Data Center Forum, Oslo 2024

Data Center Evolution: Preparing for AI Workloads

Presented by Erik Gjesdal

Life Is On Schneider Electric

AI Business Transformation

AI vs ML vs DL

Responsible AI

Today's Topic

AI disrupts Data Centers

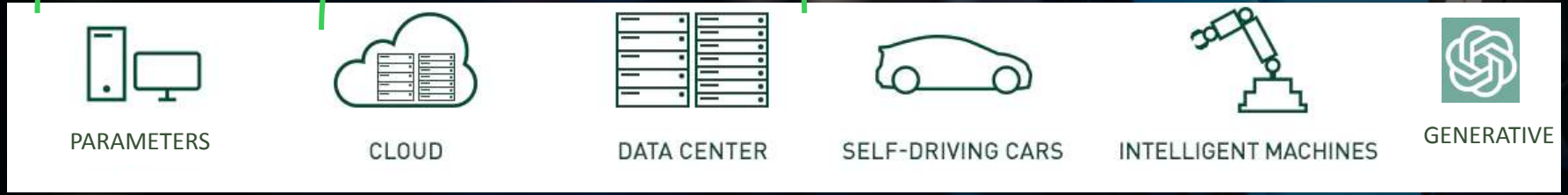
IT, Power, Cooling & Racks

Development

TRAINING

Deployment

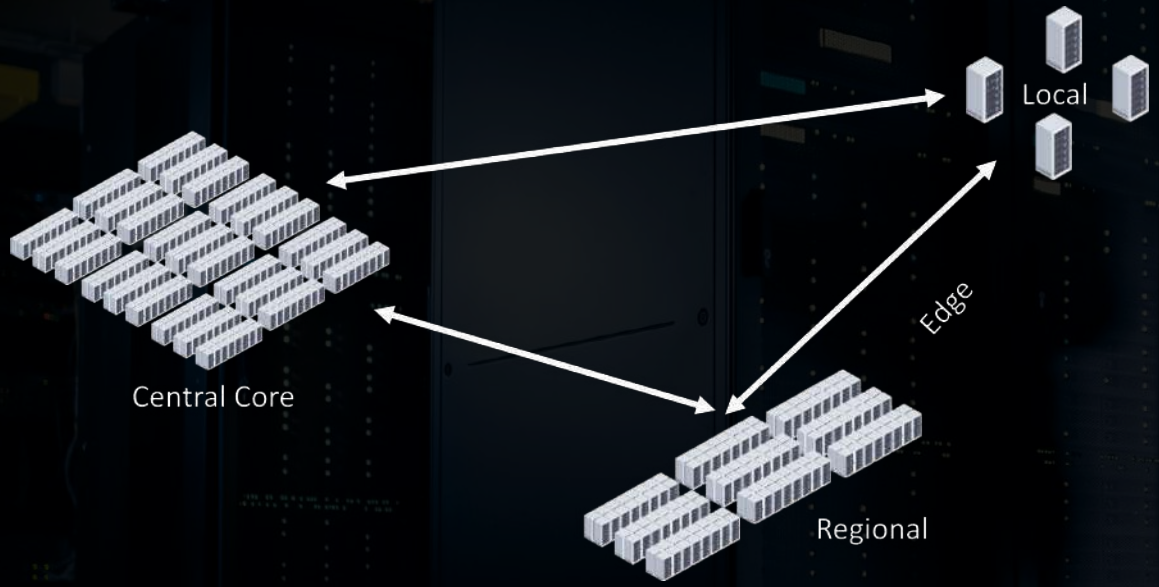
INFERENCE



Workloads developed from scratch

Workloads use **compressed models** developed during training

Training in large facilities | Inference at the edge



Power and Cooling Disruption ---------- to Massive Scaling

AI is driving data center growth at core initially - then the edge

SE Estimate	2023		2028	
AI Consumption for AI Workloads & Phys. Inf.	4.3 GW		13.5 - 18 GW	
AI, % of Total DC Power Consumption	8%		15 - 20%	
AI Workload, Central vs. Edge	95% Central	5% Edge	50% Central	50% Edge

