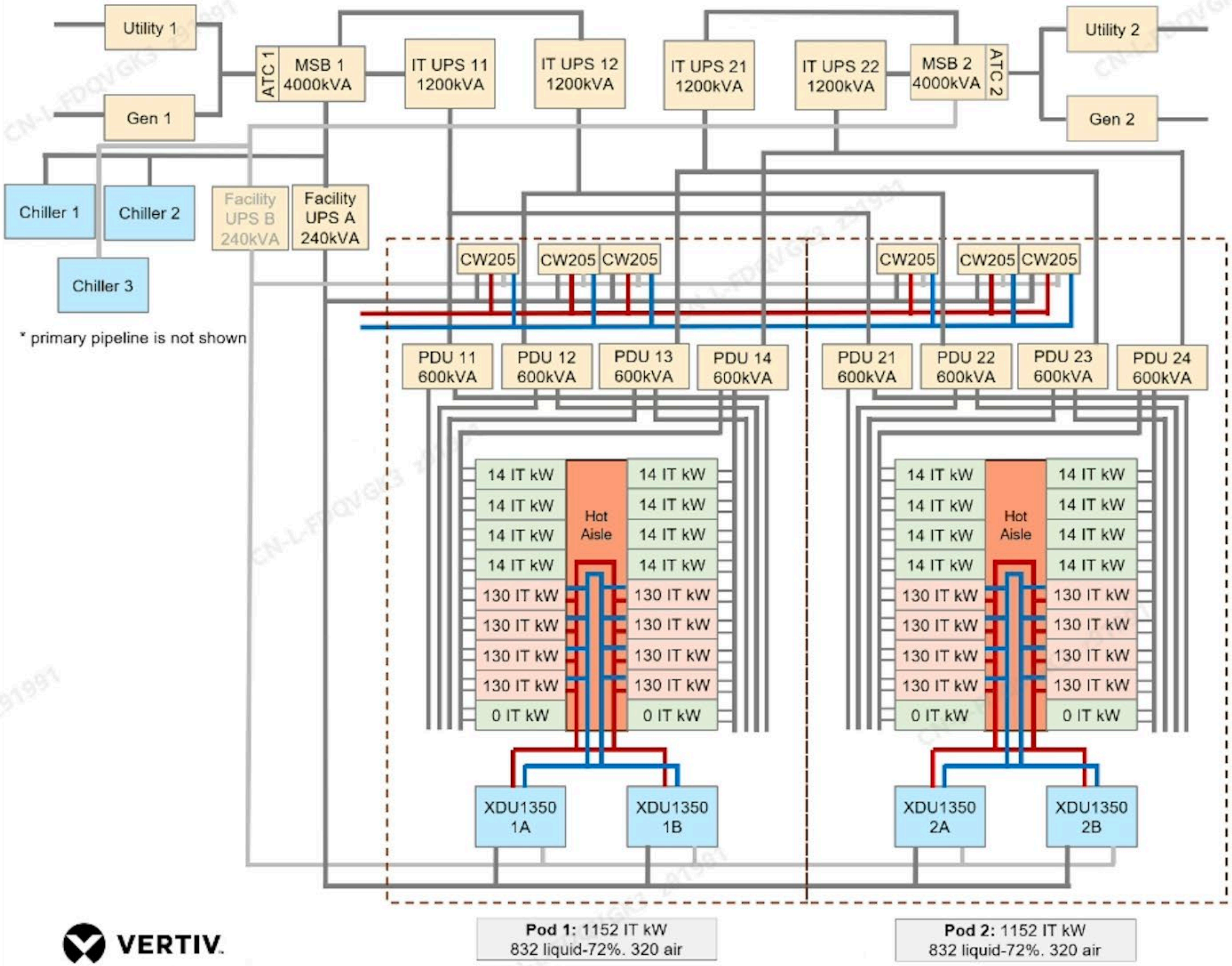
An end-to-end blueprint for today's most advanced AI factories



NVIDIA

Vertiv Reference Design: Nvidia Blackwell NVL72 @ 130 kW/Rack, liquid-to-chip (1152GPUs, 2304kW 2*PODs, 80% liquid in compute rack, distributed redundancy)





4 to make 3 configuration for 2 Pod

- 2* 4000kVA MSB for IT and facility
- 4* 1200kVA IT UPS system for two pod, 4 to make 3
- 2* 240kVA facility UPS system for two pod, 2N
- 8* 600kVA PDU

- 8* 400A 240/415V Busbar for each pod
- 2* 60A TOB per rack for each Busbar
- 8* 33kVA DC power shelf per computing rack
- 2* 30A vertical rPDU per support rack, 2N

- 2* XDU1350(N+1 pumps) for each pod, N+1
- 3* CW205s room cooling, 372kW for each pod, N+2 for two po

- 3* FH3135 with 30% propylene glycol, N+1
- 2* 240kVA facility UPS(n+1) system, 2N
  power consumption: (20.5*2+17*3)*2=184kW

VERTIV

Best Practice Project in Taiwan

Customer Requirement

- D2C Cooling
- Racks：50
- Power Density：Air35kW/R、Liquid50kW/R
- No Raised Floor

Solution

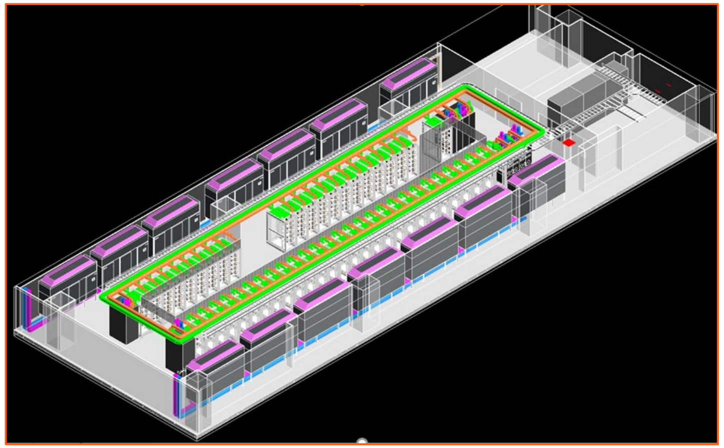
- 2sets XDU1350+14 set PW170 Chiller Fans Wall
- Chiller Cooling
- Upper Pipes design
- One rack one control

