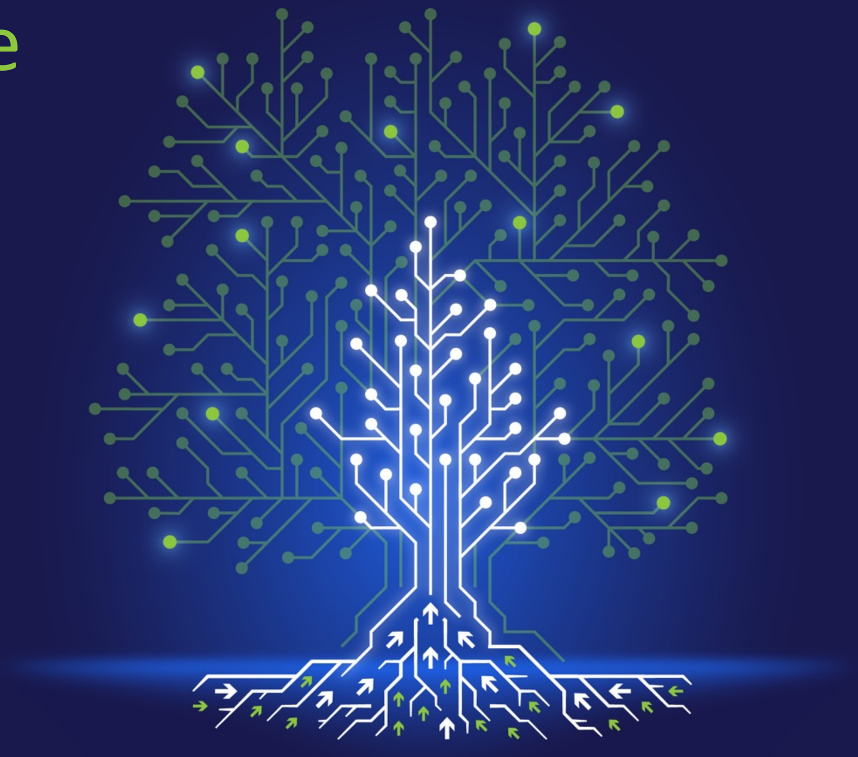
Scaling Down While Scaling Up

Design Choices That Increase Efficiency & Performance



Scaling Innovation Through Collaboration

OCP REGIONAL SUMMIT

24-25 APRIL 2024
LISBON, PORTUGAL

EMEA DEPLOYMENTS

DEPLOYMENTS

SUSTAINABILITY

Scaling Down While Scaling Up – Design Choices That Increase Efficiency & Performance

Erik Riedel, PhD, Chief Engineering Officer
Flax Computing

Hesham Ghoneim, PhD, Strategic Product Development
Sims Lifecycle Services

OPEN Compute Project SOLUTION PROVIDER®

OPEN PLATINUM®

Abstract

This talk presents a detailed design study of several OCP hyperscale systems vs previous server & rack designs, including trade-offs & choices that reduce complexity, reduce costs, increase performance, and increase scalability. We will also quantify that a clear side-effect of these choices is to reduce the carbon footprint of data center deployments & operations. These carbon savings arise from both operational & scope 3 benefits, and they continue to accrue the longer systems are productively utilized. The power of collaborative, “out of the box” design approaches that have crossed traditional industry silos – across hardware, software, & operations – have enabled innovative designs that have proven their value and influence across the industry. The fact that these designs were pursued with simplicity and efficiency in mind mean that they also inherently consume less resources in both energy, materials, and operational effort paying dividends in the short-run and long-run.

Expected Learnings

Talk attendees will learn about the details of several OCP server, storage & GPU designs that rethink the traditional data center server to provide clear benefits in reduced complexity, reduced costs, increased performance, & increased efficiency at scale. By reviewing a small subset of open hardware designs from the 10 year history of the Open Compute Project, we will outline the clear engineering trade-offs that were made, and the resulting gains in efficiency & scalable deployment. Attendees will also learn how these optimizations have direct benefits to the reduction of carbon footprints – both energy & materials consumption – when combined with modern data center infrastructure, deployment, & software architectures. The combination of creative hardware designs, software architectures, & operational approaches developed in wide industry collaborations with traditionally siloed teams and companies operating together have made these individual evolutions into a true revolution.

