The AI Revolution is Coming

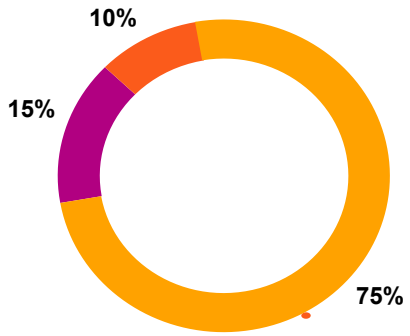
Clayton D’Souza
Senior Manager, Application Engineering

5th December 2024

Global presence, local expertise

NYSE: VRT

Market Segment[1,2]



- Data Centers
- Communication Networks
- Commercial & Industrial

Worldwide

Manuf. and Assembly Locations: **22**
Service Centers: **240+**
Service Field Engineers: **3,500+**
Technical Support/Response: **190+**
Customer Experience Centers/Labs: **19**

~$6.9B
USD revenue

~27,000
employees globally

130+
countries served

50+ years
in the industry

#1 in thermal
management[3]

#1 in 3-phase large UPS[4]
and power switching & distribution[5]

Note: [1]based on FY 2023 revenue; [2]Market segment rounded to nearest 5%; [3]Dell'Oro Data Center Physical Infrastructure reporting 2023. [4]Omdia Uninterruptible Power Systems (UPS) Hardware Tracker 2023, >250kva; [5]Omdia Data Center Power Distribution Equipment Tracker 2023. All else, company information as of December 31, 2023.

Vertiv is a leader in power and cooling, supporting impactful market segments in all world regions

Leading innovator with most complete critical digital infrastructure portfolio

Power Management

Power Train - We have all the "cars"

Medium- and Low-Voltage Switchgear/Switchboard

3-Phase Uninterruptible Power Supply (UPS) Systems

Power Distribution, Transfer Switches

Busbar

DC Power

Thermal Management

Thermal Chain - We have all the "links"

Air Handlers

Direct Expansion

Chillers

In-Room Cooling

Rear-Door Heat Exchangers

Direct-to-Chip Coolant Distribution

Immersion Cooling

IT Systems

IT Systems Set - We have the components from distributed IT to gigawatt data center sites

1-Phase UPS

Rack PDU

Racks and Enclosures

Rack Cooling

IT Mgmt. Devices, Software, High-Performance KVM

Infrastructure Solutions

Integrated Solutions
for outdoor and indoor IT whitespace

Power Modules

Smart Modules

Mega Mods

Integrated Aisles and Rows

Services

Project and Lifecycle Services
extend value of install base and maximizes market expertise

Lifecycle Services

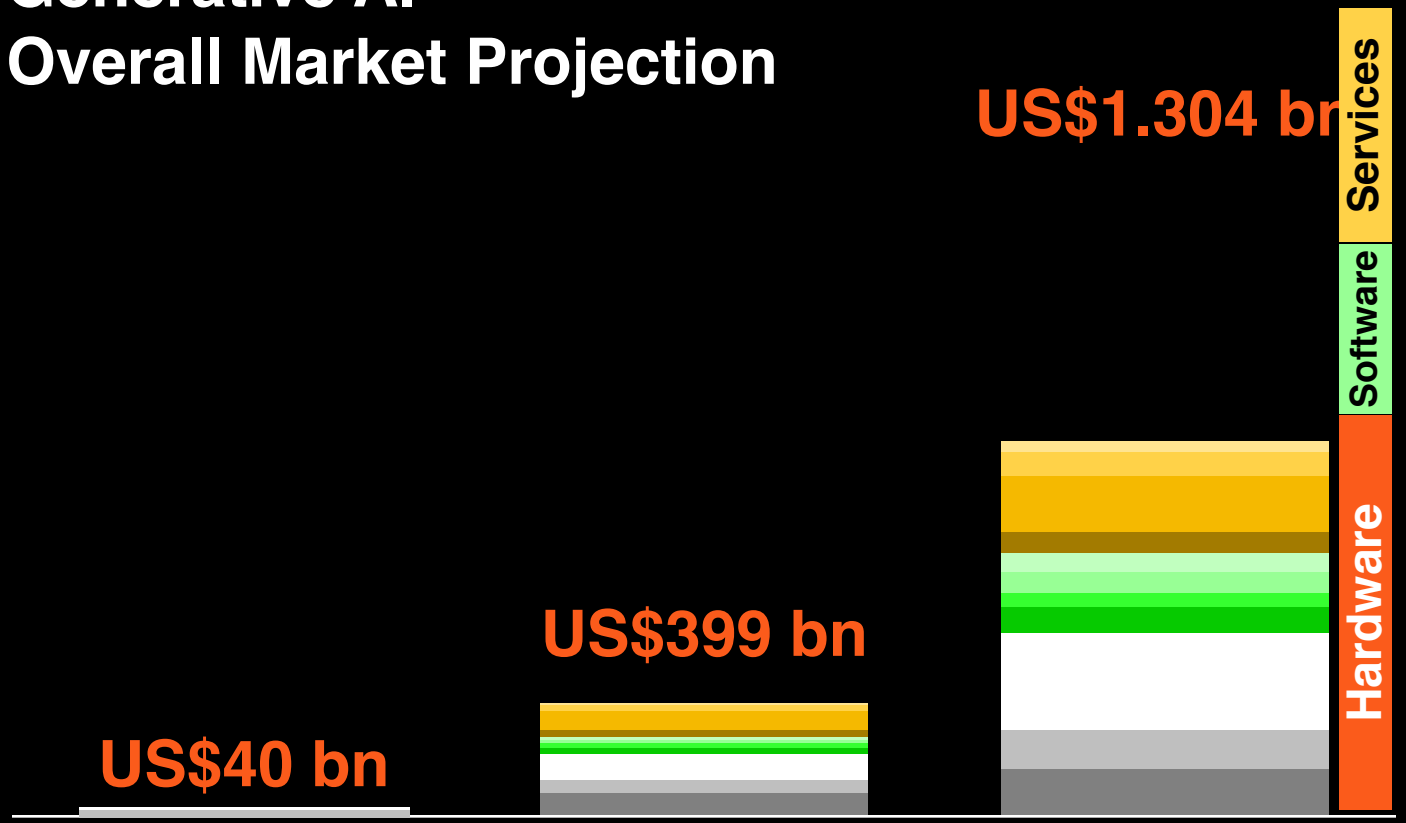
Project Services

Digital Services

Portfolio strength and uniqueness with end-to-end coverage

Bloomberg

Generative AI Overall Market Projection



1%	5%	12%
of total technology spend		

AI represents an extraordinary opportunity.

Bloomberg projects generative AI to be a $1.3-trillion business by 2032.

Source: Bloomberg Intelligence

Technology

Higher coding productivity

Consumer

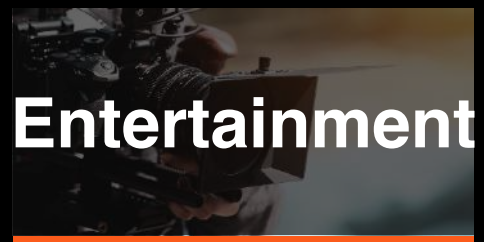
Conversion rate increase

Biopharma

Research timeline reduction

Financial Institutions

Higher fraud detection accuracy



Entertainment

Higher quality of animated images

Insurance

Customer service cost optimization

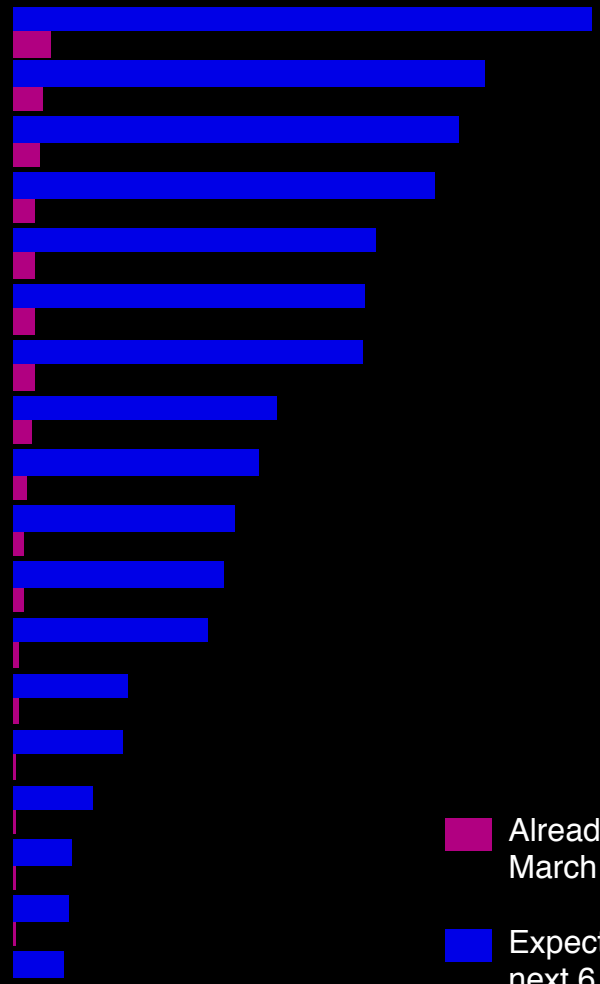
AI use-cases are already delivering encouraging results to businesses in every industry.