Core Feathers

- New HGX GH200×4
- Closed Heat Aisle
- Hybrid Cooling in one rack
- XDU Primary in/out 7/12℃
- PCW in/ out 18/28℃
- One rack one control of flow rate

Customer Value

Test result is that when cooling solution is changed from 100% air cooling to 75% liquid cooling，server fans’ power consumption decrease by 80%, TUE(Total Usage Effectiveness) improve by 15%



Range of Reference Designs for AI Data Centers

Design for new builds

Rack density	Rack count	GPU count	Design ID			Cooling technology
			NA	EMEA	ASIA	
20kW	18	248	RD002	RD002E	RD002A	Air
40kW	10	248	RD003	RD003E	RD003A	Air
40kW	10	248	RD004	RD004E	RD004A	Air
73kW	88	2304	RD006	RD006E	RD006A	Liquid + Air
73kW	110	2880	RD007	RD007E	RD007A	Liquid + Air
132kW	36	1152	RD014	RD014E	RD014A	Liquid + Air
132kW	54	1728	RD015	RD015E	RD015A	Liquid + Air
132kW	72	2304	RD016	RD016E	RD016A	Liquid + Air
75kW	8	64	RD017	RD017	RD017A	Liquid + Air
90kW	12	576	RD018	RD018	RD018A	Liquid + Air
132kW	18	576	RD019	RD019	RD019A	Liquid + Air
300kW	-	-	RD300	RD300E	RD300A	Liquid
500kW	-	-	RD500	RD500E	RD500A	Liquid

Rack & Row Standard Configurations Optimized for Retrofits

Technology Summary		Racks	Density per rack[1]	Green field / Brown field	Heat removal		Chiller Included
					From server	From room	
AI test environments, training pilots or edge inferencing							
Small HPC minimal retrofit	1L88R	1	88kW			air	-
Small HPC retrofit for chilled water system	1L100R		100kW			water/ glycol	-
AI labs, transition to AI data center							
Mid-size HPC cost-optimized retrofit	4L400R	4	100kW			refrigerant	✓
Mid-size HPC with increased heat capture	4XL400				+	water/ glycol	-
Mid-size HPC pragmatic retrofit for air cooled computer rooms	4X160R		40kW			refrigerant	✓
Mid-size HPC low complexity retrofit with air-cooling	5L500	5	100kW			water/ glycol	-
Prototype AI factory							
Large HPC preserving room neutrality	12XL1200	12	100kW		+	water/ glycol	-
Large HPC building towards scale	14L1400	14				water/ glycol	-