Social 280

Scaling Down While Scaling Up – learn about OCP designs that reduce complexity, costs & carbon footprints, while increasing performance & efficiency. By innovating across industry silos, these collaborations have made a series of design evolutions into a true revolution.

Outline

- Intro – carbon footprint of computing
- Design Footprints – compare OCP to traditional server designs
- Materials Footprints – material + carbon impacts
- Materials Reuse / Recycle – extend life, reduce carbon impacts
- Server Density – can we improve further
- Workload / Carbon – example: memory density
- Conclusions + Call To Action


natural
resources
carbon
footprint
demand
growth

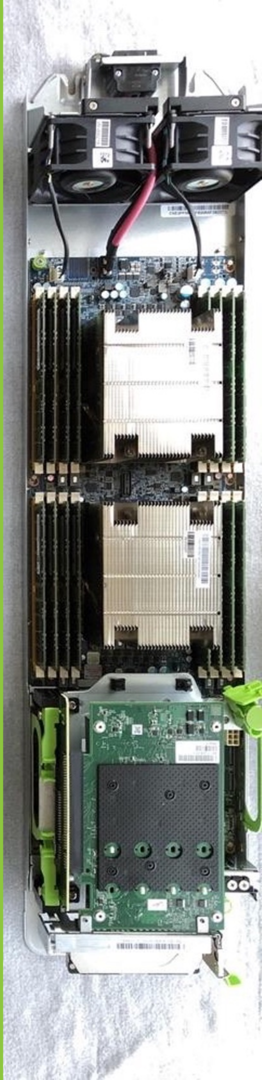
design footprints

2 servers / 2U
- dual CPUs
- 24x memory sockets
- two 1U heatsinks
- twelve 1U fans
- eight SATA drives 2.5”
- two PSUs

(PowerEdge R630)

(OCP Leopard)

3 servers / 2 OU
- dual CPUs
- 16x memory sockets
- two 2U heatsinks
- two 2U fans
- six NVMe drives M.2
- shared PSUs



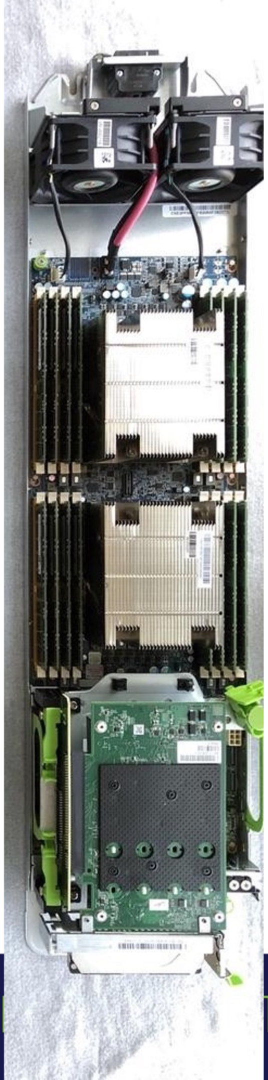
Dell R630

- 2 servers / 2U
    - dual CPUs
    - 24x memory sockets
    - two 1U heatsinks
    - twelve 1U fans
    - eight SATA drives 2.5”
    - two PSUs
- 40 servers / rack

	server	rack
cpus	2	80
dimms	24	960
heatsinks	2	80
fans	12	480
2.5” SATA	8	640
psus	2	80

