Smarter technology for all

Managing power in the AI Data Centre with Lenovo Neptune®

Jim Roche, Senior HPC Technical Architect
Sept 2025
jroche@lenovo.com

Lenovo

Powering the data center: the environmental impact

Until the world's power grid transitions to **low/no emission** sources, the largest contributor to IT carbon emissions is power consumption.
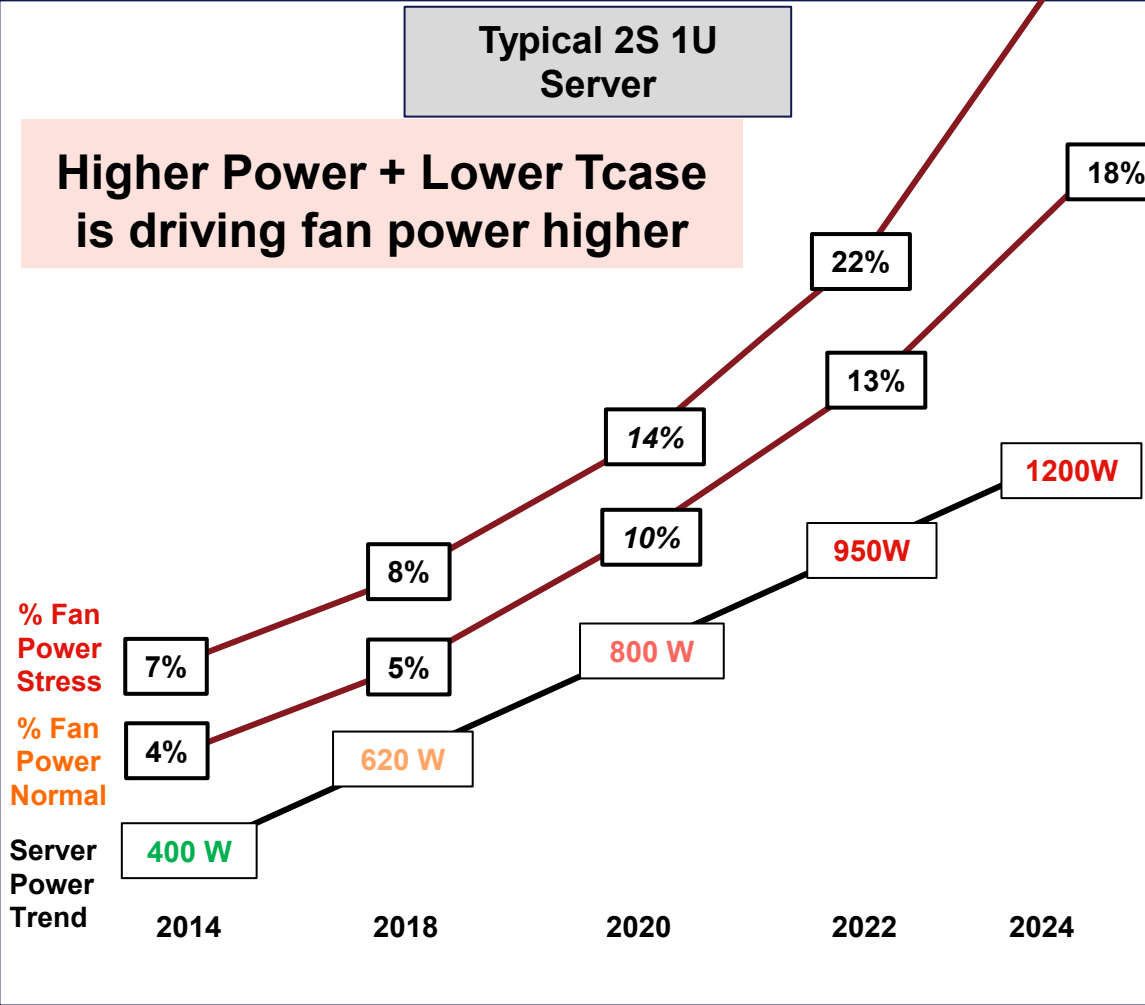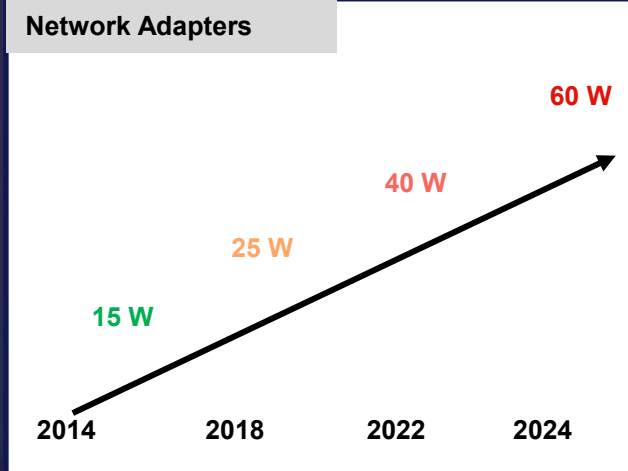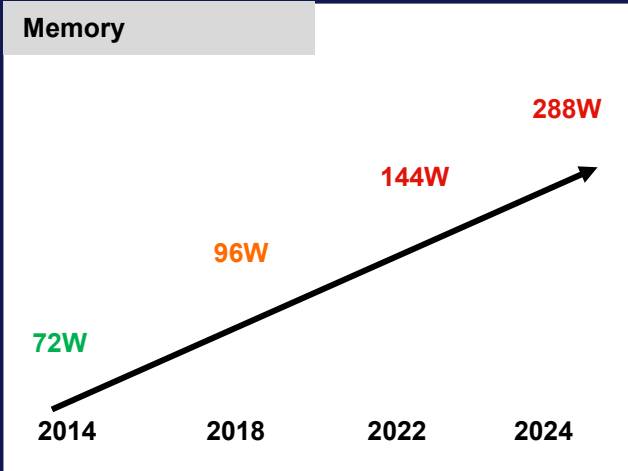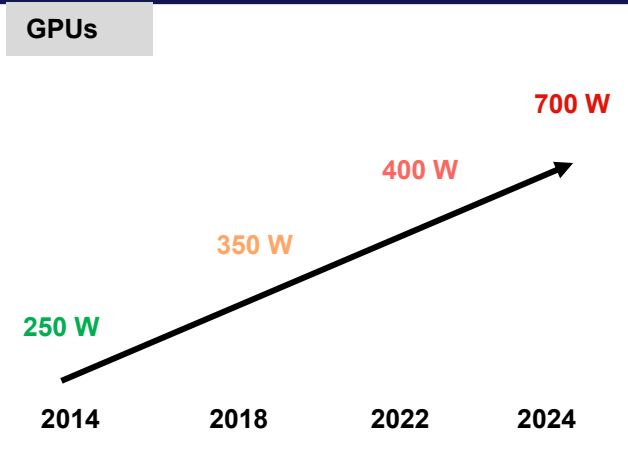
80% of the lifetime CO2

Manufacturer's Influence

What's is Consuming All the Power?