Row Solution with heat capture for AI labs and IT white space

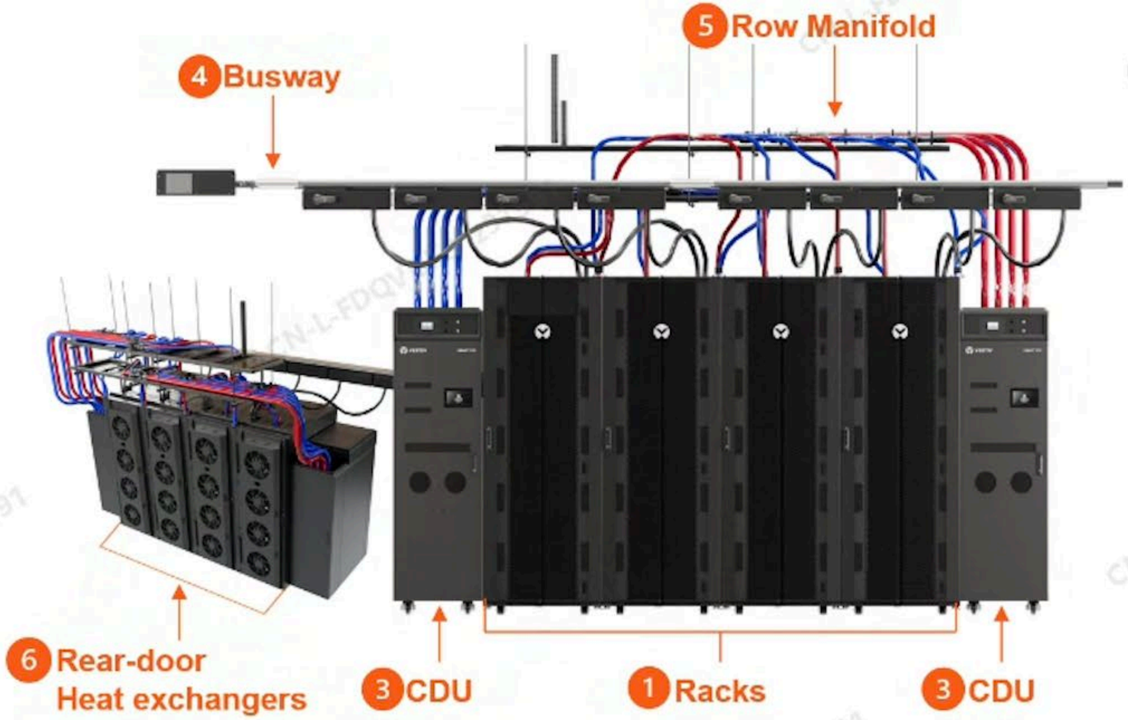
4 *Rack(s)*

400kW *Total solution capacity*

Liquid Cooling Type:
Liquid-to-liquid

Chilled water loop required

Model Number:
4XL400

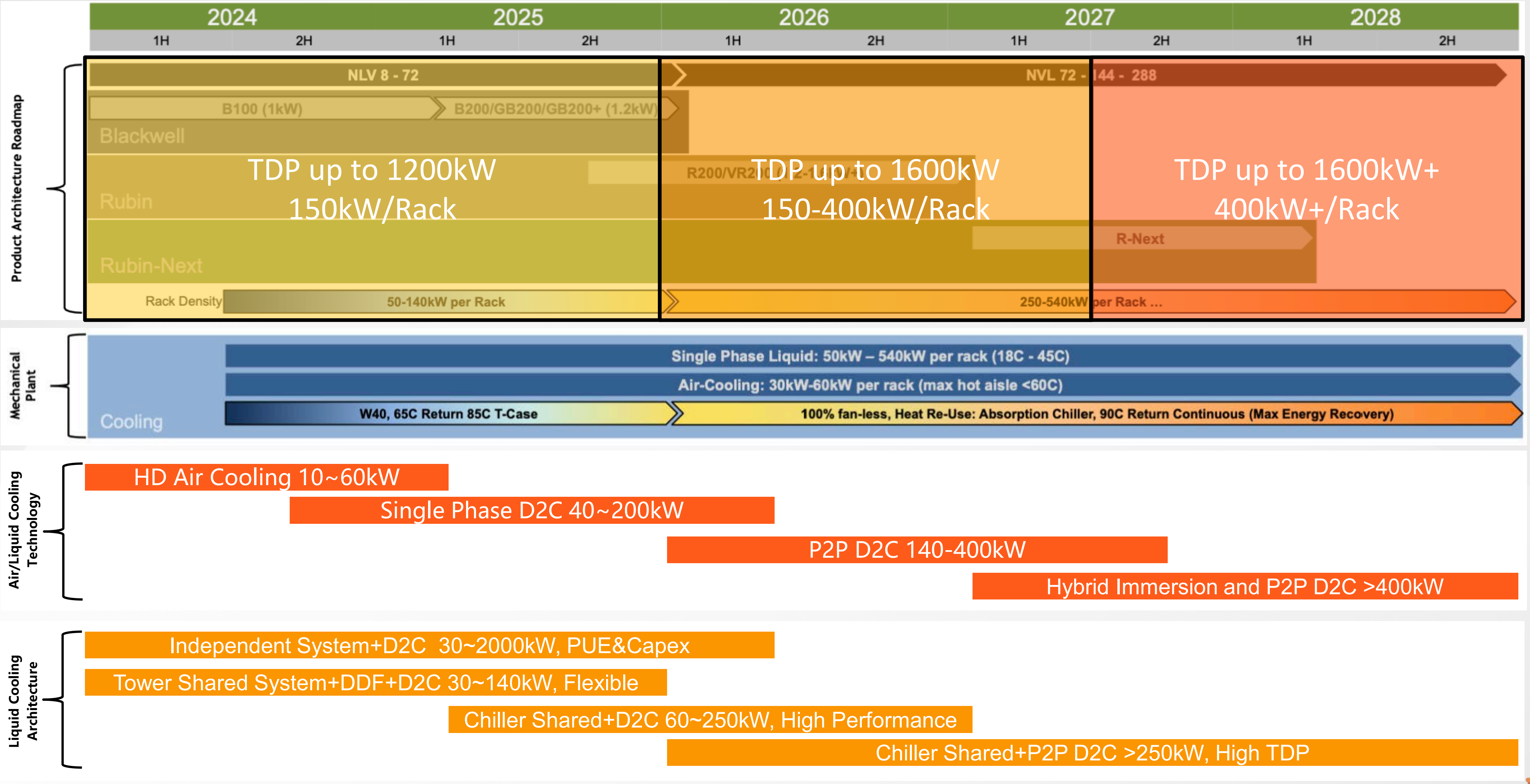


What's Included

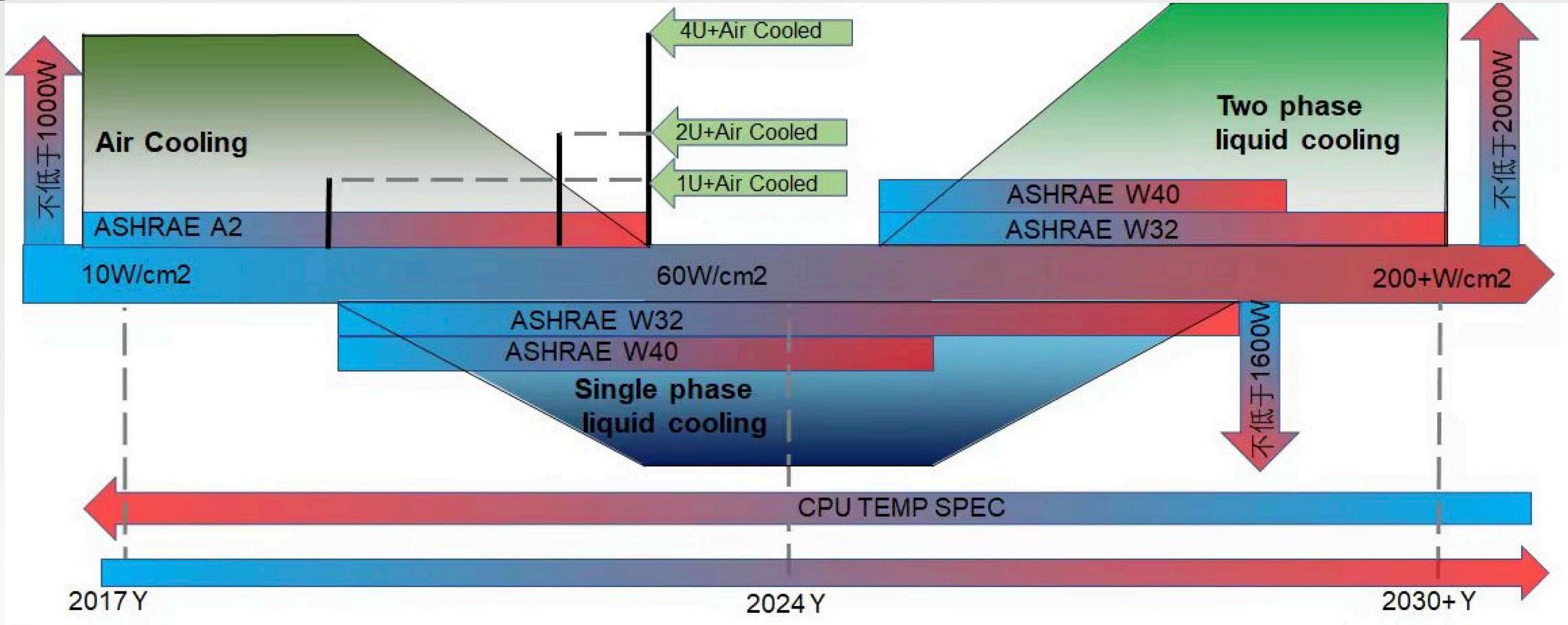
1. Rack Enclosures
2. Rack PDU (2 per rack)
3. CDU's (2)
4. Busway (with taps and endcap)
5. Rack & Row manifolds
6. Rear-door heat exchangers (4)
7. TH Sensors (2 per rack)
8. Remote Management
9. Deployment + Commissioning
10. Maintenance