OCP Leopard/v2

- 3 servers / 2OU
  - dual CPUs
  - 16x memory sockets
  - two 2U heatsinks
  - two 2U fans
  - six NVMe drives M.2
  - shared PSUs
- 48 servers / rack



	server	rack
cpus	2	96
dimms	16	768
heatsinks	2	96
fans	2	96
M.2 NVMe	6	288
psus	--	6

materials footprints

	server	weight (g)	rack	total (kg)	
cpus	2	54	80	4.3	
dimms	24	10	960	9.6	
heatsinks	2	165	80	13.2	
motherboard	1	875	40	35.0	
fans	12	20	480	9.6	
2.5” SATA	8	90	320	28.8	
2.5” drive caddy	8	54	320	17.3	
NIC 4x 10G	1	320	40	12.8	
PCI riser (NIC)	1	90	40	3.6	
PCI riser (unused)	1	130	40	5.2	
psus	2	565	80	45.2	
chassis	1	17.6 kg	40	704.0	
pdu	--	35 kg	1	35.0	
power cables	2	120	80	9.6	**TOTAL**
rack	--	145 kg	1	145.0	**1078 kg**

	server	weight	rack	total	
cpus	2	54	96	5.2	
dimms	16	10	768	7.7	
heatsinks	2	185	96	17.8	
motherboard	1	325	48	15.6	
fans	2	30	96	2.9	
M.2 NVMe	6	75	288	21.6	
AVA board	2	307	96	29.5	
NIC 2x 25G mezz	1	154	48	7.4	
PCI riser (x16 + x8)	1	92	48	4.4	
--	--	0	48	0.0	
psus	--	5.5 kg	6	33.0	
chassis	1	7.5 kg	48	360.0	
power shelf	--	25 kg	1	25.0	
busbar	--	15 kg	1	15.0	**TOTAL**
rack	--	175 kg	1	175.0	**720 kg**

	server	weight (g)	rack	total (kg)	
cpus	2	54	80	4.3	1,120
dimms	24	10	960	9.6	cores
heatsinks	2	165	80	13.2	
motherboard	1	875	40	35.0	
fans	12	20	480	9.6	
2.5” SATA	8	90	320	28.8	just
2.5” drive caddy	8	54	320	17.3	over 1
NIC 4x 10G	1	320	40	12.8	core
PCI riser (NIC)	1	90	40	3.6	per kg
PCI riser (unused)	1	130	40	5.2	
psus	2	565	80	45.2	
chassis	1	17.6 kg	40	704.0	
pdu	--	35 kg	1	35.0	
power cables	2	120	80	9.6	TOTAL
rack	--	145 kg	1	145.0	1078 kg

	server	weight (g)	rack	total (kg)	
cpus	2	54	80	4.3	0.4%
dimms	24	10	960	9.6	0.9%
heatsinks	2	165	80	13.2	
motherboard	1	875	40	35.0	
fans	12	20	480	9.6	
2.5” SATA	8	90	320	28.8	
2.5” drive caddy	8	54	320	17.3	
NIC 4x 10G	1	320	40	12.8	
PCI riser (NIC)	1	90	40	3.6	
PCI riser (unused)	1	130	40	5.2	
psus	2	565	80	45.2	
chassis	1	17.6 kg	40	704.0	
pdu	--	35 kg	1	35.0	
power cables	2	120	80	9.6	TOTAL
rack	--	145 kg	1	145.0	1078 kg

only 1.3% of rack is the “computing” elements

	server	weight	rack	total	
cpus	2	54	96	5.2	1,344
dimms	16	10	768	7.7	cores
heatsinks	2	185	96	17.8	
motherboard	1	325	48	15.6	
fans	2	30	96	2.9	
M.2 NVMe	6	75	288	21.6	almost
AVA board	2	307	96	29.5	2 cores
NIC 2x 25G mezz	1	154	48	7.4	per kg
PCI riser (x16 + x8)	1	92	48	4.4	
--	--	0	48	0.0	
psus	--	5.5 kg	6	33.0	
chassis	1	7.5 kg	48	360.0	
power shelf	--	25 kg	1	25.0	
busbar	--	15 kg	1	15.0	TOTAL
rack	--	175 kg	1	175.0	720 kg