CPUs

- 500 W
- 300 W
- 240 W
- 145 W

2014 — 2018 — 2022 — 2024

GPUs

- 700 W
- 400 W
- 350 W
- 250 W

2014 — 2018 — 2022 — 2024

Memory

- 288W
- 144W
- 96W
- 72W

2014 — 2018 — 2022 — 2024

Network Adapters

- 60 W
- 40 W
- 25 W
- 15 W

2014 — 2018 — 2022 — 2024

Typical 2S 1U Server

Higher Power + Lower Tcase is driving fan power higher

33%
18%
22%
13%
14%
10%
8%
1200W
950W
7%
5%
800 W
4%
620 W
400 W

% Fan Power Stress

% Fan Power Normal

Server Power Trend

2014 — 2018 — 2020 — 2022 — 2024

AI Will Drive More Focus on The Complete Energy Picture

- The data center can't keep up
- Sustainability mandates a change

80-150kW+
(100kW Rack)

>60%

40-90kW/
Rack

>60%

30-50kW/
Rack

>10%
For the server's fans
10KW+
+
Cooling with a PUE of
1.4
40KW

50KW
to move air /
reject heat on a
100KW rack

The Good Ole Days….
(Grantley, Purely, V100)
2015~2021

Current Dense
(EGS, Genoa; H100)
2021~2024

What's Next
(Bx00, GraceHopper, Mi300X)
2025~2028
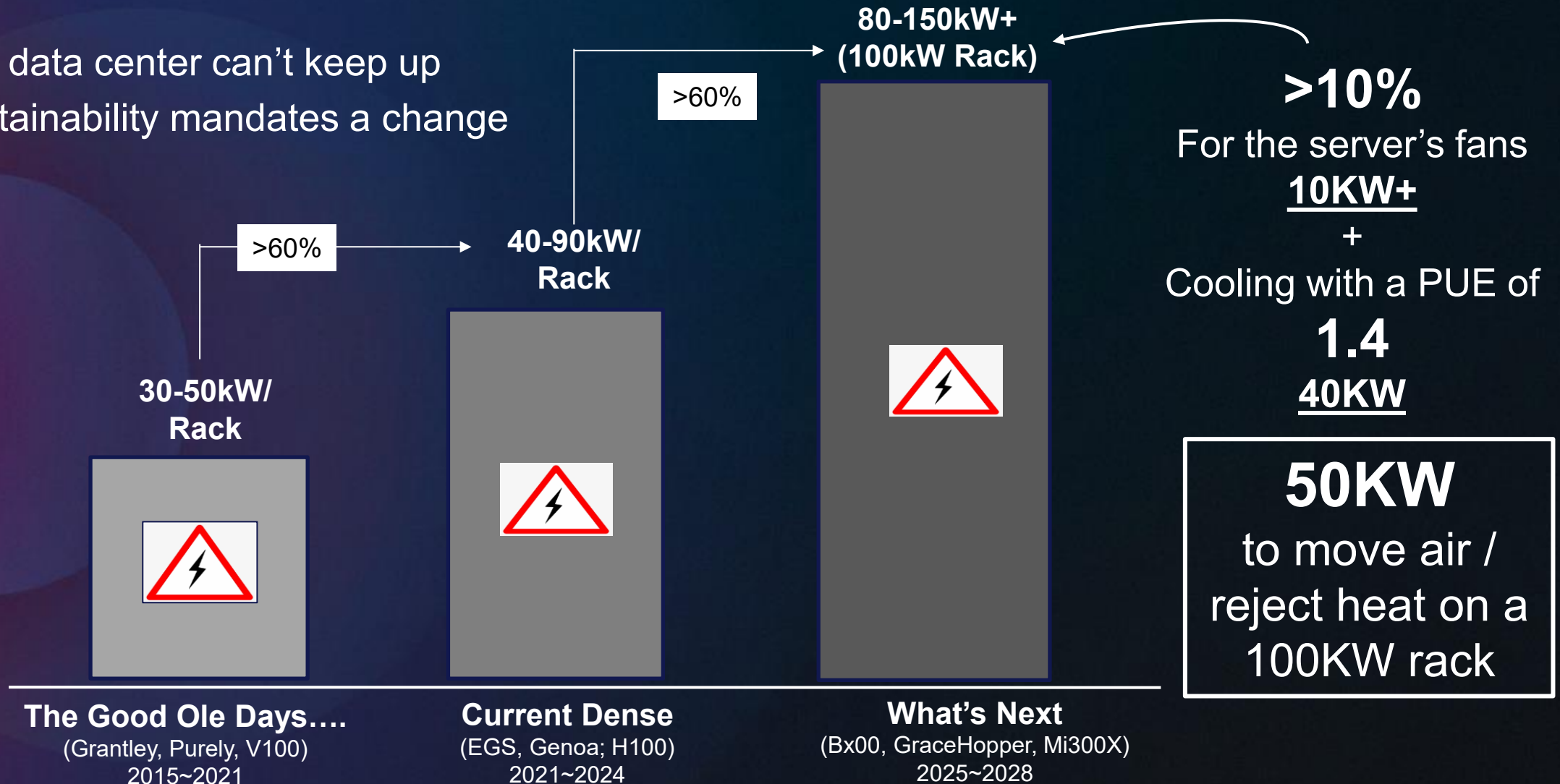
Lenovo Neptune® Technology Family

Neptune®

- The original, market-leading **100%** direct water cooling
- Warm water up to 45°C inlet
- #1 performance density platform
- ExaScale to EveryScale™ design
- Lowest PUE

Neptune® Air

- Liquid Assisted HS Cooling to **Air**
- Closed Loop with Radiator Resign
- Best air-cooling efficiency
- Full configurability; easiest to deploy
- Highest PUE; no DC heat rejection to liquid