← Back

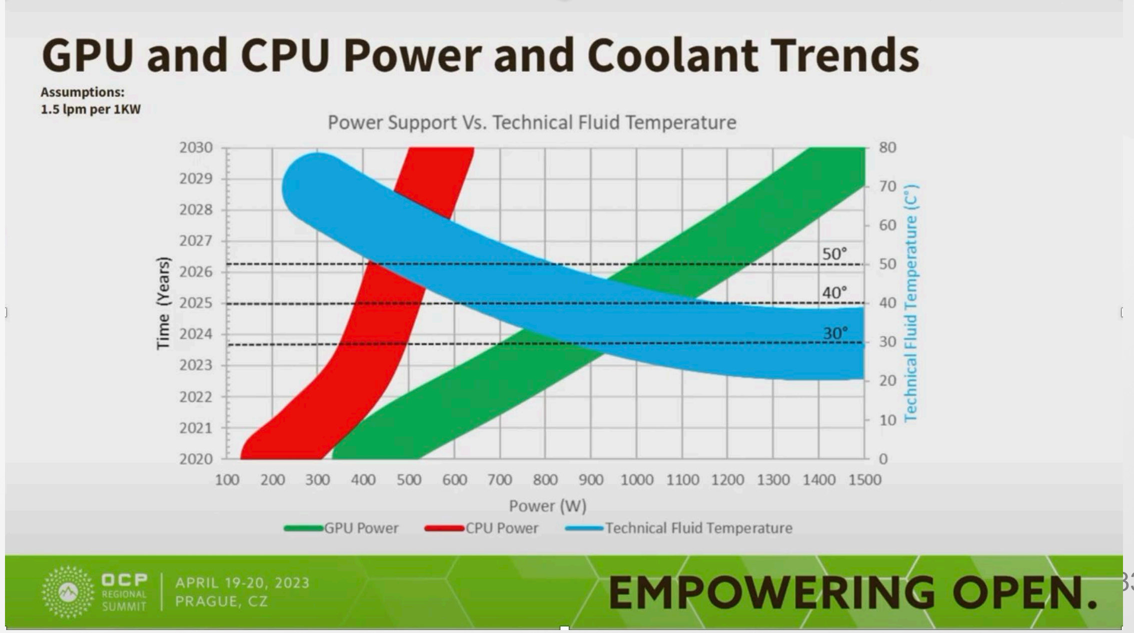
Cooling Tech and Architecture Trend Aligning



The Design Temperature Guide from ASHRAE TC909 and OCP

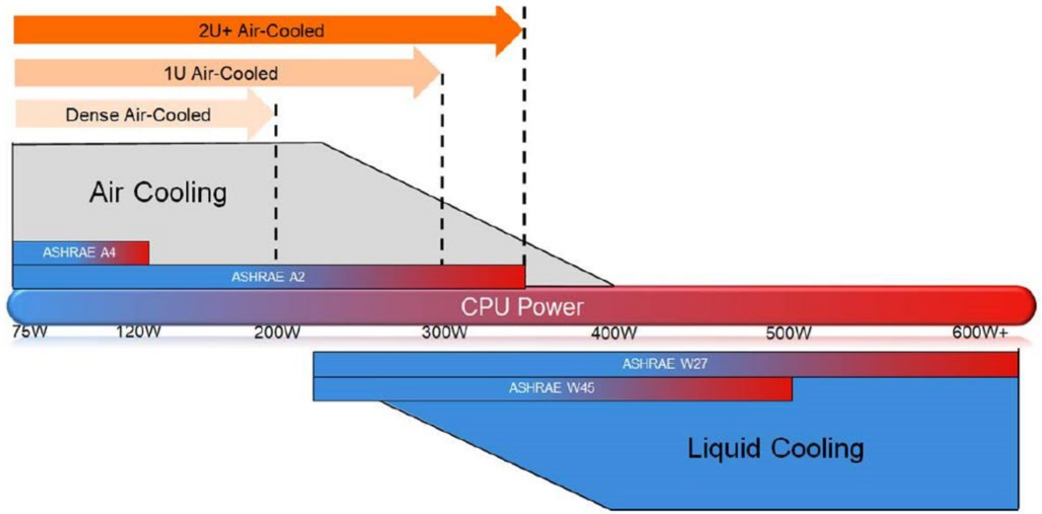


TCS Fluid Class	Typical Infrastructure Design		Maximum TCS Supply Temperature
	Common FWS Facilities	TCS Facilities	
S30	Chiller / Cooling Tower	CDU	30°C (86°F)
S35	Chiller / Cooling Tower		35°C (95°F)
S40	Cooling Tower		40°C (104°F)
S45	Cooling Tower / Dry Cooler		45°C (113°F)
S50	Dry Cooler		50°C (122°F)