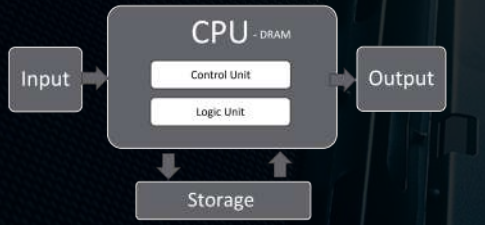
Source: Schneider Electric™ Internal Analysis

All AI data centers need to be deployed in a sustainable way

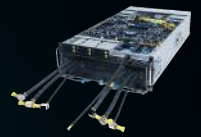
- **Set a bold**, actionable strategy
- **Implement efficient** data center designs
- **Drive efficiency** in operations
- **Transition to** renewable energy
- **Decarbonize** supply chain

AI Training requires new server components and architecture

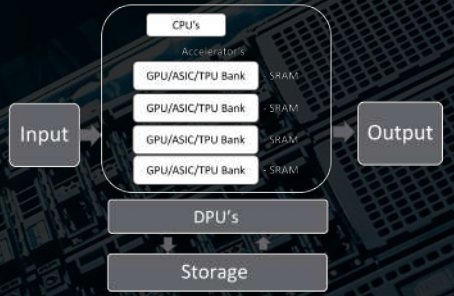
2-4U 3-5 kW



Von Neumann bottleneck, memory wall plus limited CPU throughput will not work for AI training



8-10U 6-12 kW*



Leveraging pipelined back propagation parallelism

Thermal design power (TDP) is trending upwards…

AI Training

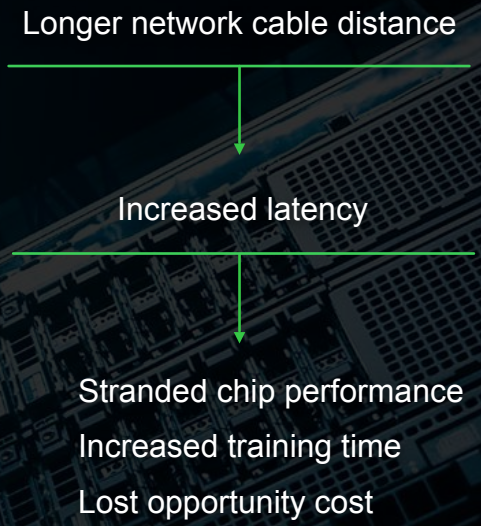
GPU	TDP (W)	TFLOPS (Training)	Performance over V100	TOPS (Inference)	Performance over V100
V100 SXM2 32GB	300	15.7	1X	62	1X
A100 SXM 80GB	400	156	9.9X	624	10.1X
H100 SXM 80GB	700	500	31.8X	2,000	32.3X

…leading to **higher server densities**

AI Training

Impractical to spread the load

Data Center Fabric network latency presents a constraint that drives up density



AI training racks are deployed in “clusters”

Legacy 800kW DC

100 – 150 Racks

120-400kW

800kW AI Training DC

8 Racks

240-800kW

AI disrupts Data Centers

Chip

Server

Rack

Cooling

Electrical Distribution

Monitoring

POWERING AI servers in these dense applications has challenges

Challenge	Guidance
120/208V distribution is impractical	415V distribution
Small power distribution blocks waste space	Increase block size
Standard 60/63A rPDUs are impractical	Multiple or custom rPDUs
Arc flash risks complicate work practices	Analysis & mitigation measures
Load diversity increases tripping risks	Design peak-to-average = 1
High rack temperatures increase risk of failures	High-temp rated connectors

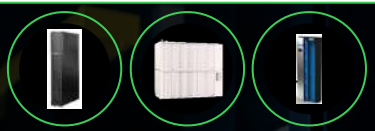
Challenges in POWERING AI servers in these dense applications EXAMPLE