Neptune® Core
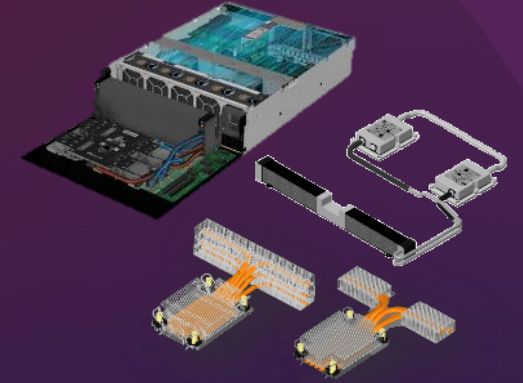
- Targeted at the **"hottest"** Components
- Open Loop Direct Water Cooling
- PUE ~1.2, 80% heat capture to liquid
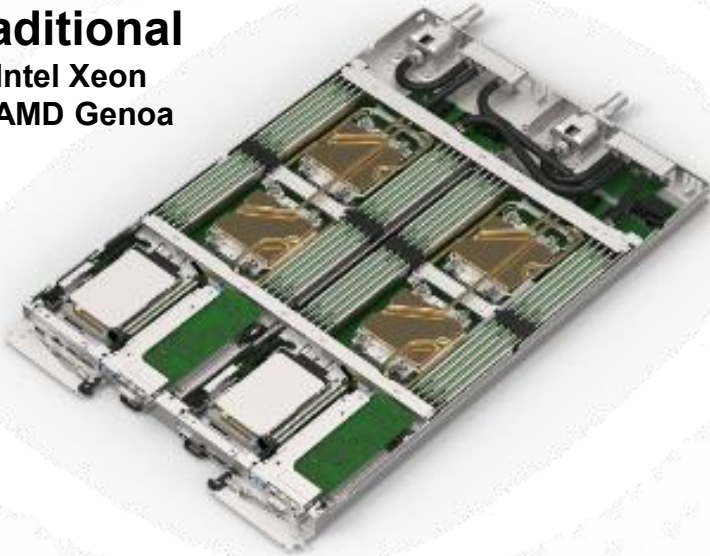- Maintain most configurability
- Low airflow requirement
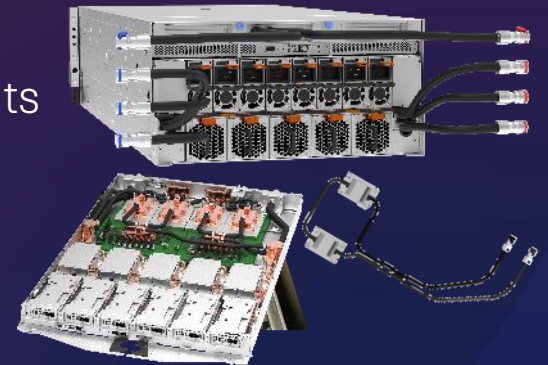
Lenovo Neptune· offers the right choice for your datacenter at every scale.

Lenovo Neptune™ Direct Liquid Cooling

Traditional
2 x Intel Xeon
2 x AMD Genoa
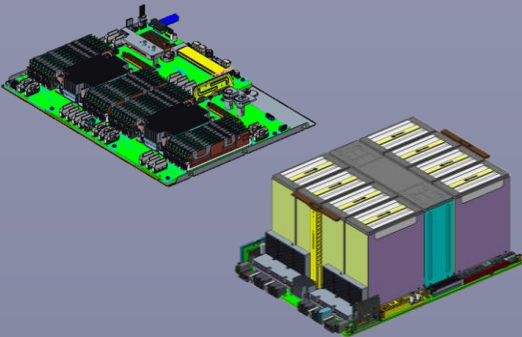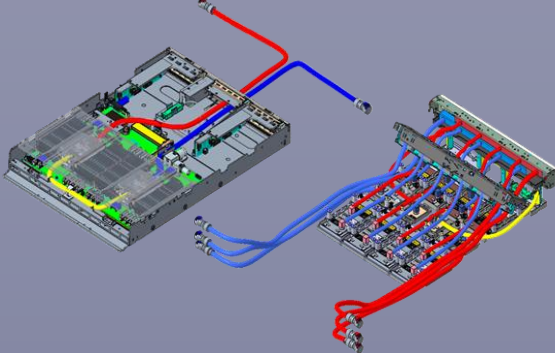
Accelerated
2 AMD + 4 NVIDIA H100
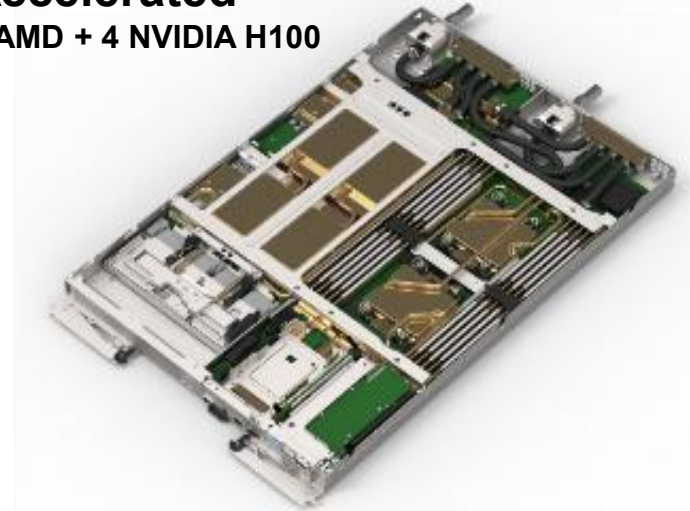
- Removal of fans reduces power inside the IT itself
- Liquid carries >97% of heat from the data center – reduces need for specialized AC / air handling
- Uses unchilled/warm water for cooling (>40C inlet) – chillerless operation, saves $ and water
- Supports PUEs <1.1 providng more kWs to the IT load not the overhead
- High temperature effluent allows for future recycling of heat energy
- Uses conditioned water – easy, safe, stable

B200 Air-Cooled vs Liquid-Cooled Baseboard

Models	Air-Cooled GPU + Air-Cooled CPU	Liquid-Cooled GPU + Liquid-Cooled CPU
Cooling Modules		
Max. B200 UBB* TDP	9335 W	9335 W
Max. CPU TDP	350W	350W
Liquid Capture	—	85%
Node Power	13654W	⬇ **11632W**
Fan Power Saving per node	—	⬇ **95% (≈1939W)**
DC 42U Rack 5 node Power Saving	—	**10.1KW**

Note：Draft data based on Ambient 22C；Liquid inlet 45C @ 12LPM per system

*UBB = Universal Base Board

Lenovo ThinkSystem N1380 Neptune

ThinkSystem SC750 V4

2X Intel Xeon 6th Generation "Granite Rapids"
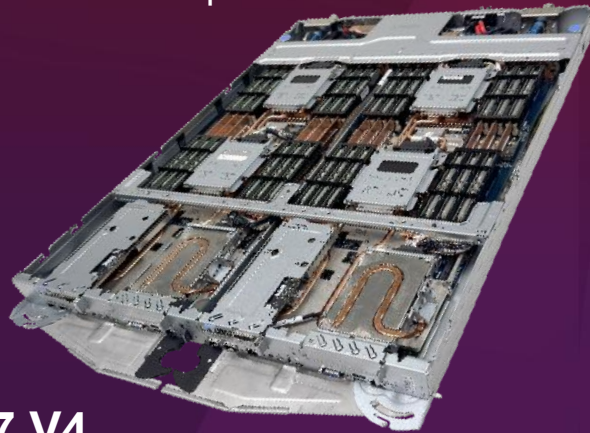
ThinkSystem SC777 V4