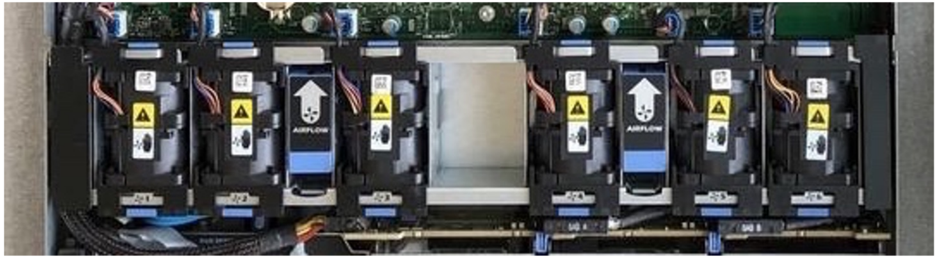
	server	weight	rack	total	
cpus	2	54	96	5.2	0.7%
dimms	16	10	768	7.7	1.1%
heatsinks	2	185	96	17.8	
motherboard	1	325	48	15.6	
fans	2	30	96	2.9	
M.2 NVMe	6	75	288	21.6	
AVA board	2	307	96	29.5	
NIC 2x 25G mezz	1	154	48	7.4	
PCI riser (x16 + x8)	1	92	48	4.4	
--	--	0	48	0.0	
psus	--	5.5 kg	6	33.0	
chassis	1	7.5 kg	48	360.0	
power shelf	--	25 kg	1	25.0	
busbar	--	15 kg	1	15.0	TOTAL
rack	--	175 kg	1	175.0	720 kg

only 1.8% of rack is the "computing" elements

	server	weight	rack	total	
cpus	2	54	96	5.2	0.7%
dimms	16	10	768	7.7	1.1%
heatsinks	2	185	96	17.8	only 1.8% of rack is the “computing” elements
motherboard	1	325	48	15.6	
fans	2	30	96	2.9	
M.2 NVMe	6	75	288	21.6	3.0%
AVA board	2	307	96	29.5	to 4.8% with storage elements included
NIC 2x 25G mezz	1	154	48	7.4	
PCI riser (x16 + x8)	1	92	48	4.4	
--	--	0	48	0.0	
psus	--	5.5 kg	6	33.0	
chassis	1	7.5 kg	48	360.0	
power shelf	--	25 kg	1	25.0	
busbar	--	15 kg	1	15.0	TOTAL
rack	--	175 kg	1	175.0	720 kg