Are you already seeing the impact of AI in your business?

Source: BCG, GitHub, Vue.AI, Nature, ScienceDirect, BusinessWire

Share of companies using various forms of AI applications

Goldman Sachs

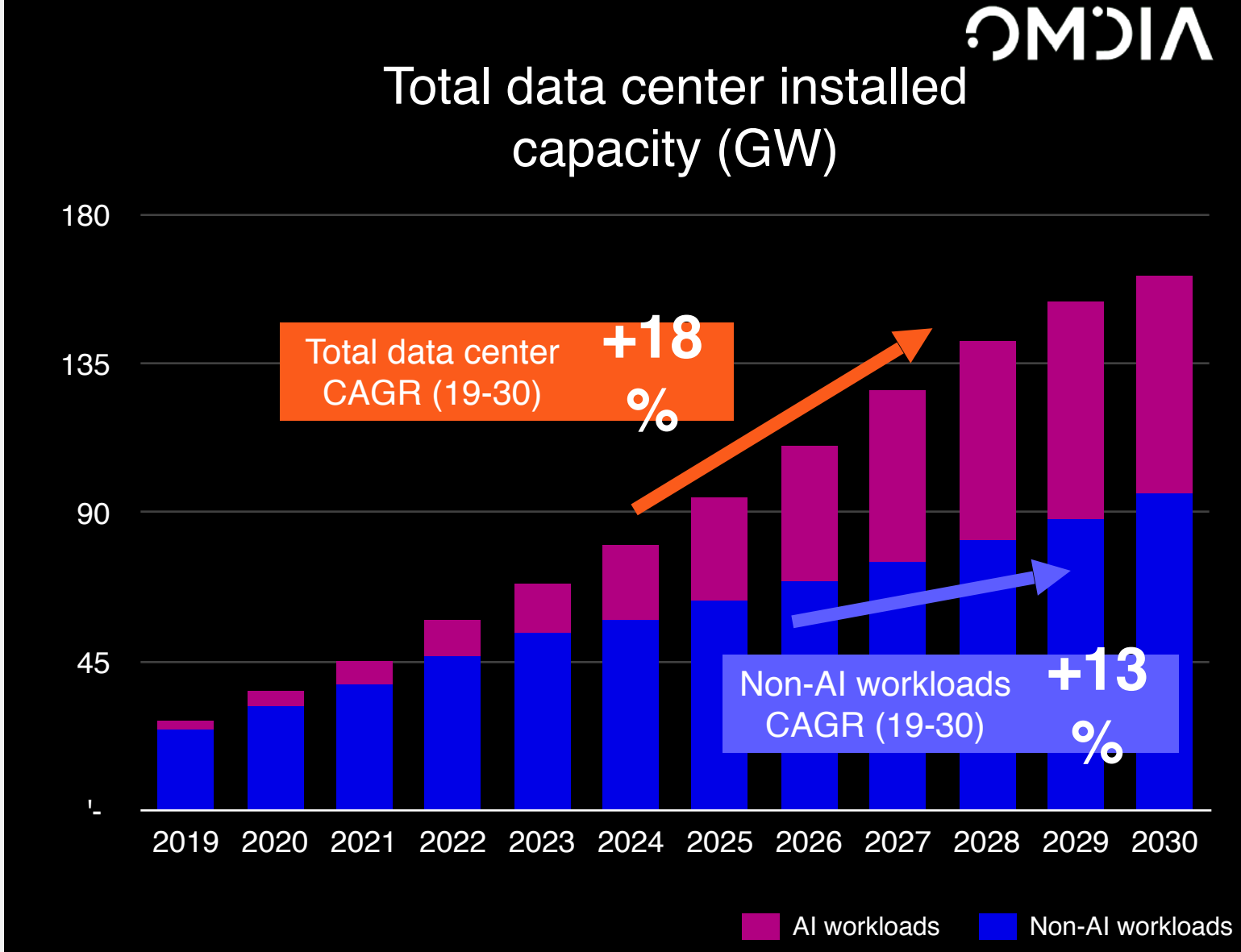


Source: Goldman Sachs

Companies are racing to adopt AI and not be left behind competition.

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

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

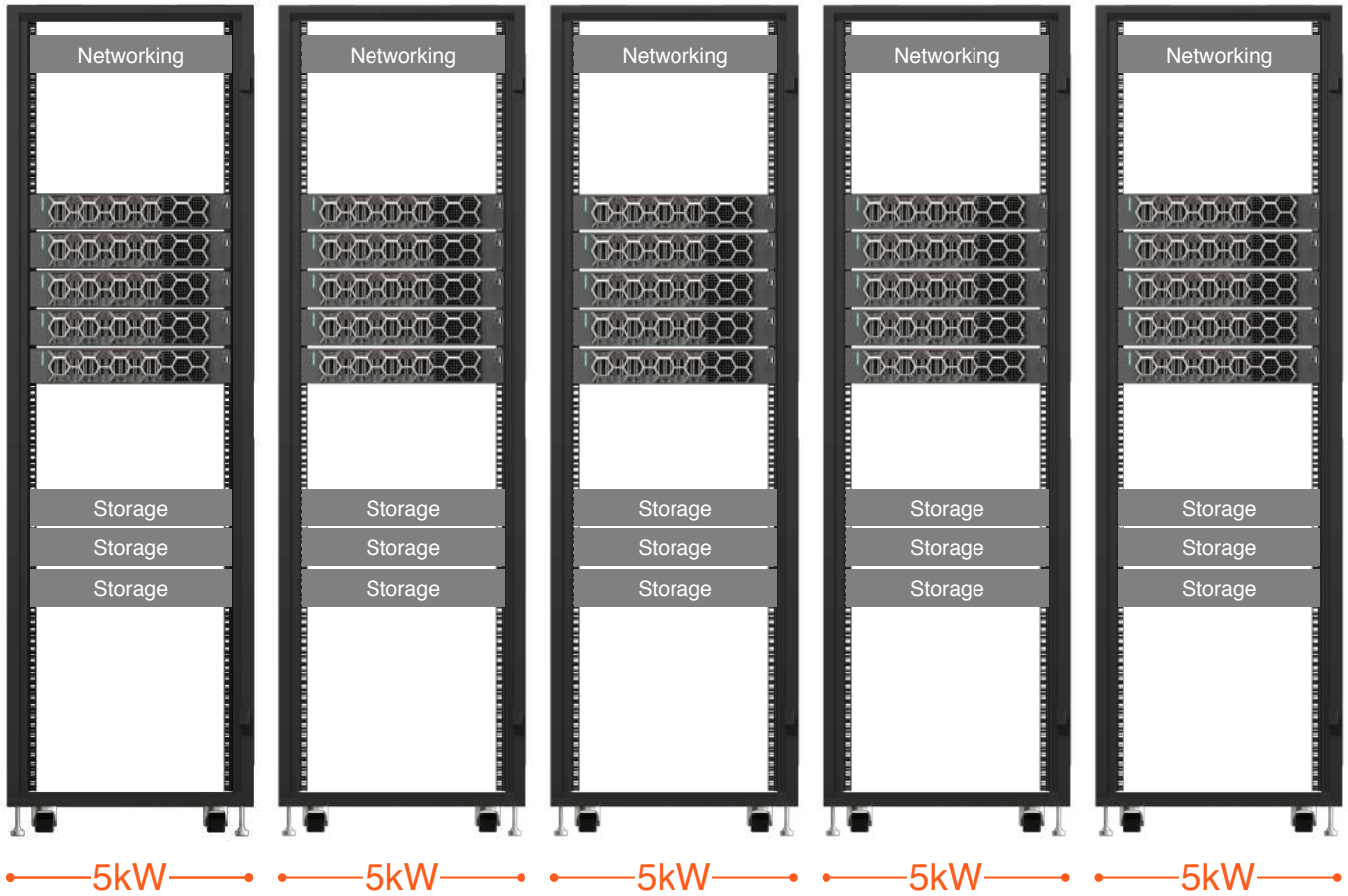


Source: Omdia

50kW

workload?

Conventional IT solution.



Traditional workloads were able to keep densities low by splitting loads across multiple racks.