2X NVIDIA GH200
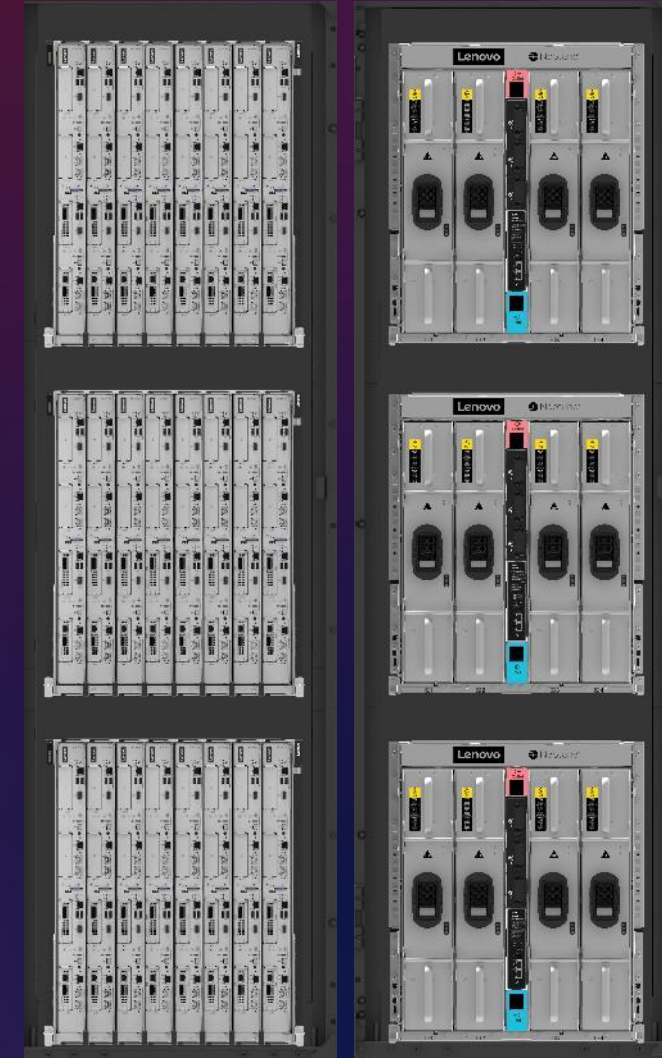
H200 Air-Cooled vs Liquid-Cooled Server

Models	Air-Cooled GPU + Air-Cooled CPU	Fan-less DLC Server Node
Cooling Modules		
Max. H200 SXM TDP	2800 W	2800 W
Max. CPU TDP	350W	350W
Liquid Capture	—	97%
Node Power	4940W	⬇ **4385W**
Fan Power Saving per node	—	⬇ **100%** (≈555W)
DC 42U Rack 36 node Power Saving	—	**20kW**

Note：Draft data based on Ambient 22C；Liquid inlet 45C @ 12LPM per system

When Power is the Constraint…

Power Grid

1MW

Data Centre

400kW

Cooling, UPS, Lights

400kW

Waste

600kW

Servers, Network, Storage

90kW

510kW

Productive Work
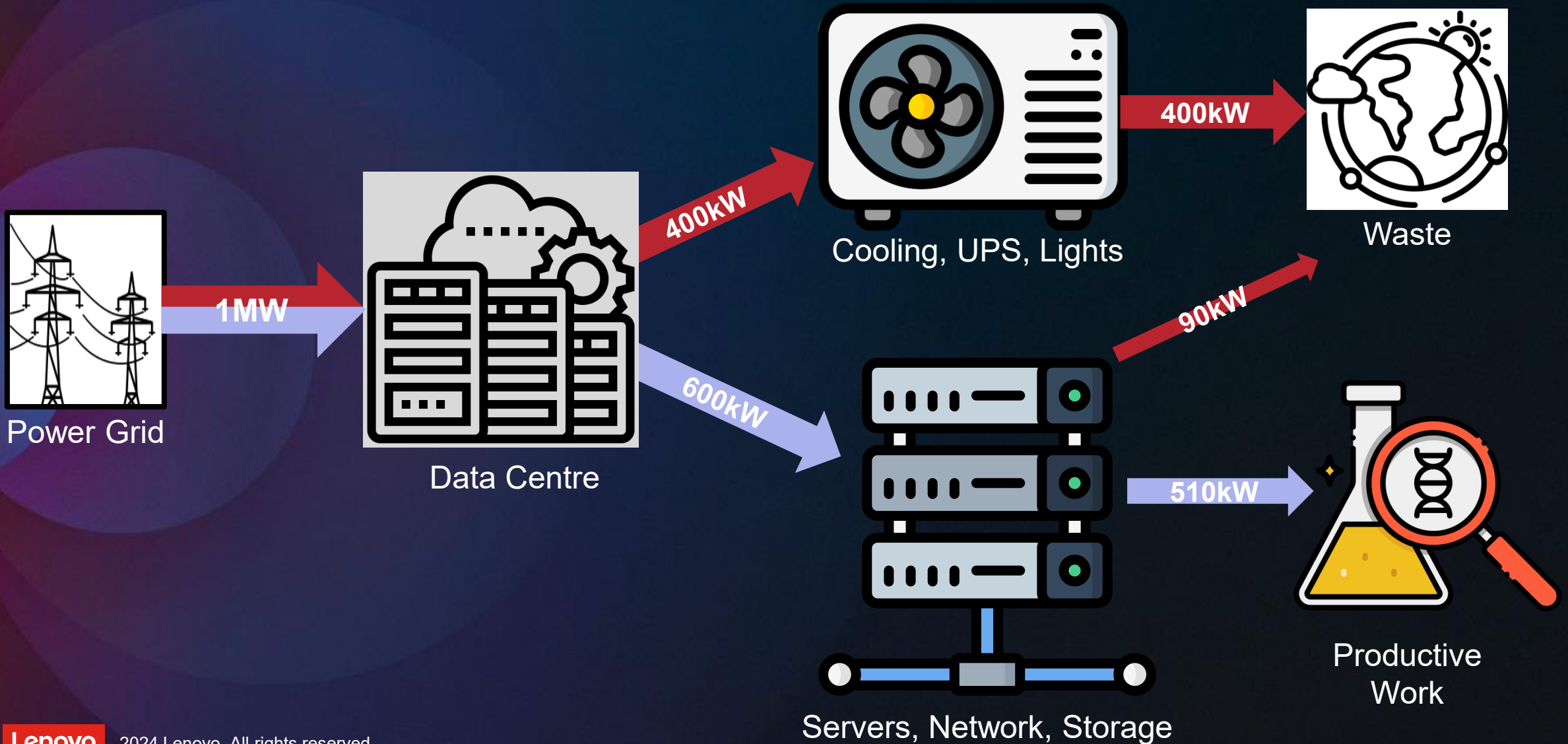
Liquid Cooling and Power