Example – Fans

- twelve 1U fans



- two 2U fans

Example – Fans

- twelve 1U fans

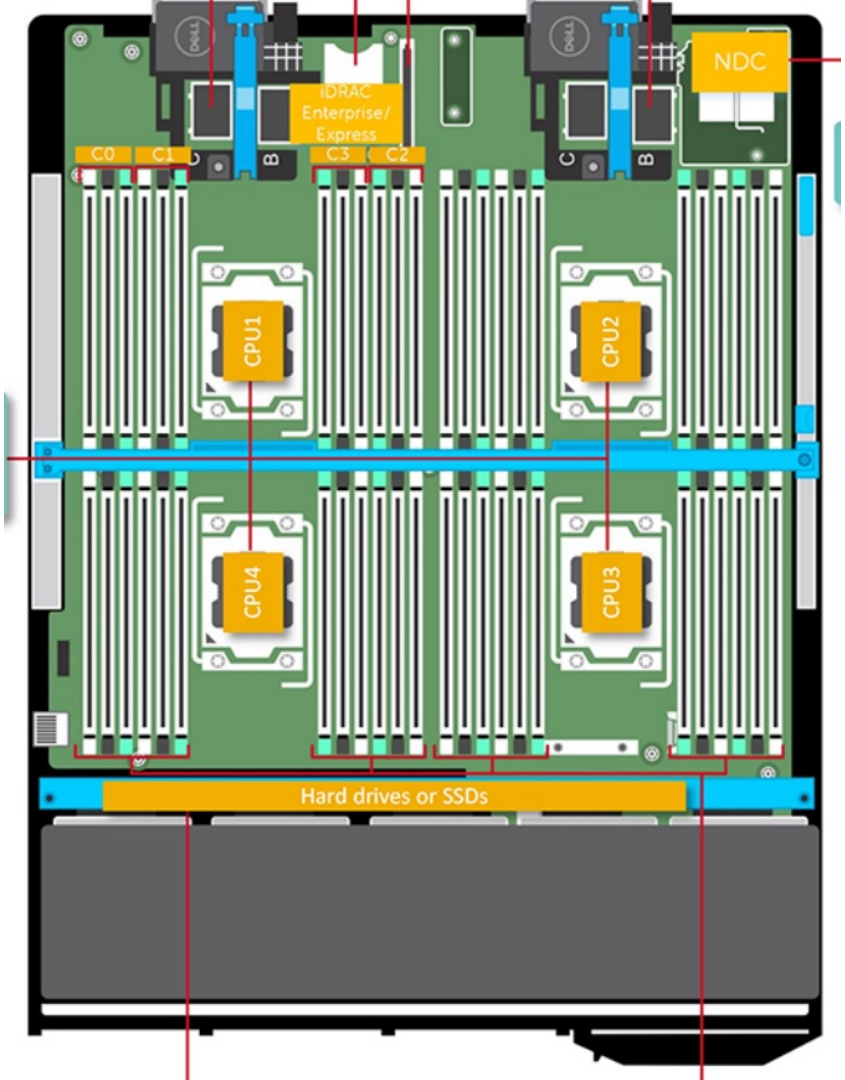
24 / 2U

- two 2U fans

6 / 2OU

…but wait, there's more…

(even denser server options)



2 servers / 2U
- quad CPUs
- 48x memory sockets
- four 1U heatsinks
- twenty 1U fans
- eight SAS/SATA drives 2.5”
- two PSUs

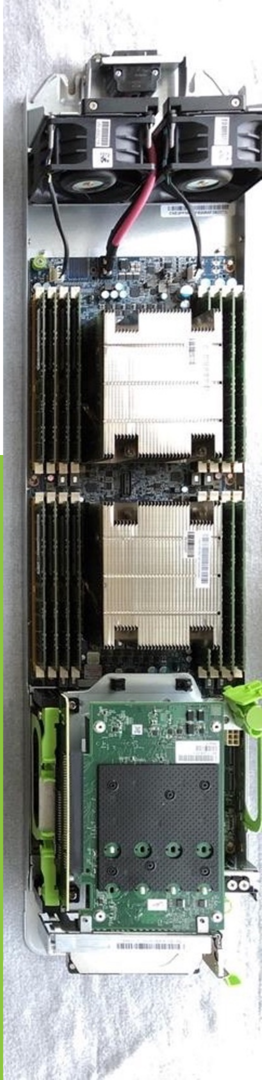
8 sockets,
96 dimms

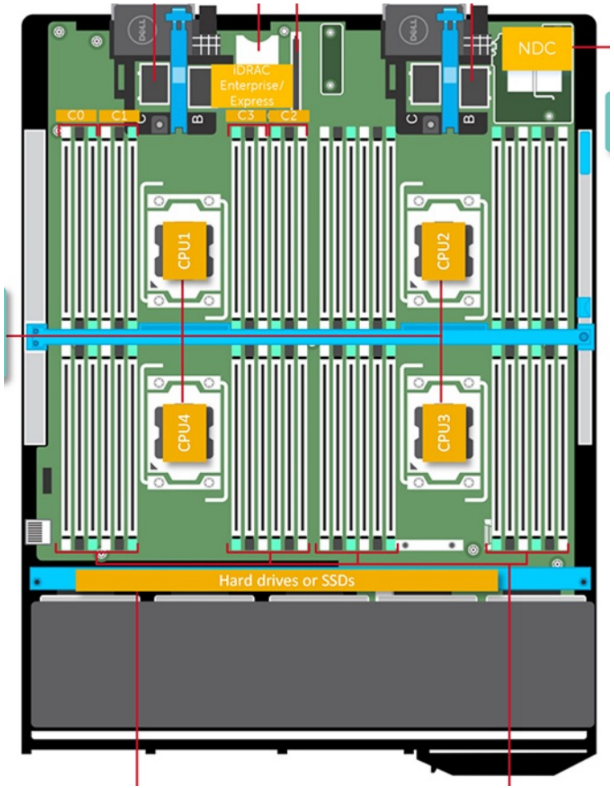
(PowerEdge FC830)