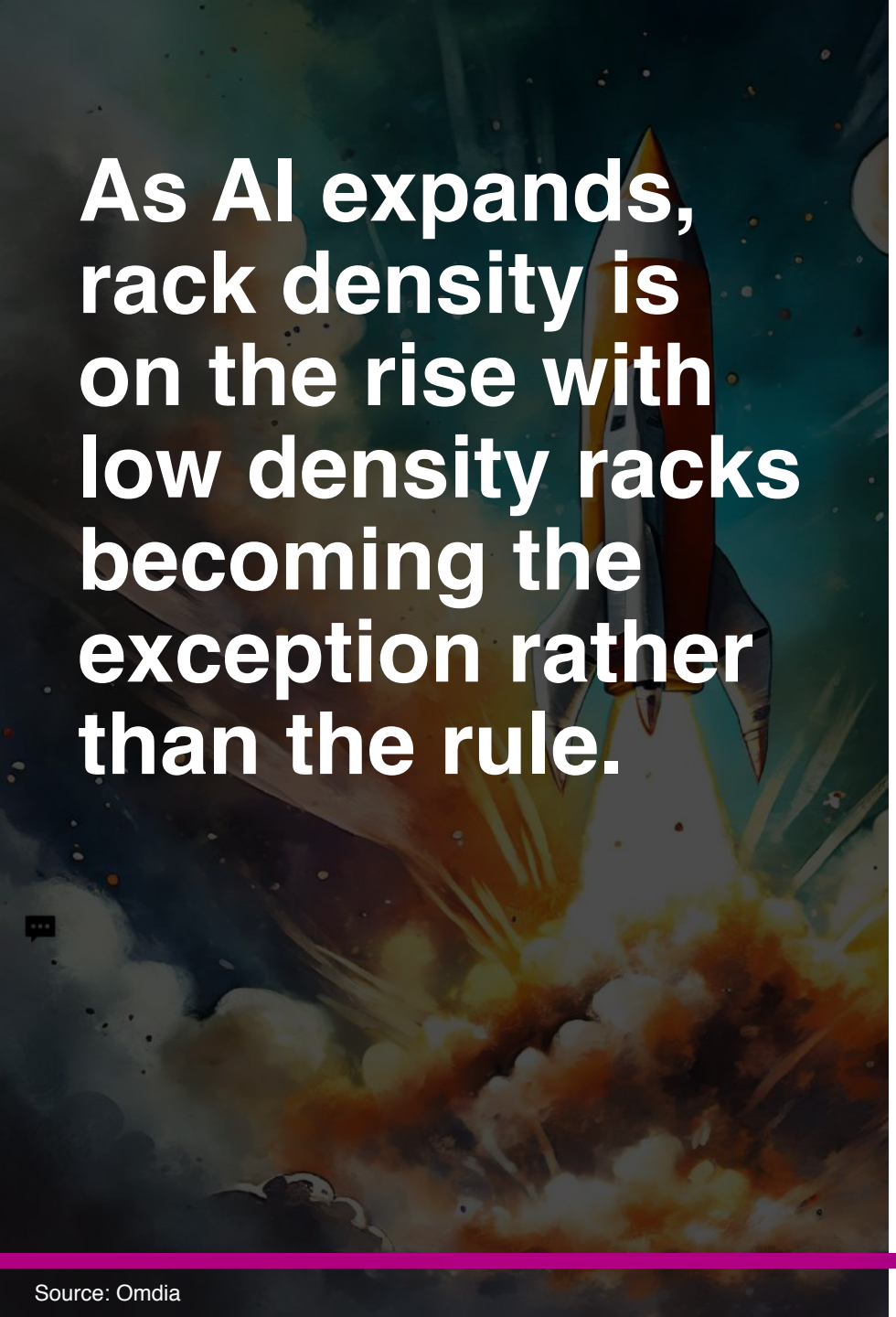
50kW

workload?

AI solution.

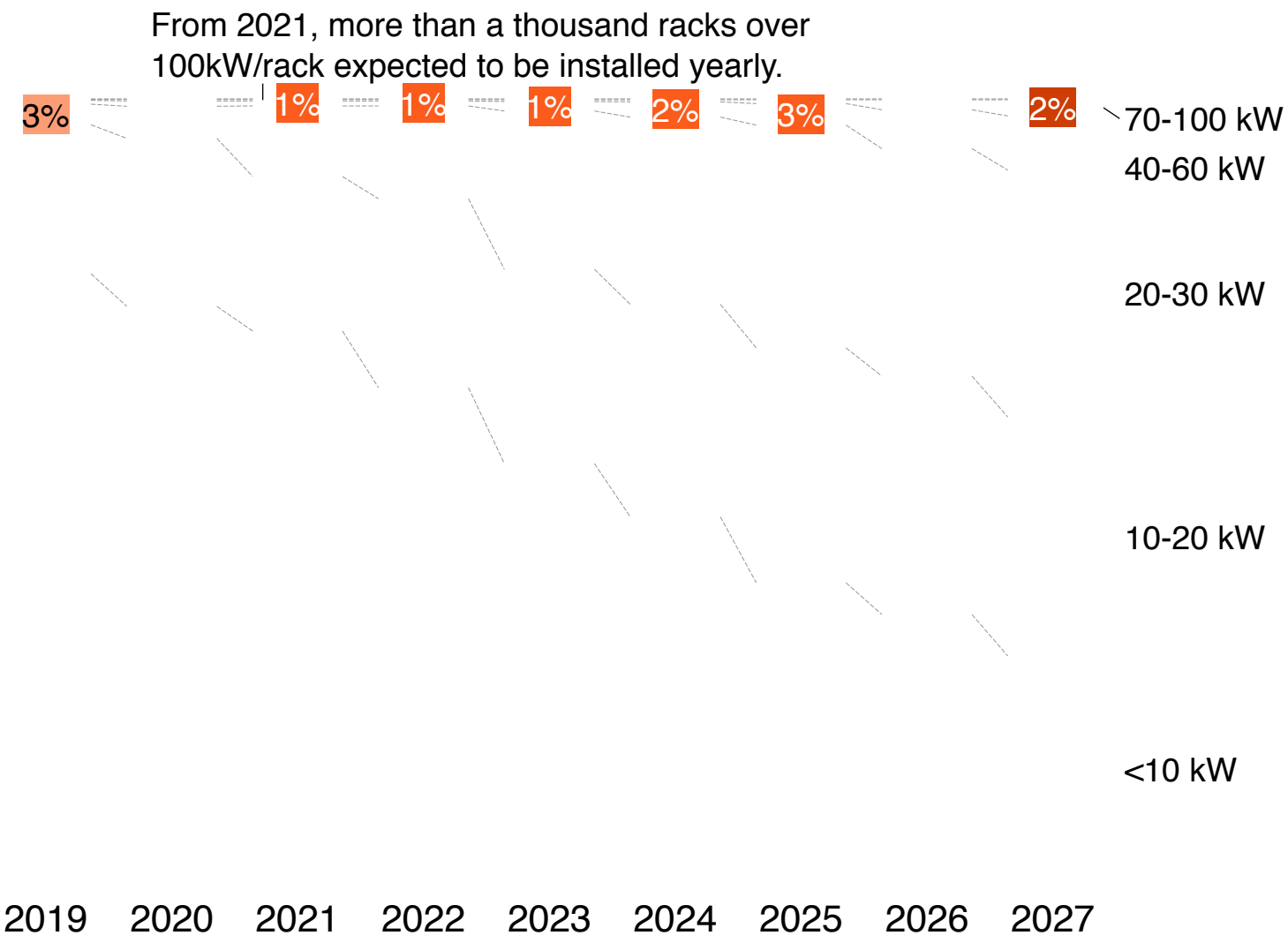
AI workloads are driving rack density up: servers are drawing more power per rack unit, and they need to be closer together to minimize network latency and costs.

Traditional workloads were able to keep densities low by splitting loads across multiple racks.



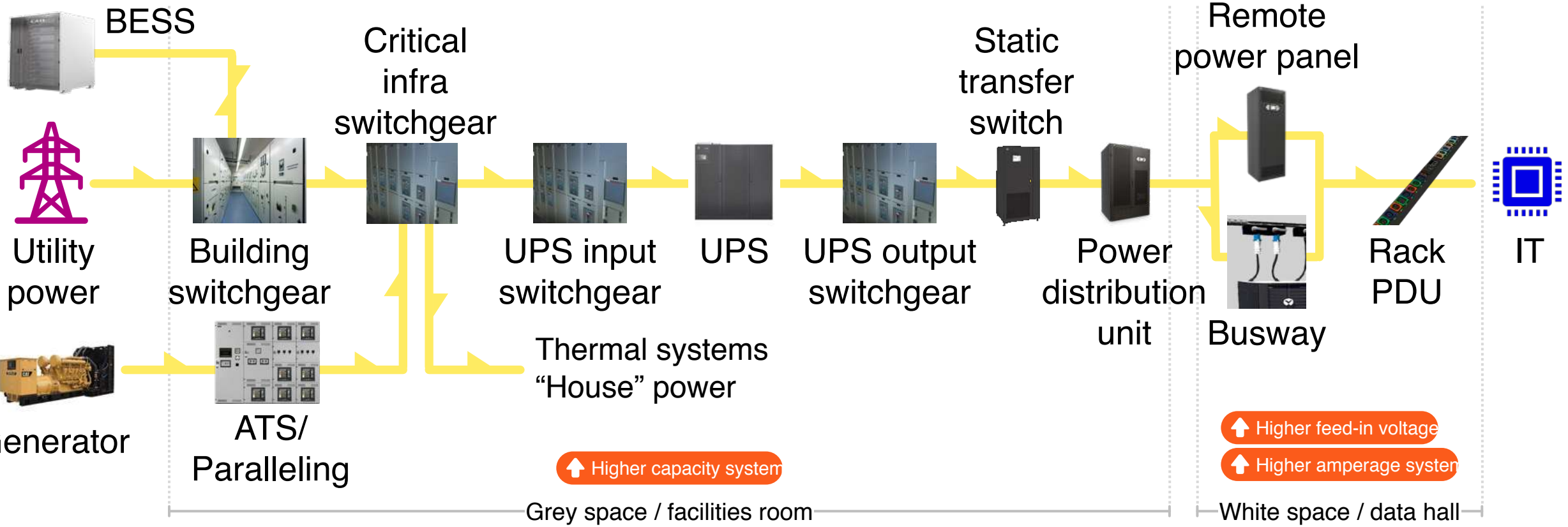
As AI expands, rack density is on the rise with low density racks becoming the exception rather than the rule.