POWERING AI servers in these dense applications has challenges

Challenge	Guidance
Air cooling is not suitable > 20 kW	Use liquid-cooled AI servers
Customized designs complicate deployment	Seek experts to assess design of proposed LC loads and facility
Unknown future TDPs risk obsolescence of designs	Design for both air and liquid cooling, and for ease of scalability
Inexperience complicates operation/maintenance	Seek experts to develop SOPs and MOPs for day-to-day operations
Liquid cooling increases risk of leaks in racks	Perform pressure testing & apply leak detection + leak prevention
Limited fluid options exist to operate sustainably	Use direct-to-chip with water to avoid GWP issues

Liquid cooling architecture

COMPLEMENTARY AIR COOLING



HEAT REJECTION UNITS

COOLANT DISTRIBUTION UNIT (CDU)

Standard RACKS present challenges in supporting AI applications

48U minimum height

750mm minimum width | 1200mm minimum depth

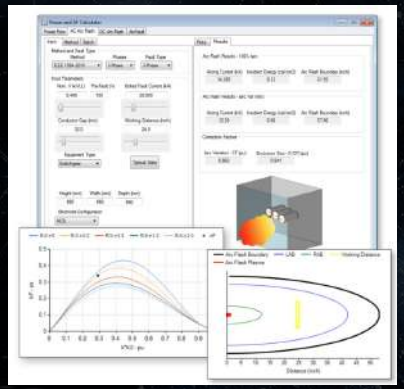
GUIDANCE

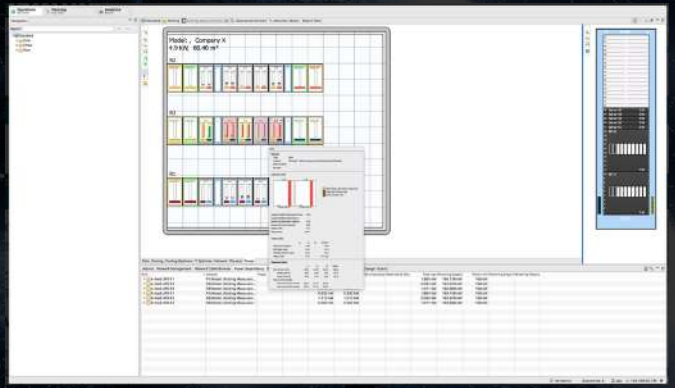
Weight-bearing
Static: 1,800 kg (3,968 lb)

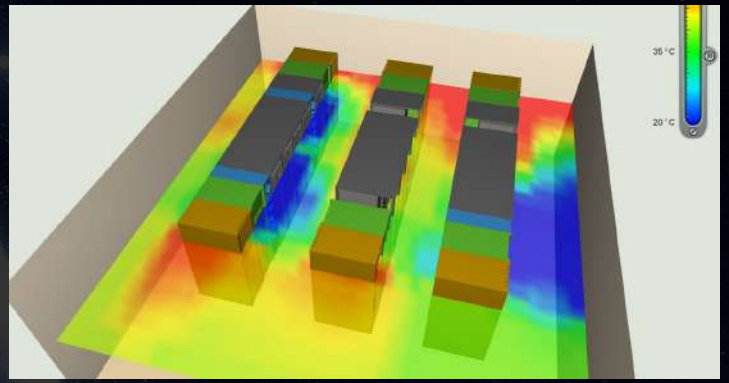
Dynamic:
1,200 kg (2,646 lb)

SOFTWARE TOOLS help address AI challenges

EXAMPLES







1. Power design tool coordinates breakers and analyzes arc flash risk

2. DCIM planning & modeling avoids exceeding available capacities

3. CFD ensures AI clusters don't impact nearby IT

AI driving disruption and scale

1. AI will drive rapid growth in all sizes of data centers globally

2. Training AI disrupts physical infrastructure:
power capacity, power distribution, cooling and racks

3. The next big challenge is deploying large scale inference working models at the edge in a sustainable way

White Papers for DataCenter Operators

Navigate the new landscape

WP67

28 reporting important metrics

White Paper 67: A Guide to Environmental Sustainability Metrics for Data Centers

WP110

AI: the game-changer for data center

White Paper 110: The AI Disruption. Challenges and Guidance for Data Center Design

WP133

Liquid cooling / AI workloads

White Paper 133: Liquid cooling distribution for high density / AI workloads

Get the Guides


Life Is On
Schneider
Electric