Power Grid — **1MW** → Data Centre

Data Centre — **100kW** → Cooling, UPS, Lights — **100kW** → Waste

Data Centre — **900kW** → Servers, Network, Storage — **900kW** → Productive Work

⚡ **76% more power available for IT.**

13
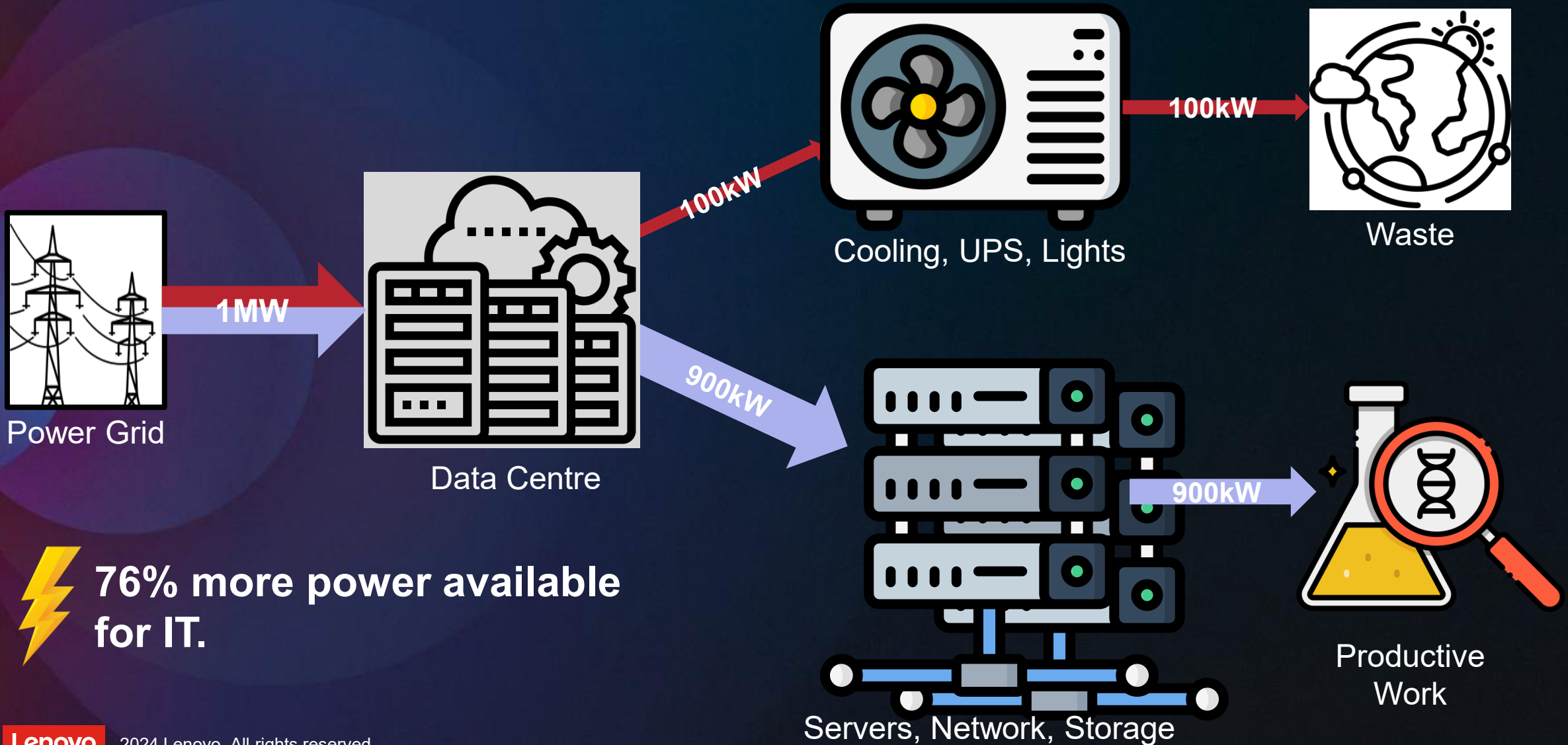
DC Operator's Influence

SYSTEMS AND DATACENTER COOLING
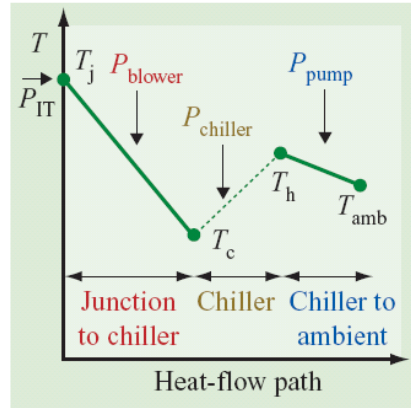
Evaporative Tower Fans