Number of racks installed by rack density.



Source: Omdia

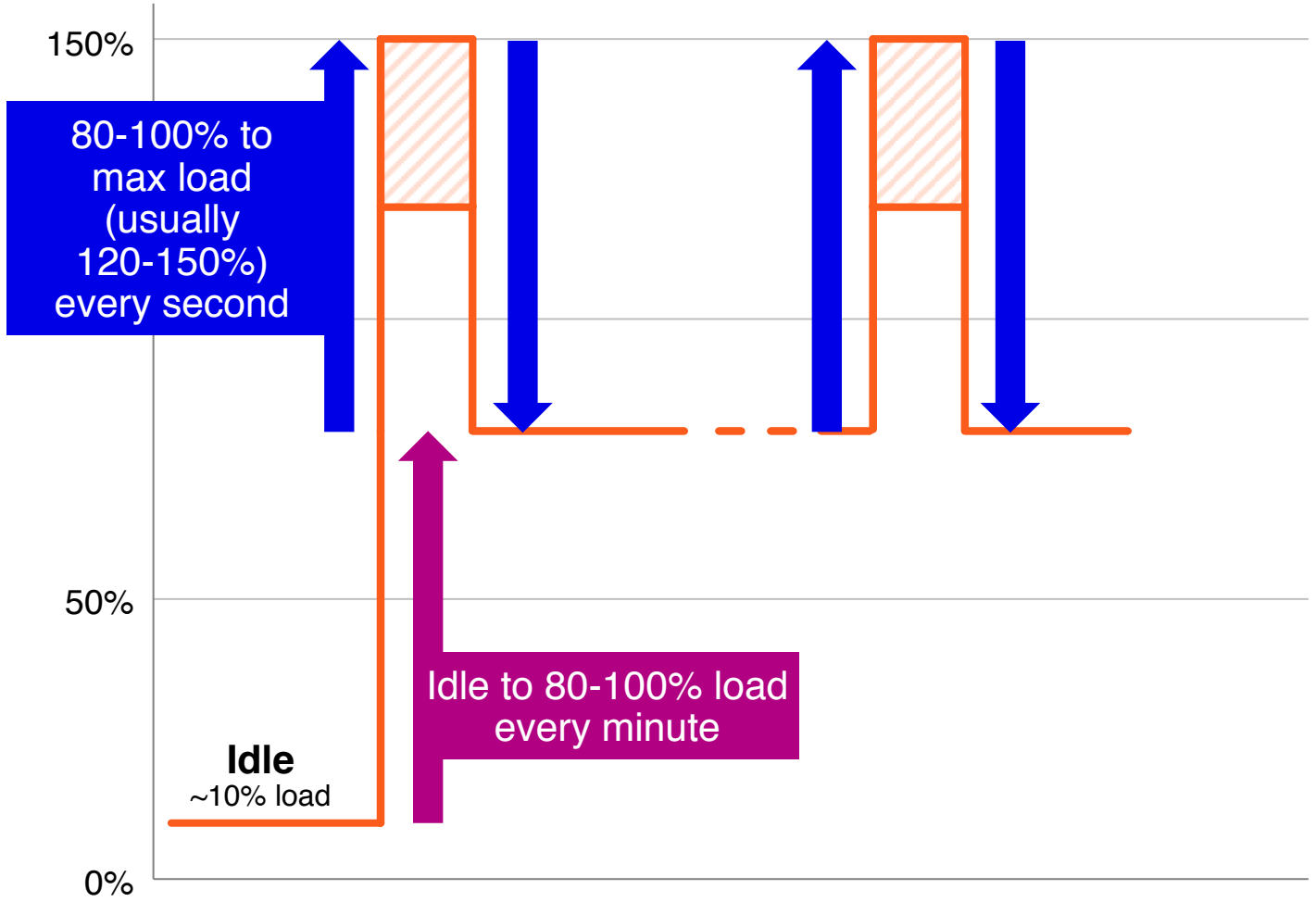
Power Train



Power is critical to the success of AI deployments. IT and facilities teams must collaborate and partner with specialists who understand the power train end-to-end.

As GPUs become ubiquitous in data halls, they bring in challenging load profiles for supporting power infrastructure.

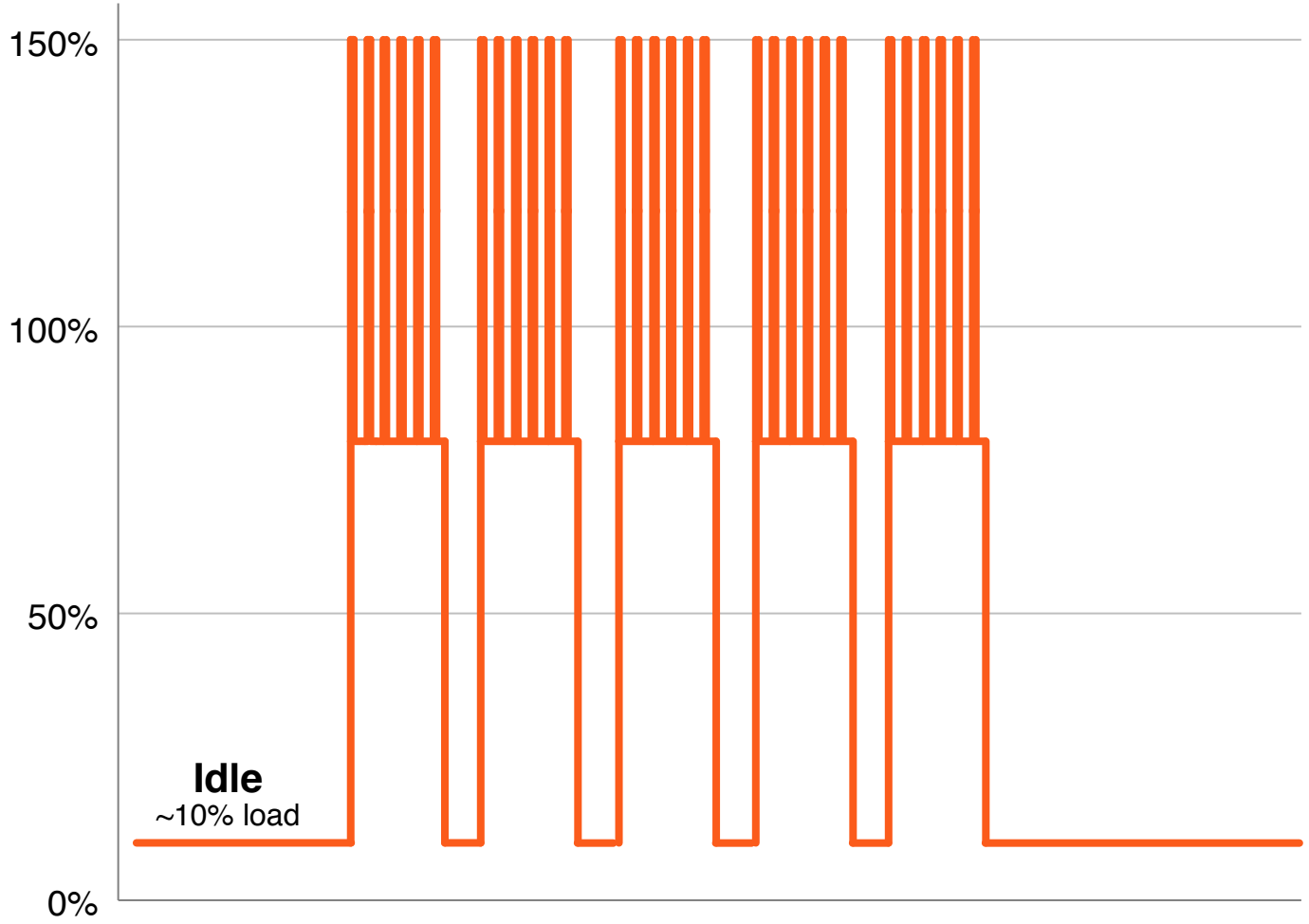
AI load profile



Load profile of GPUs running AI workloads is unique and very different from traditional IT equipment

As GPUs become ubiquitous in data halls, they bring in challenging load profiles for supporting power infrastructure.

AI load profile

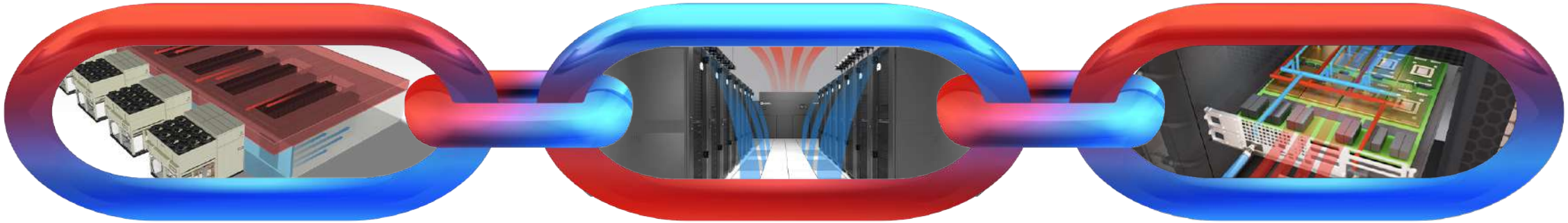


Pulses result in system instabilities, UPS must be able to avoid triggering battery and generator

Thermal Chain

Complete High-Density Thermal Chain, from Chip to Reuse

AI and high-density workloads may require extensive capacity upgrades across the entire power train.