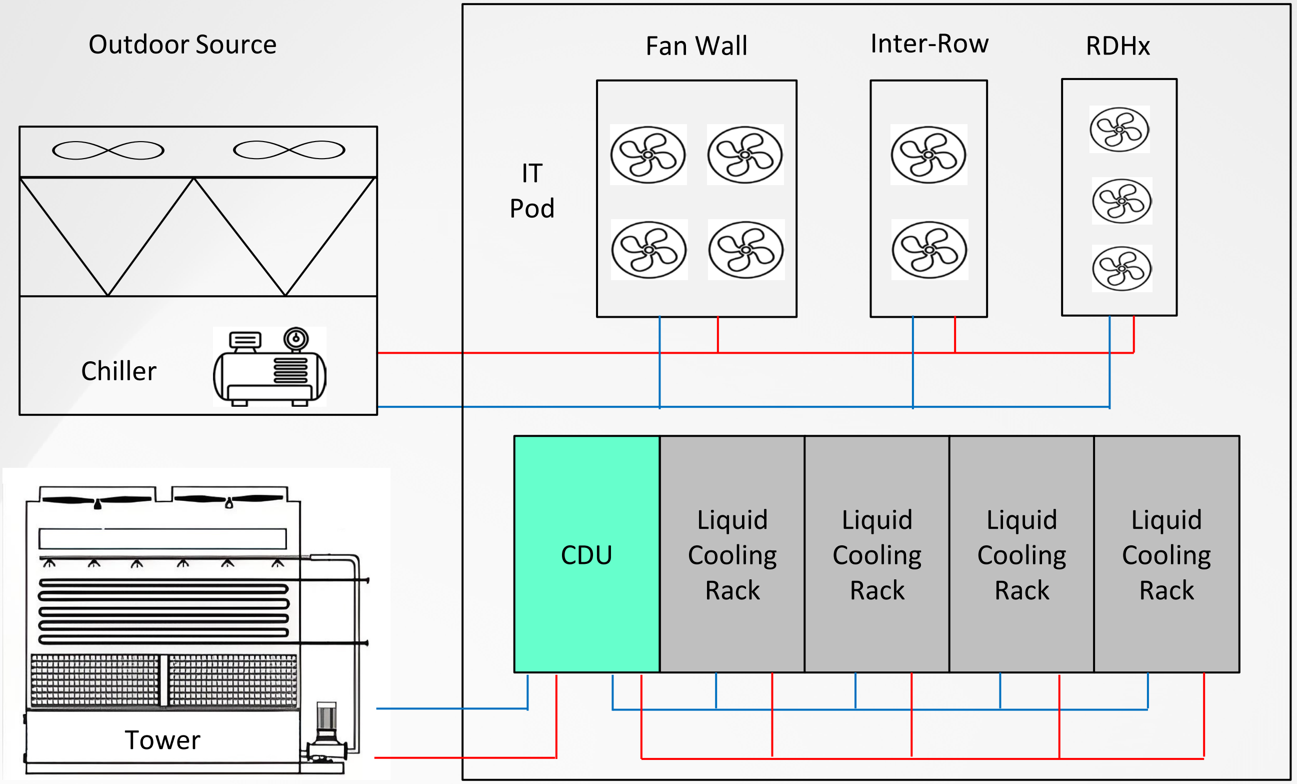
Benefits of Liquid Cooling over Air

- IT Equipment is Running too hot for Air
  - Increased Cooling Capacity
- Save Space (Higher Densities Achieved through Liquid Cooling)
- Reduce Energy Consumption
  - More efficient heat dissipation
  - More stable temperatures throughout the equipment
- Reduce Noise in the Data Center



Fluid Types	Specific Heat Capacity [J/(kg*K)]	Thermal Conductivity [W/(m*K)]
Water	4200	0.61
Air	1000	0.026
Oils	1670-2200	0.13
Single Phase Fluorocarbons	1100-2300	.059-.067