35°C

Chillers

Heat-flow path

T, T_j, P_{blower}, P_{pump}, P_{IT}, $P_{chiller}$, T_h, T_c, T_{amb}

Junction to chiller — Chiller — Chiller to ambient

Facility Pumps

Condenser

Heat Transfer Water to Refrigerant

Heat Transfer Chilled Air to Chilled Fluid

Heat Transfer Refrigerant to External coolant

6-8°C

25°C
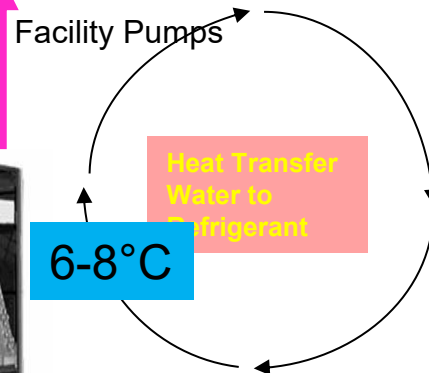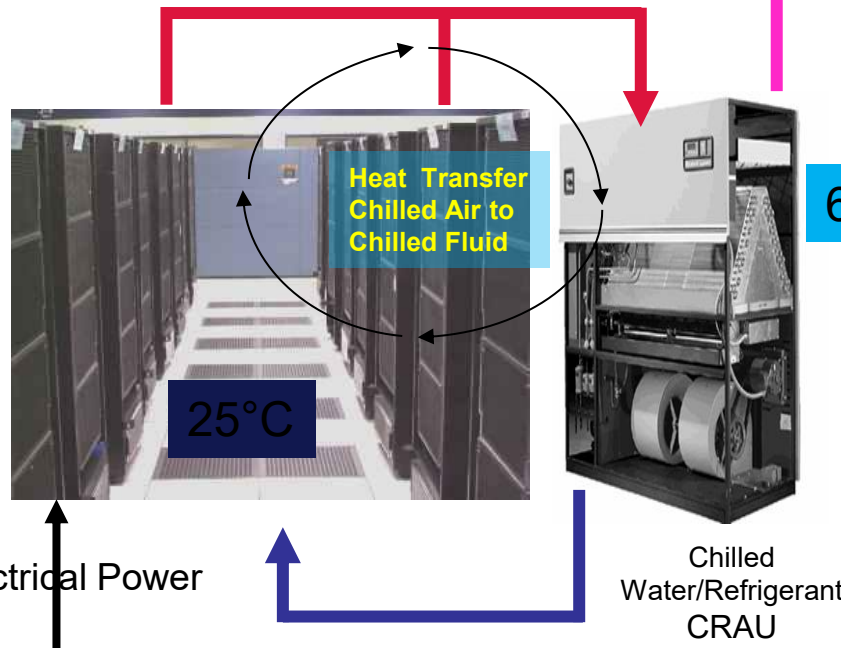
Energy Sources:
System Fan
CRAU Blower
Chiller Compressor
Chiller Pump
External Facility Pump
External Tower Fans

Electrical Power

Chilled Water/Refrigerant CRAU

Chiller 23%
Humidifier 3%
CRAC/CRAH 15%
Electrical Power IN
IT Equipment 47%
INDOOR DATA CENTER HEAT
Heat OUT
PDU 3%
UPS 6%
Lighting / aux devices 2%
Switchgear / generator 1%
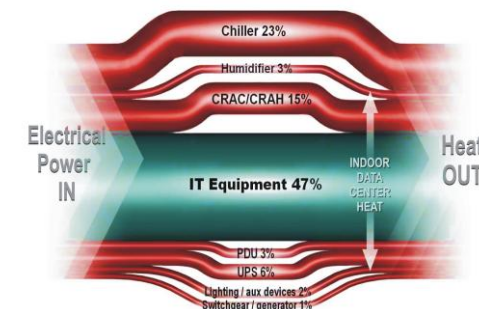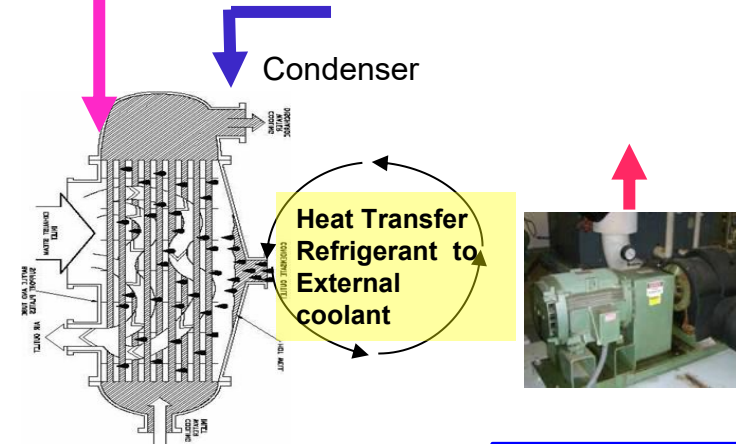
Power Usage Effectiveness

1.4 – 1.6

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

Typical Data Center Cooling

Direct-to-Chip Cooling in the Datacenter – Warm Water *

35°C

40°C

PUE~1.1

(optional)

Chiller

Cooling Tower

Datacom Equipment Center

External CDU

Rack

ITE

< 1% of Server Power

45°C

- Liquid-to-Liquid Heat Exchanger
- Multiple racks per CDU

- Servers employing open-loop liquid cooling

Condenser Water System (**CWS**)

Facilities Water System (**FWS**)

Technology Cooling System (**TCS**)
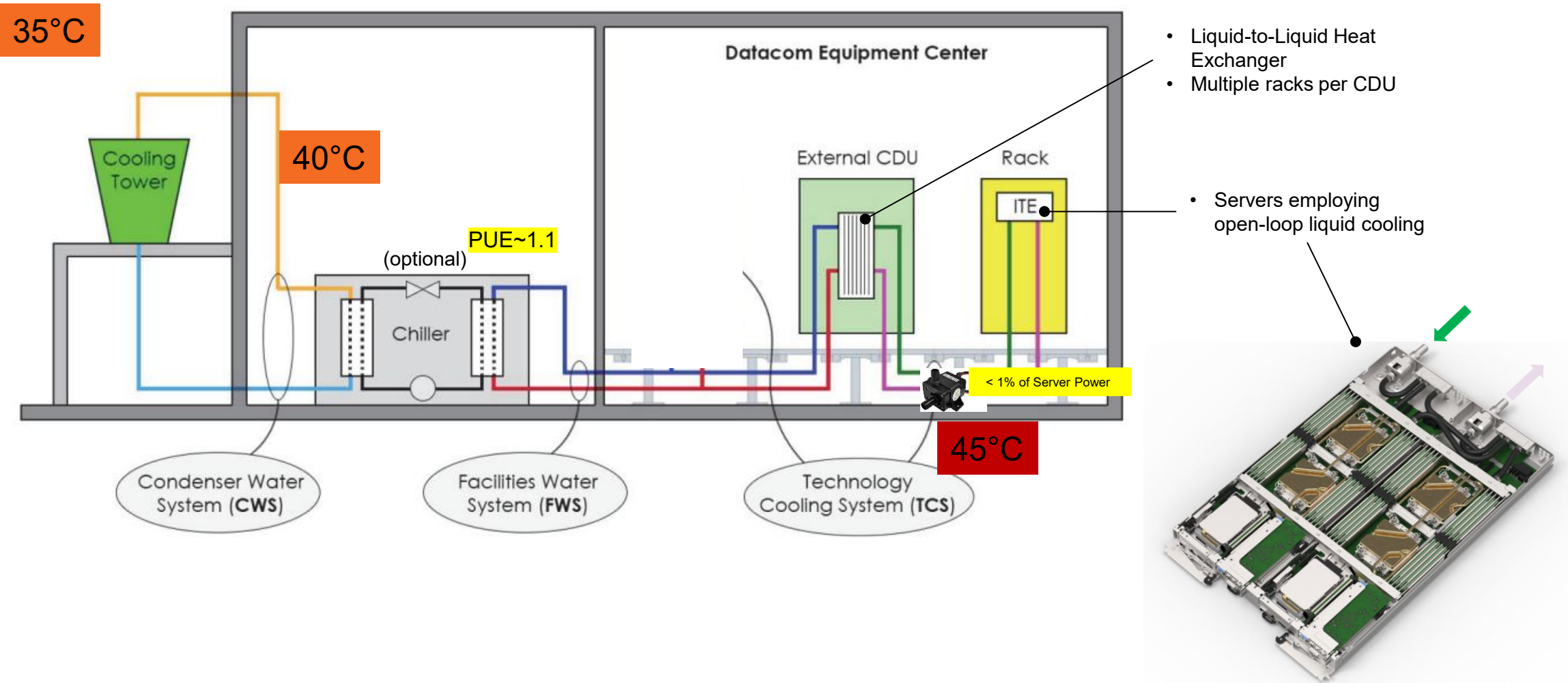
Synonyms:
- Direct Water Cooled (DWC)
- Direct-to-Chip Cooling (DTC)