Facility/Outdoor

Heat Rejection

Air Handlers

Chillers

Room/Row

Heat Collection

Direct Expansion

In-Room Cooling

Rack Cooling

Rack/Server

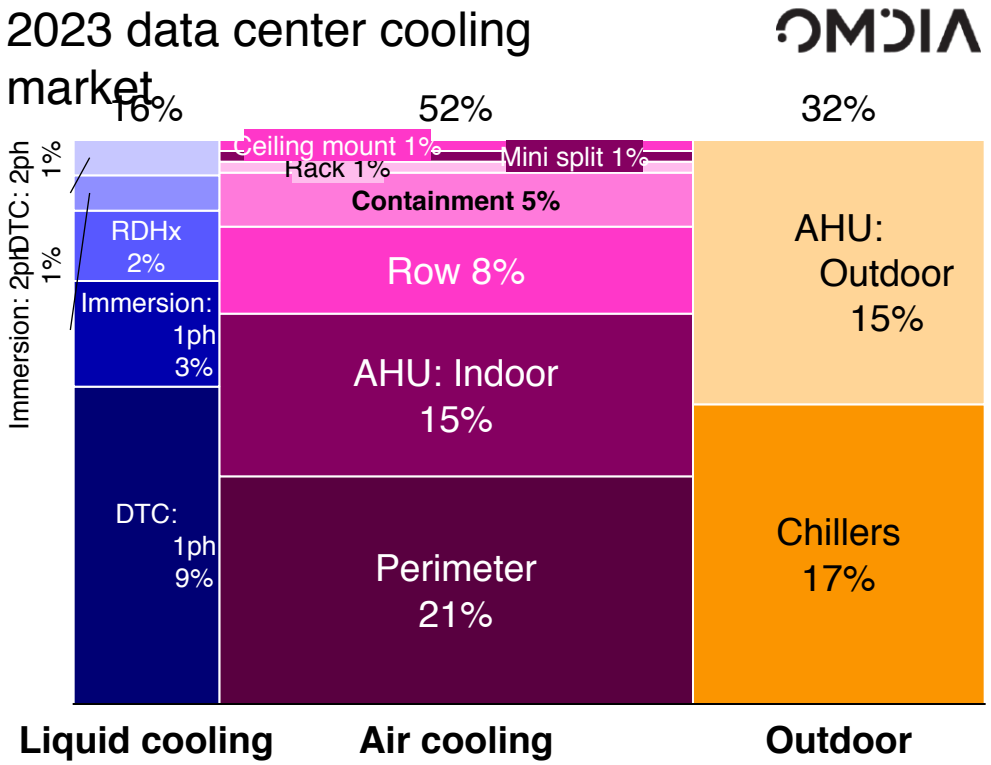
Heat Collection

Rear-Door Heat Exchangers

Direct-to-Chip Coolant Distribution

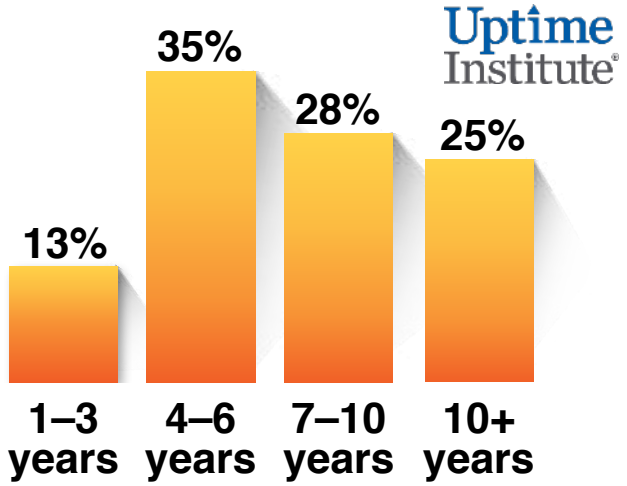
Immersion Cooling

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.



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?

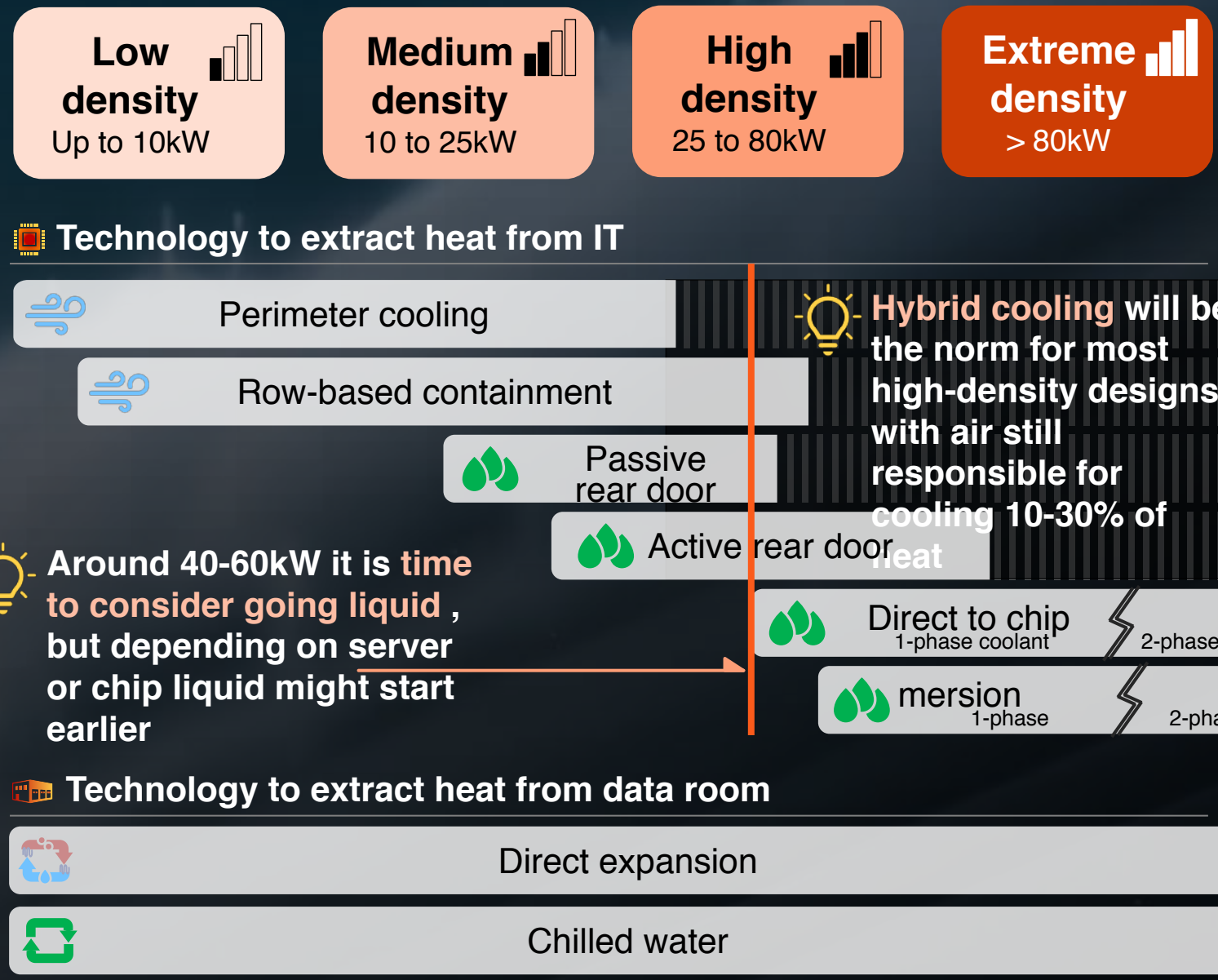


Source: Omdia, Uptime Institute

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.

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.