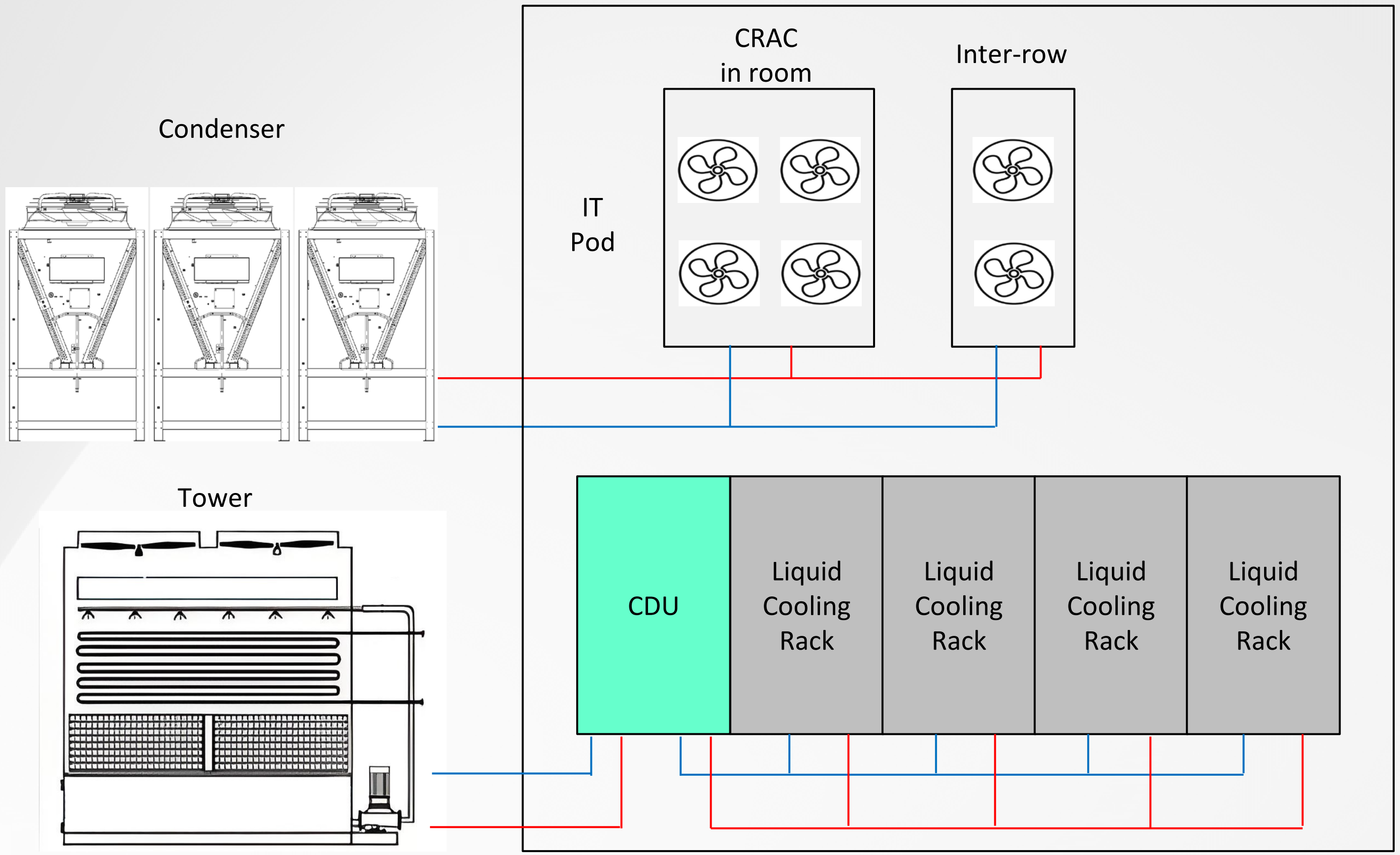
Independent System1: Chiller + D2C



Structure Features

1. Suitable for chiller data center retrofitting
2. System is independent and Simple
3. Liquid cooling sub-system is easy to O&M
4. Various air cooling terminals to choose

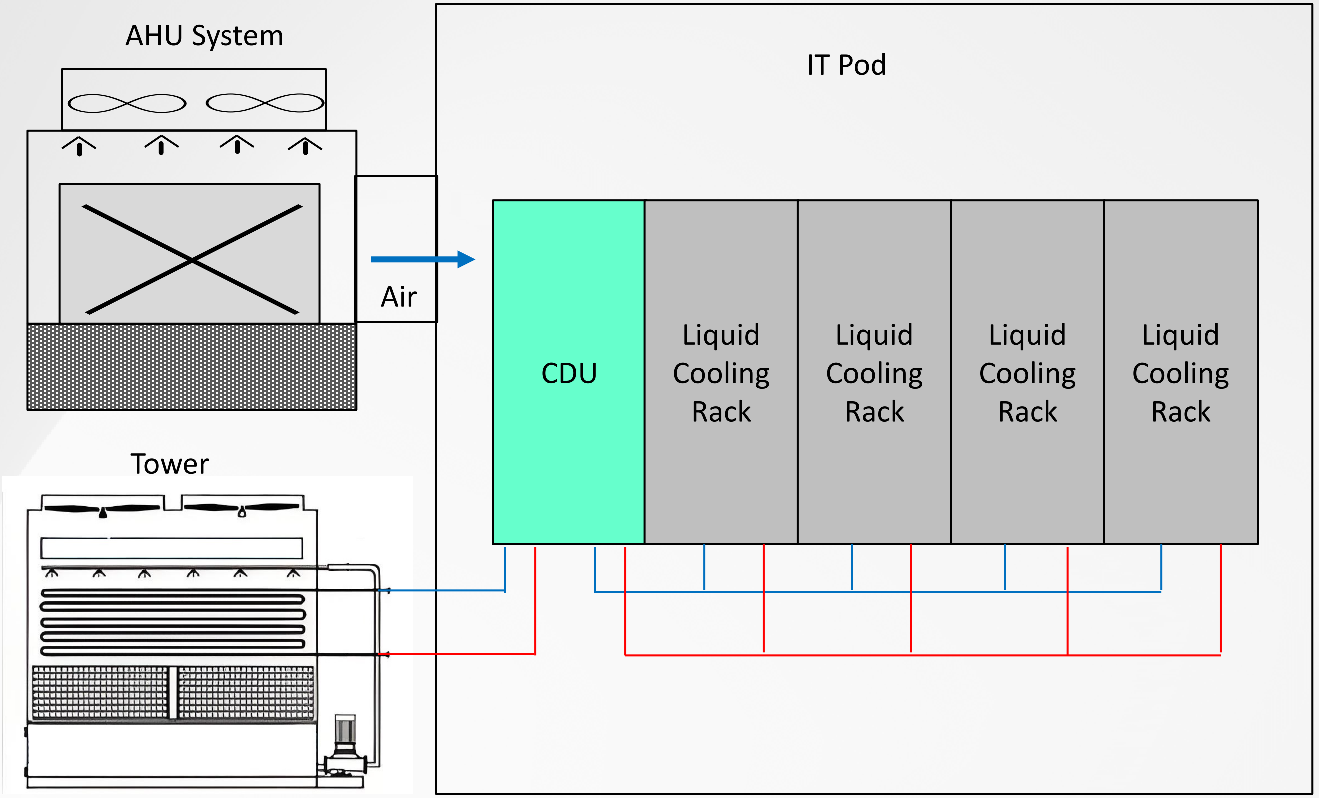
Independent System2: CRAC + D2C



Structure Features

1. Suitable for CRAC data center retrofitting
2. Water free and simple system
3. Liquid cooling sub-system is easy to O&M
4. In winter add fluorine pump modular to improve PUE