Source: ASHRAE
https://www.ashrae.org/File%20Library/Technical%20Resources/Bookstore/WhitePaper_TC099-WaterCooledServers.pdf

16

Packaged CDU WCT Solution

Lenovo Services provided solution through vendor partner in region

Every solution differs due to customer resiliency requirements, existing infrastructure, location, and IT solution
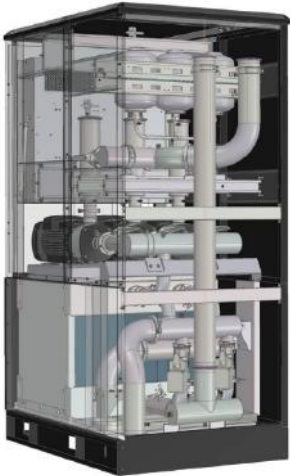
FS1350
(DN100 hygienic flange connections)

FS600
(DN65 hygienic flange connections)

COOLTERA

RM100
(DN40 hygienic flange connections)

WCT racks and/or RDHX

CS — Supply — Circuit setter flow control valves
Shutoff valves — CS
Return

Secondary side temperature feedback for controlled water temperature to specification

Pressure relief valve (based on maximum pressure application)

Quick-connect couplings

Facility Primary Water or Dry-cooler

Building chilled water

Supply

Return

Distribution manifolds

Flexible hoses, maximum length 15.24 meters (50 feet)
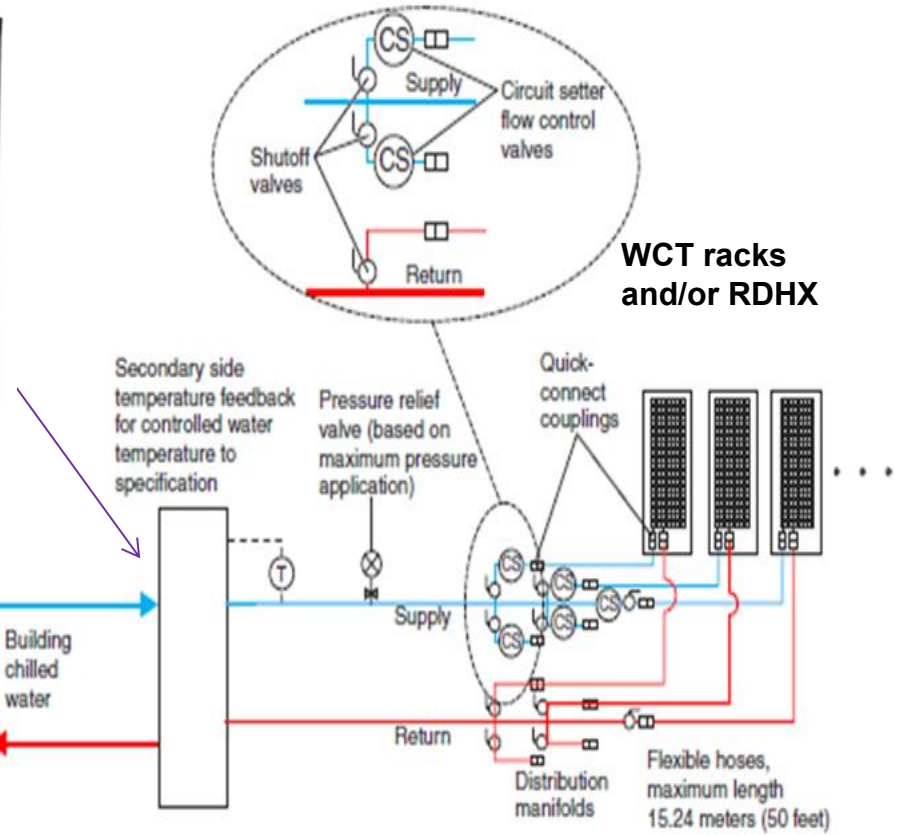
Supplier-built coolant distribution unit (CDU) suggested features:
- Temperature and flow metering (monitoring)
- Leak detection or water level sensing and shutdown
- Local and remote monitoring and control
- Access port for filling and water treatment

Lenovo 2024 Len

Beyond Carbon Neutral to Carbon Negative

GOAL: Deliver computing that is operationally carbon negative
Requires energy efficiency + green power + energy re-use