Cold plate

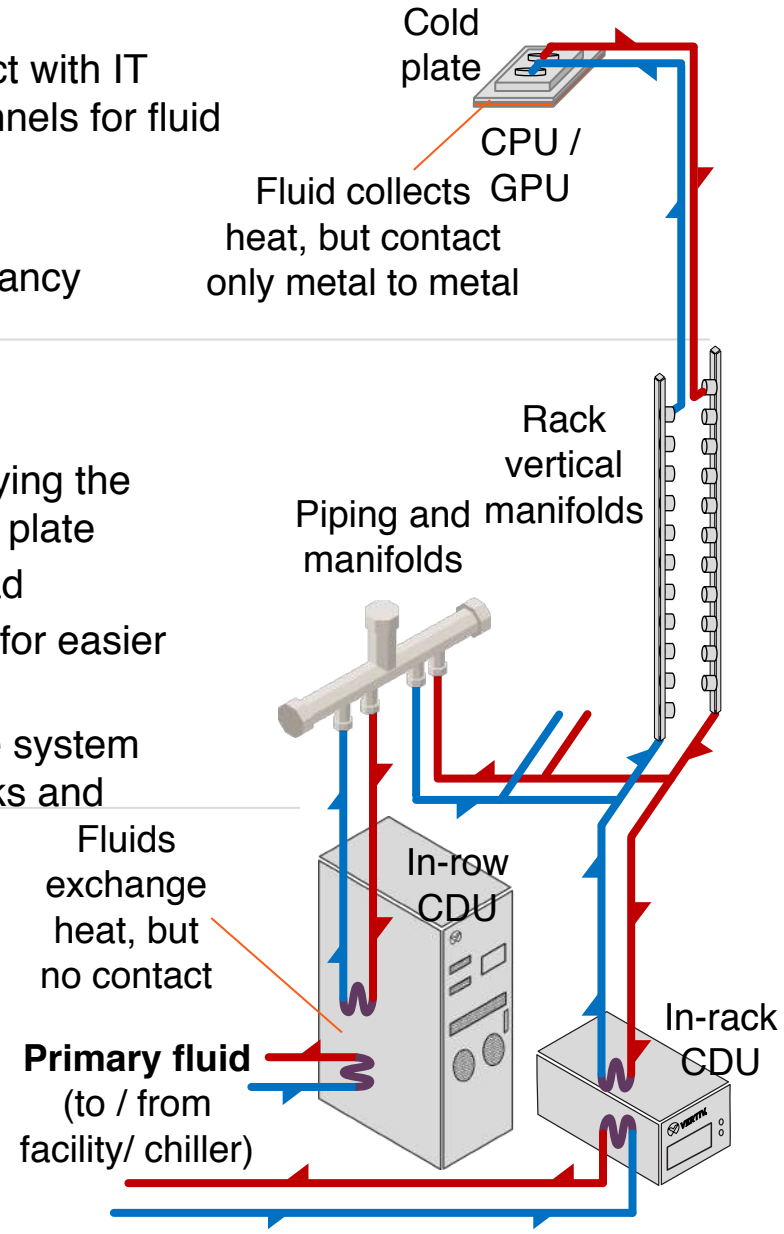
- Highly conductive metal in contact with IT equipment pierced by micro-channels for fluid to go through and collect heat
- Considerable variety in designs
- Critical with little room for redundancy

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.

CDUs are the beating heart of liquid cooling systems, collecting heat from SFNs and transferring it to primary fluid loops.

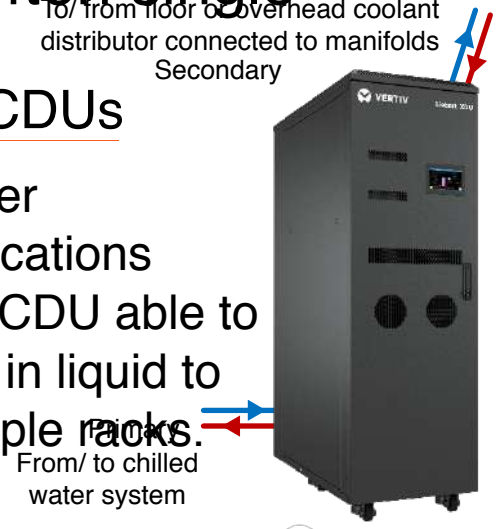
- Critical component **distributing coolant** to entire system via SFN to cold plates.
- Intelligence of cooling solution **controlling flow rates**, ensuring **steady flow rates and inlet temperatures** to IT with embedded monitoring and controls.
- **Filtration** to extracting impurities out of SFN.
- Pump and power are key points of **redundancy** in the secondary fluid network (cold plate and manifolds are often single points of failure).

In-rack CDUs

- Designed for smaller loads or for modular scale-up.
- Able to contain blast rates in case of faults.

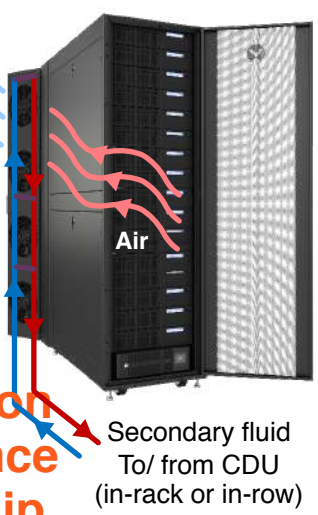
In-row CDUs

- Larger applications with CDU able to feed in liquid to multiple racks.



Liquid cooling design options also include rear-door heat exchangers and immersion cooling.

Rear-door heat exchangers



- Highly efficient **air-to-liquid** solution.
- Available as **standalone cooling solution** or **air-assistance in direct-to-chip** designs providing room neutrality.
- Back-of-the-rack attachments. **adaptable to existing rack architectures**.
- Not as sensitive to contaminants as cold plates, **less challenging fluid management**.

Immersion cooling

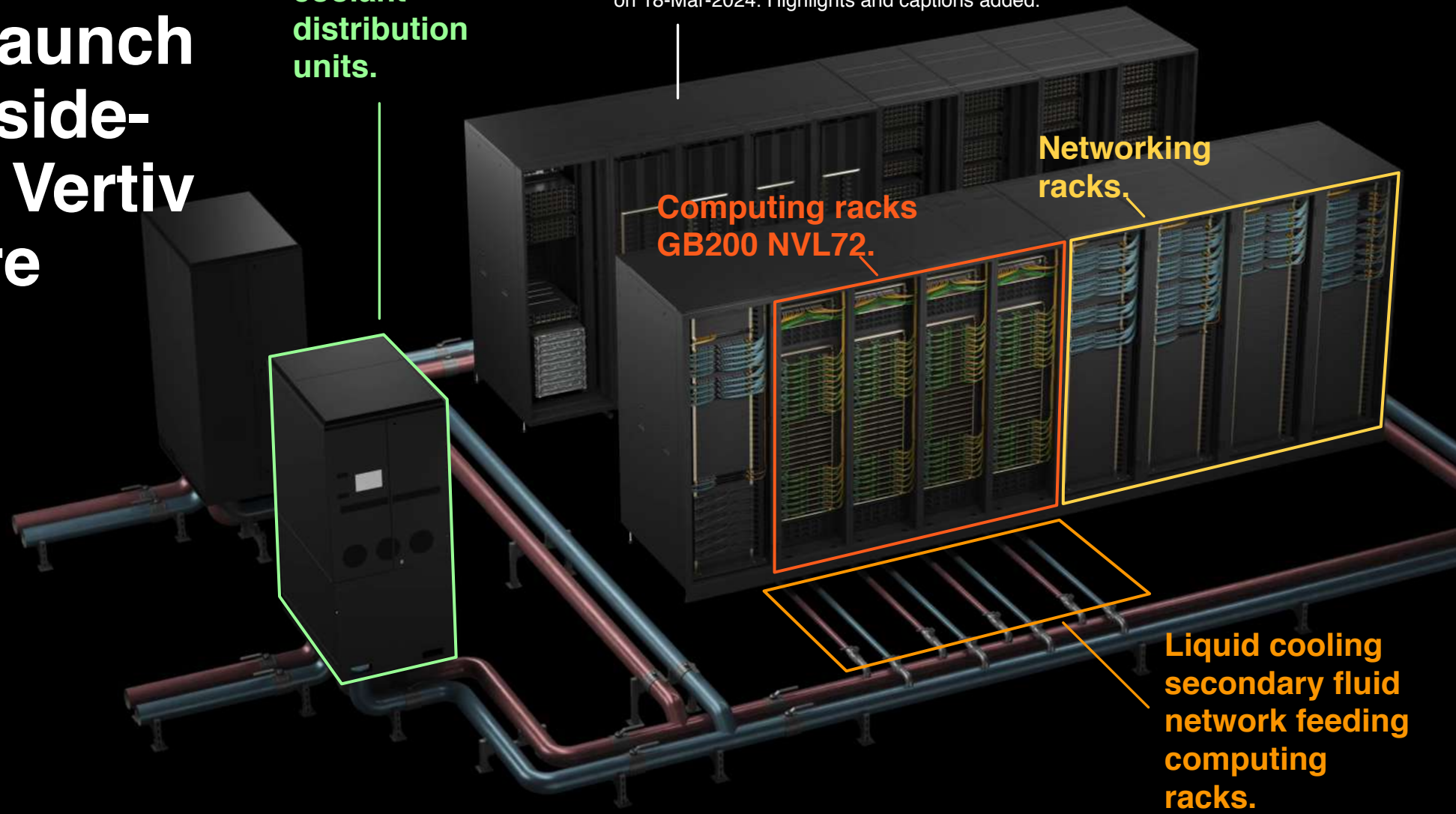
- Highest **heat rejection efficiencies** amongst liquid cooling technologies, with **no need of air cooling** in most designs.
- Higher adoption in smaller **standalone** or **pilot** applications.
- Scalable technology with **two-phase fluid** options being able to handle extreme

There is not one-size-fits-all answer in liquid cooling. Best technology will depend on site and application.

NVIDIA introduced its flagship launch at GTC2024 side-by-side with Vertiv infrastructure solutions.

NVIDIA DGX SuperPOD™ with GB200 systems

Rendering as presented by Jensen Huang at GTC2024 on 18-Mar-2024. Highlights and captions added.



NVIDIA Partner

Source: NVIDIA GTC2024 Keynote Speech

Vertiv is working side-by-side with Intel to develop power and cooling infrastructure and support its highest-grade AI solutions.

- Joint development of a product strategic for Intel's growth in the AI space.

- Two different liquid cooling design options:
  - Refrigerant-to-air-cooled solutions handling one rack up to 40kW.

  —or—

  - Refrigerant-to-liquid cooling system able to remove up to 160kW of heat load.