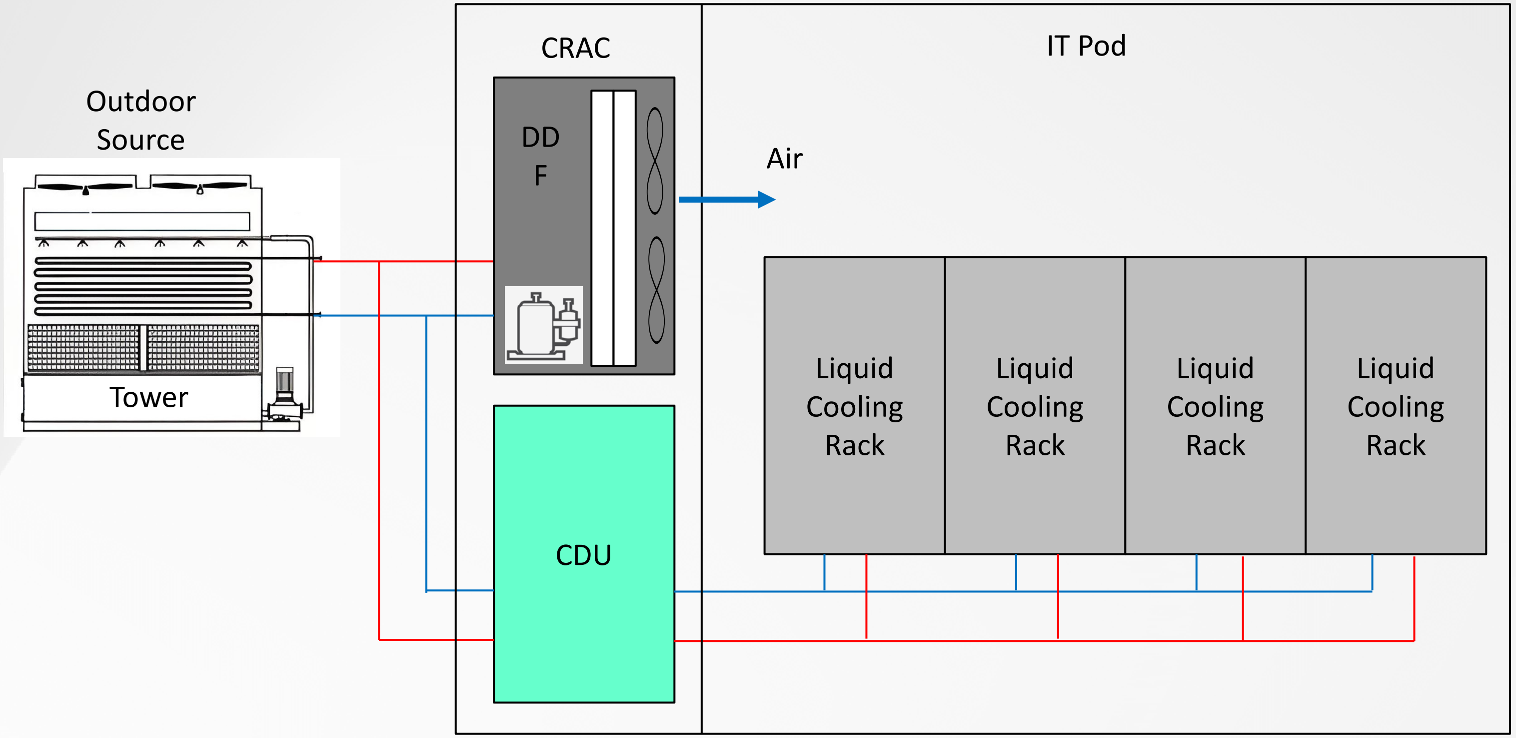
Independent System3: AHU + D2C



Structure Features

1. Suitable for hyperscale data center with large flat building
2. High efficiency
3. Prefabrication production to reduce installing work
4. WUE is close to 2, need more water