No/Low Emission Power:
Now more affordable than ever before

GREEN POWER IN

Direct Liquid Cooling:
5th Generation ThinkSystem SD650 and SD650-N

Neptune™

GREEN "HEAT" OUT

Liquid to Liquid Exchanger:
Transfer heat from IT to facilities

RECYCLED EMISSION FREE ENERGY

Facility Reuse:
Hot Water, Heat, Pool
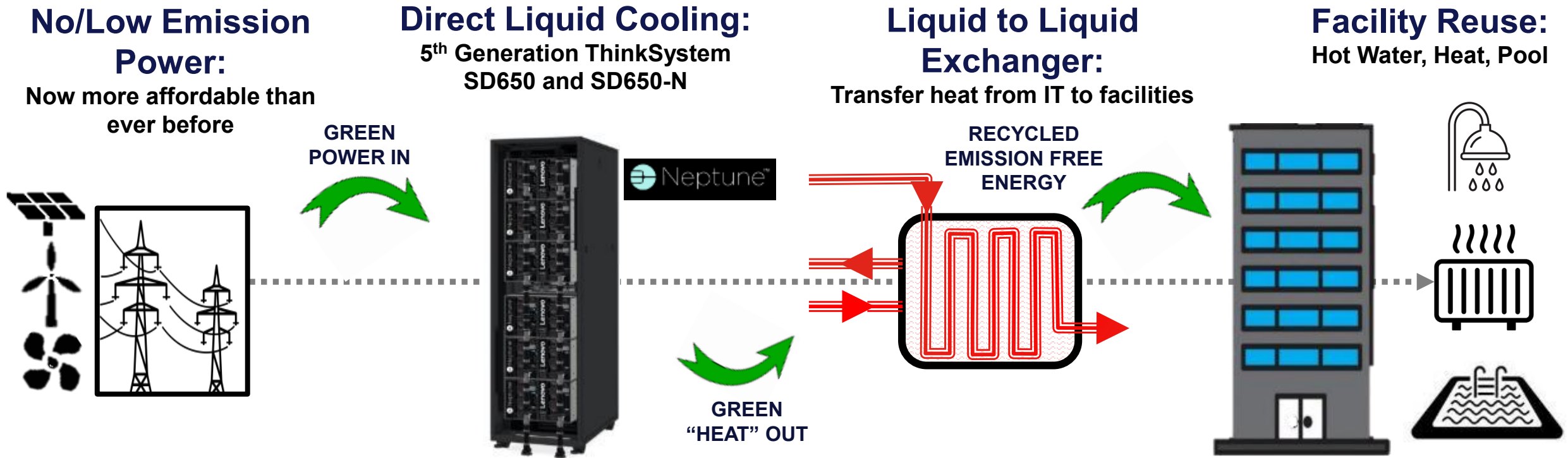
User's Influence

EAR SYSTEM SOFTWARE
FOR ENERGY EFFICIENT DATA CENTERS

Luigi Brochard
contact@eas4dc.com

September 2025

EAR: SYSTEM SOFTWARE FOR ENERGY EFFICIENCY

Energy models and policies for CPU/Memory/GPU frequency selection

CPU &GPU Optimization

Data reporting for accounting, billing, and system analytics

Accounting

Performance and Power metrics for system management and job analysis

Data Center Monitoring

Analysis and classification of job metrics for energy and power optimization

Analysis

Power control to guarantee data center operational limits

Powercap

www.eas4dc.com/

DYNAMIC ENERGY OPTIMIZATION

➢ ADAPT THE POWER CONSUMPTION TO THE APPLICATION ACTIVITY

CPU & GPU Optimization

Accounting

Analysis

Data Center Monitoring

Smart Powercap

www.eas4dc.com/

AI APPLICATIONS
RESULTS WITH PYTORCH AND TENSORFLOW ON NVIDIA H100

AI applications: GPU Energy savings and efficiency average

Legend: GPU Power Reductions, GPU Energy Savings, Performance per Watt Gain

5% Threshold: GPU Power Reductions 22.0, GPU Energy Savings 18.0, Performance per Watt Gain 20.5

10% Threshold: GPU Power Reductions 31.5, GPU Energy Savings 25.0, Performance per Watt Gain 31.4

Y-axis: % compared to the nominal

www.eas4dc.com/

23

AI APPLICATIONS
MORE RESULTS WITH PYTORCH AND TENSORFLOW

GPU Energy savings and efficiency on H100-SXM5

10% Threshold

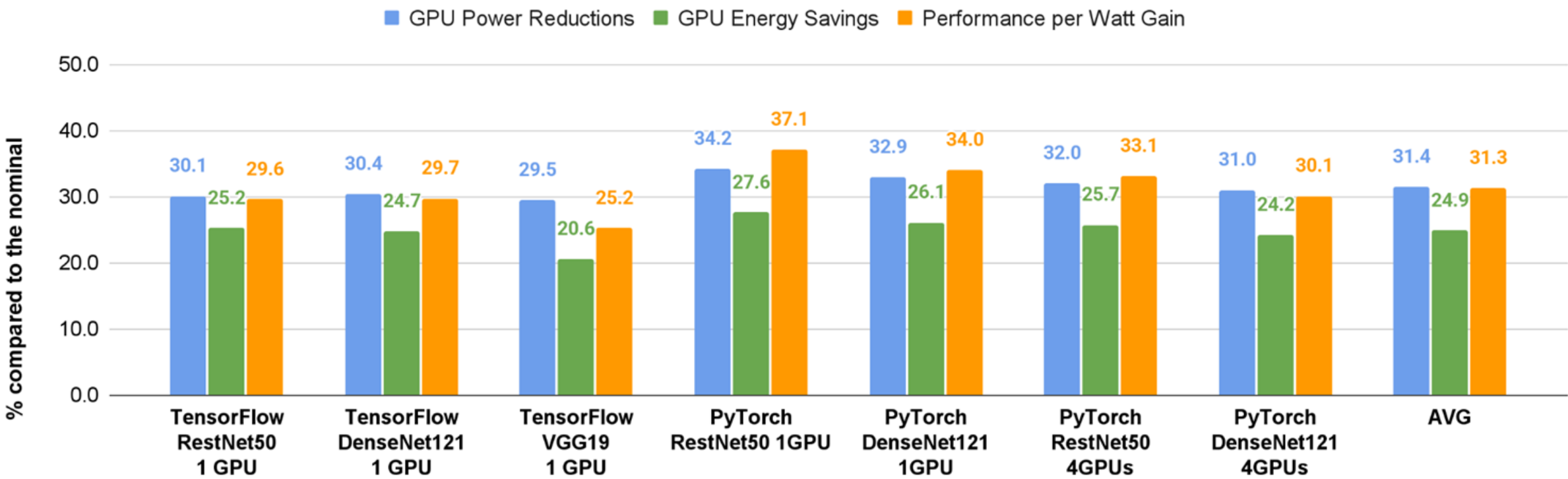Legend: ■ GPU Power Reductions ■ GPU Energy Savings ■ Performance per Watt Gain

% compared to the nominal

Category	GPU Power Reductions	GPU Energy Savings	Performance per Watt Gain
TensorFlow RestNet50 1 GPU	30.1	25.2	29.6
TensorFlow DenseNet121 1 GPU	30.4	24.7	29.7
TensorFlow VGG19 1 GPU	29.5	20.6	25.2
PyTorch RestNet50 1GPU	34.2	27.6	37.1
PyTorch DenseNet121 1GPU	32.9	26.1	34.0
PyTorch RestNet50 4GPUs	32.0	25.7	33.1
PyTorch DenseNet121 4GPUs	31.0	24.2	30.1
AVG	31.4	24.9	31.3

www.eas4dc.com/

thanks.

Smarter
technology
for all

Lenovo