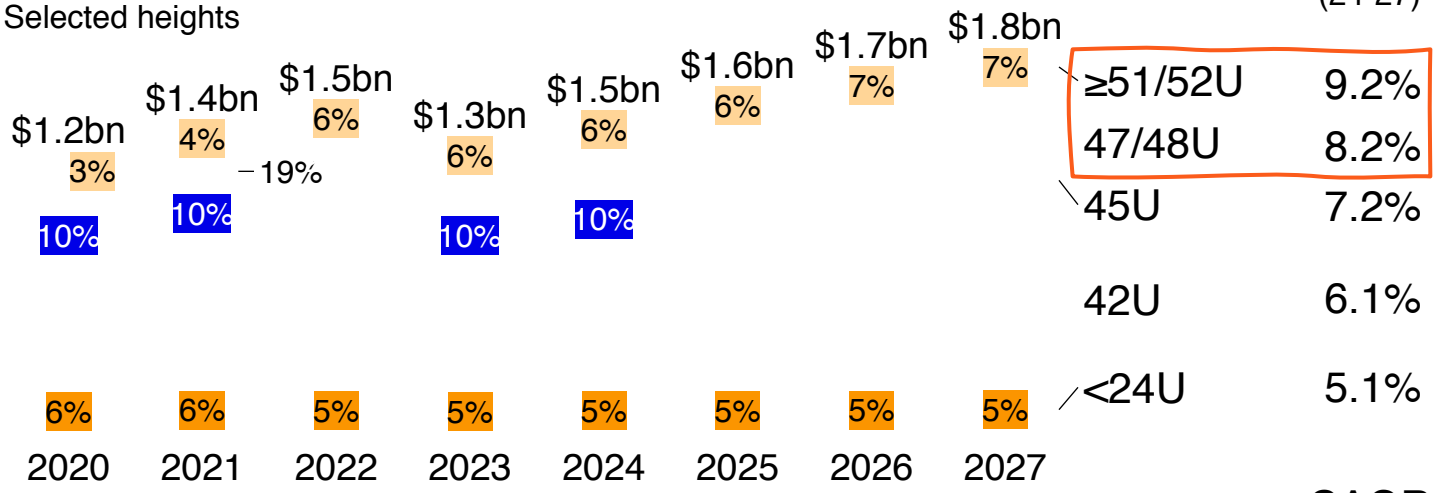
> "To support increasing thermal design power and heat flux for next-generation accelerators, Intel has worked with Vertiv and other ecosystem partners to enable an innovative cooling solution that will be critical in helping customers meet critical sustainability goals."
>
> — Devdatta Kulkarni, Principal Engineer intel

Scan me

42U still the dominant rack height, but taller racks slowly expanding with faster growth than shorter sizes.

Global rack enclosures market by rack height

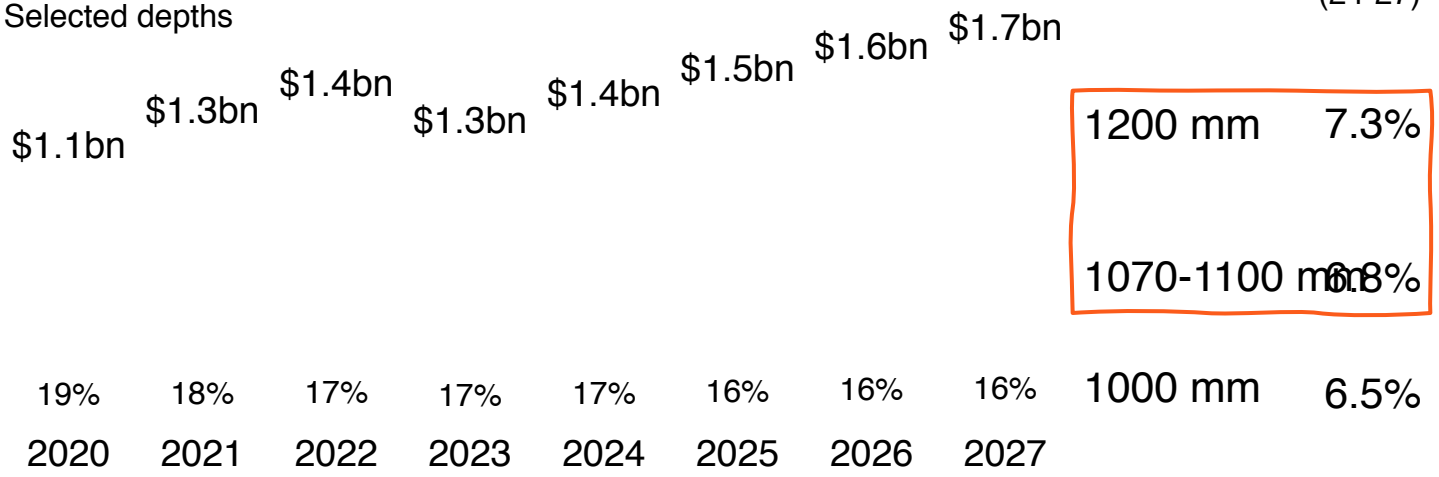
Selected heights



	CAGR (24-27)
≥51/52U	9.2%
47/48U	8.2%
45U	7.2%
42U	6.1%
<24U	5.1%

Global rack enclosures market by rack depth

Selected depths



	CAGR (24-27)
1200 mm	7.3%
1070-1100 mm	6.8%
1000 mm	6.5%

Source: Omdia

AI is changing the IT rack, making them taller, wider and deeper.

Higher networking racks are being redesigned to handle heavier IT gear.

- **Stronger racks** to handle additional weight of IT gear and liquid volume and torque.
- Larger racks to allow for **HD rPDUs, manifolds**, **larger cables**, inlet/ outlet **hoses**.
- **Retrofitting existing datacenters may become challenging** due to change in rack footprint.
- Racks must have **drip pans with sensors** that communicate to a central management controller.
- **Smart shutoff valves** to rack and row must be used to automatically shut off liquid flow in severe leak situations.

Source: NVIDIA, Supermicro

Vertiv™ 360AI

Rack solutions

Inferencing & Edge AI

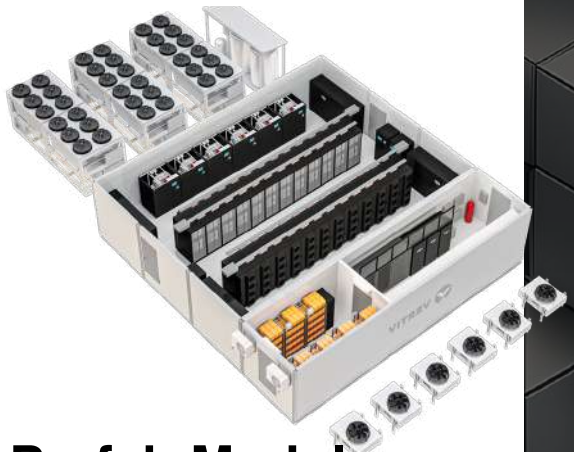
- AI testing and pilots
- Edge inferencing

Row and room Solutions

Enterprise & Data Center AI

- AI lab
- Model training and inferencing
- AI row within existing data hall

Facility & data center solutions



Prefab Modular AI Data Centers

- AI model training at scale
- AI factory

Modularity and repeatability are key elements to improve ROIs of AI deployments, no matter their size.

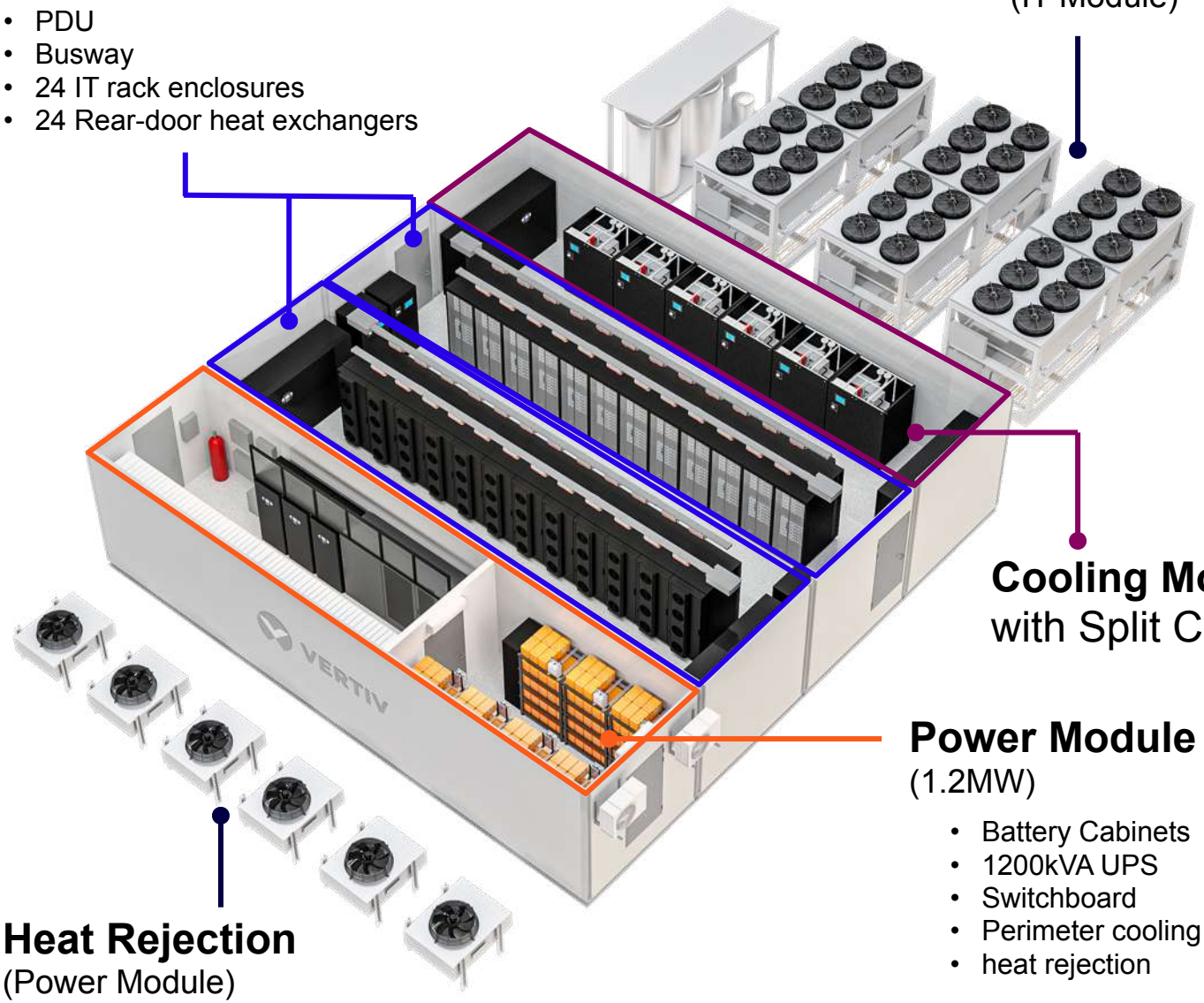
High-Density Modular Reference Designs – 1MW at 50kW per Rack