Tower Shared and Dynamic Dual Free-Cooling+D2C



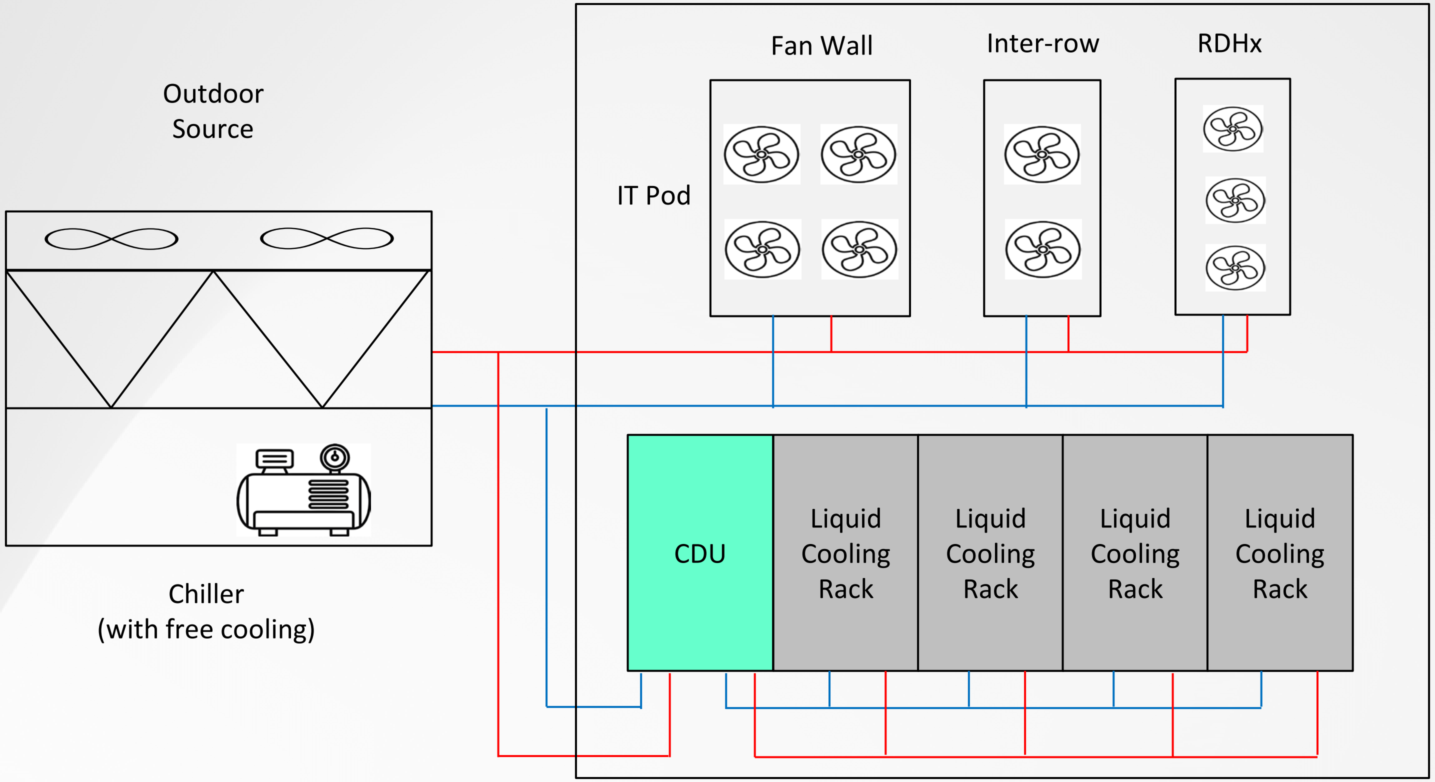
Structure Features

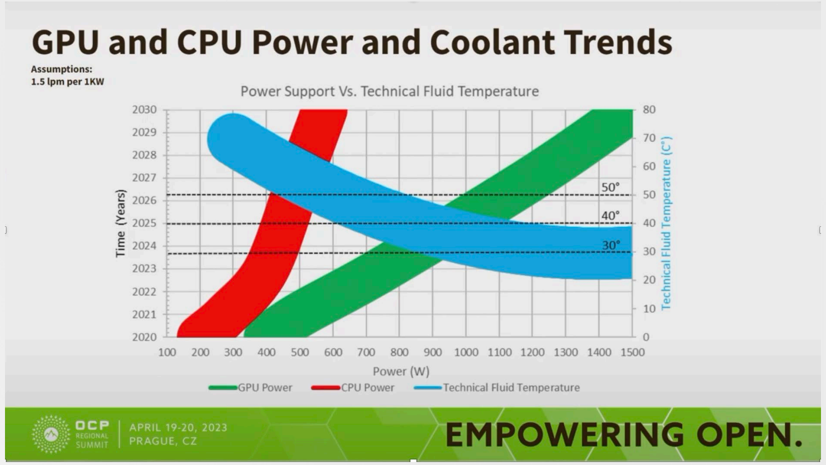
1.Water free cooling and fluorine pump for supplement

2. Easy to change the cooling capacity between air cooling and liquid cooling

3. Lower Capex and flexible of IT cooling types suitable for colocation

Chiller Shared + D2C

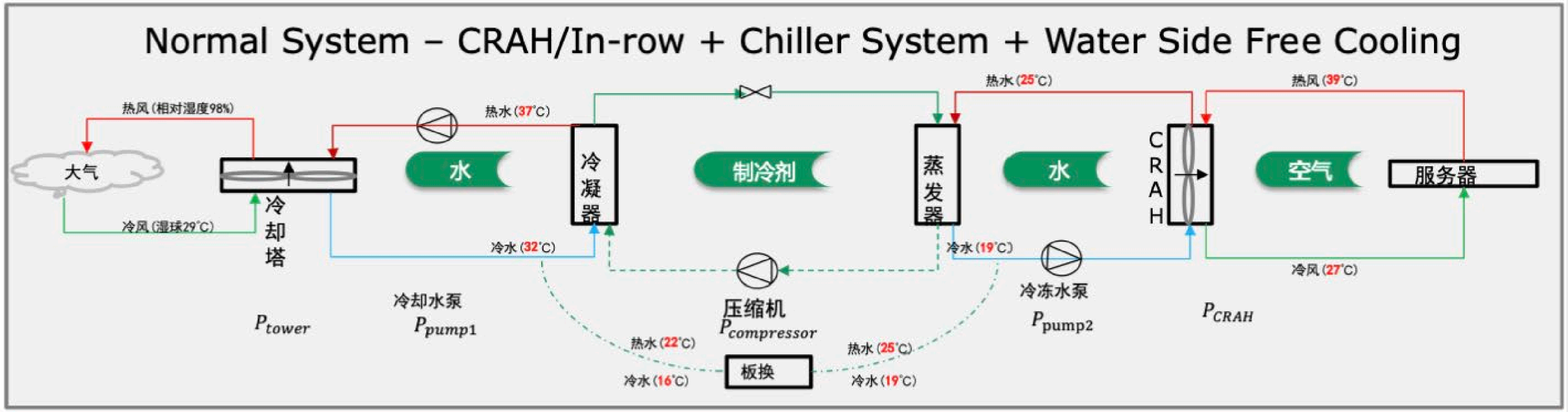




Structure Features

1. Broad environment adaptability with chiller
2. Suitable for HD over 100kW/Rack and higher TDP of Chips with lower secondary inlet water Temp.
3. Various air cooling terminal to choose
4. With free cooling modular could improve PUE

The Valuable Cooling Tech We Could Follow up



CONTECT US

Linked In — DeepKnowledgeCommunity

Web Ad — https://www.deepknowledgecommunity.com

Email — dww@deepknowledgecommunity.com

DeepKnowledge WeChat Public Platform

EnergyKnowledge WeChat Public Platform