Hybrid Modular Design with
4 Modules Bayed Together

2x IT Modules
(50kW per rack)

- PDU
- Busway
- 24 IT rack enclosures
- 24 Rear-door heat exchangers

Heat Rejection
(IT Module)



Cooling Module
with Split Chillers

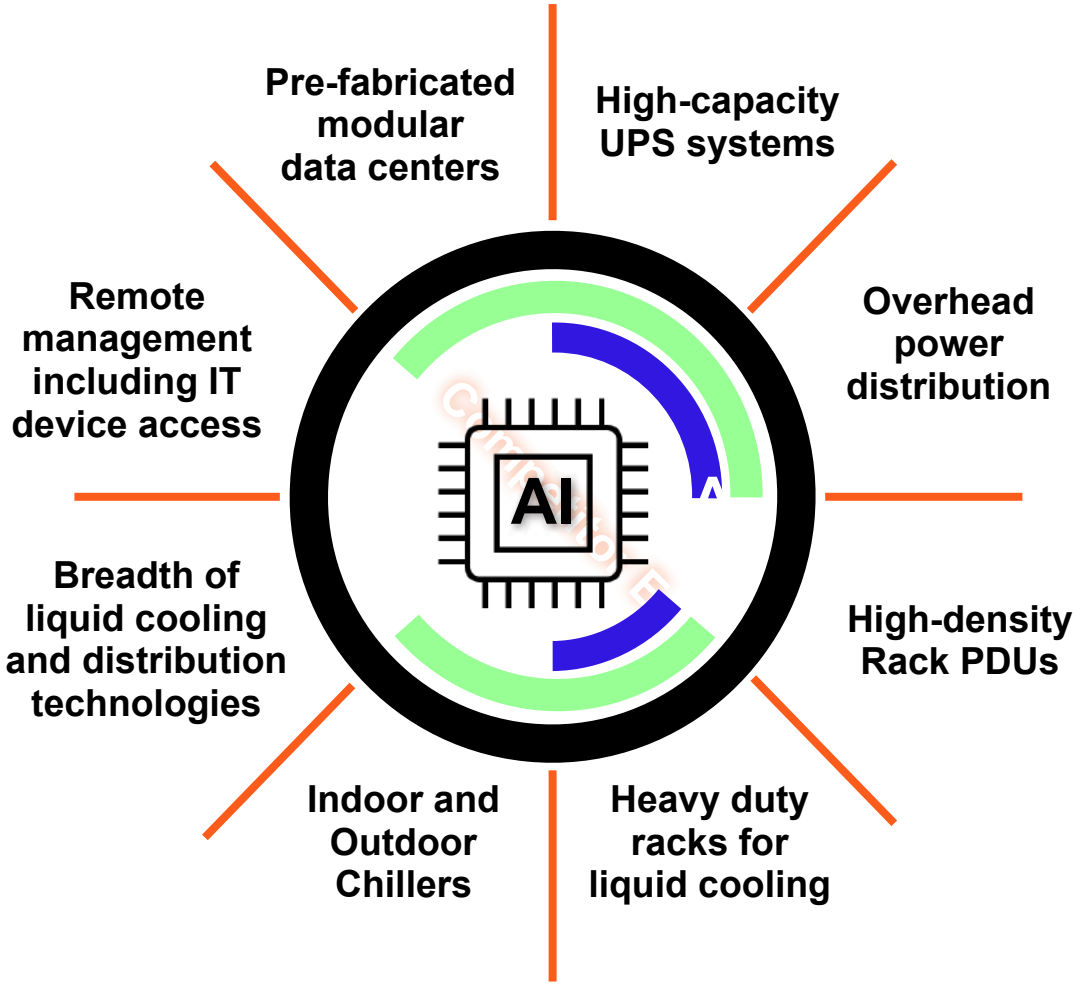
Power Module
(1.2MW)

- Battery Cabinets
- 1200kVA UPS
- Switchboard
- Perimeter cooling units +
- heat rejection

Heat Rejection
(Power Module)

Only Vertiv™ Can Power and Cool AI Workloads from End-to-End

Cross-disciplined Engineers and Solution Architects available to help tailor solutions



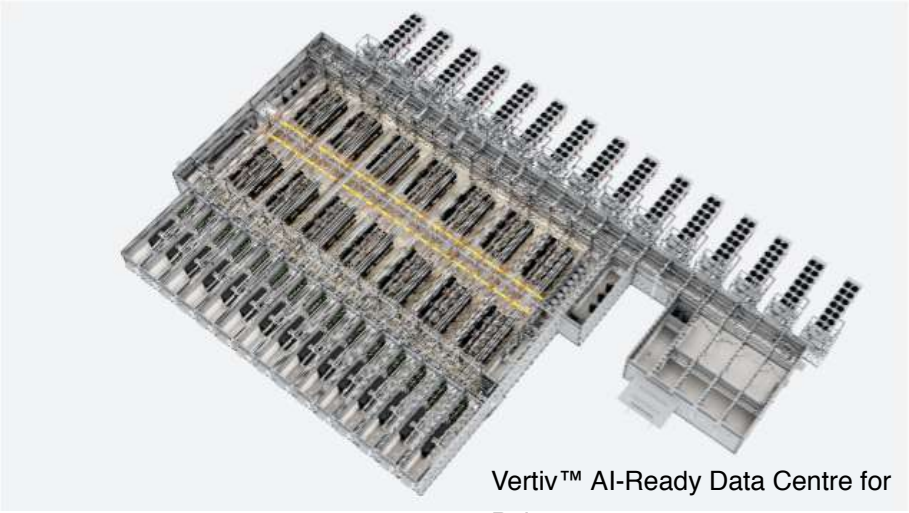
End-to-End Solution Capabilities

- ☑ Evaluate
- ☑ Design & Model
- ☑ Prefabrication
- ☑ Deployment
- ☑ Commissioning
- ☑ Maintenance

Vertiv™ AI-Ready Data Centre for Polar

Developing the future of AI in Norway

Polar is partnering with Vertiv to drive innovation and resilience in their Norwegian data centres, utilising Vertiv's trusted infrastructure solutions for enhanced efficiency.

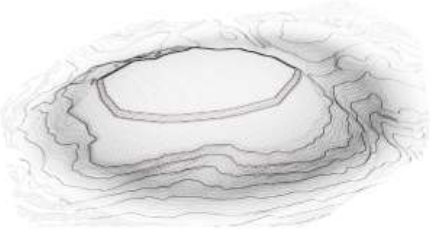


Vertiv™ AI-Ready Data Centre for Polar

Site location: Norway

Resilient Interconnectivity

100% operating on hydro-power



The solution is made for IT loads of 12MW and up to 115kW per Rack (D2C), with N+1 topology on both electrical and thermal systems.
All made for quick and easy assembly on-site in Norway.

Vertiv™ CoolChip CDU is ideal for high-density environments

Polar is partnering with Vertiv to drive innovation and resilience in their Norwegian data centres, utilising Vertiv's trusted infrastructure solutions for enhanced efficiency. This collaboration aims to boost performance for AI and HPC applications.

This data centre structure provides cutting-edge reliability, scalability, and safety to meet the demands of modern high-performance IT operations.

Vertiv is providing Polar with a complete modular AI solution consisting of a data hall with 28 modules, 15 power modules, 15 chillers and skibbed buffer tanks, 2 MMR modules, double hydromodules and logistic modules; all interconnected with corridor modules.


VERTIV™


NVIDIA
Partner