(OCP Leopard)

3 servers / 2 OU
- dual CPUs
- 16x memory sockets
- two 2U heatsinks
- two 2U fans
- six NVMe drives M.2
- shared PSUs

6 sockets,
48 dimms





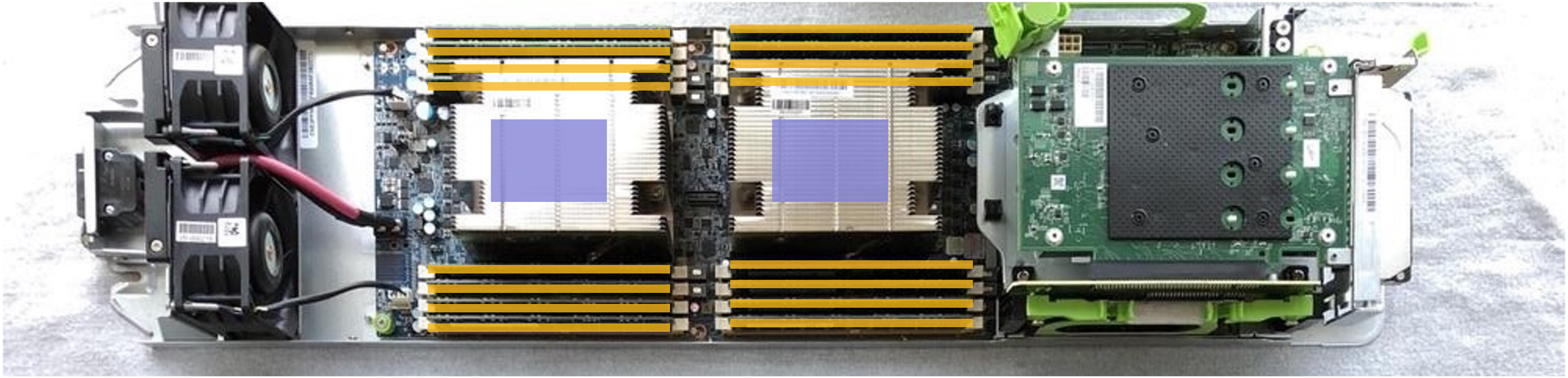
	server	weight (g)	rack	total (kg)	
cpus	4	54	160	8.6	3,520
dimms	48	10	1920	19.2	cores
heatsinks	4	165	160	26.4	
motherboard	1	875	40	35.0	
fans	12	20	480	9.6	
2.5” SATA	8	90	320	28.8	over 3
2.5” drive caddy	8	54	320	17.3	cores
NIC 4x 10G	1	185	40	7.4	per kg
PCI riser (NIC)	1	65	40	2.6	
PCI riser (unused)	3	65	120	7.8	
psus	1	565	40	22.6	
chassis	1	18.4 kg	40	736.0	
pdu	--	35 kg	1	35.0	
power cables	1	120	40	4.8	TOTAL
rack	--	145 kg	1	145.0	1106 kg

materials reuse

	server	weight	rack	total	
cpus	2	54	96	5.2	
dimms	16	10	768	7.7	
heatsinks	2	185	96	17.8	
motherboard	1	325	48	15.6	
fans	2	30	96	2.9	
M.2 NVMe	6	75	288	21.6	
AVA board	2	307	96	29.5	
NIC 2x 25G mezz	1	154	48	7.4	
PCI riser (x16 + x8)	1	92	48	4.4	
psus	--	5.5 kg	6	33.0	
chassis	1	7.5 kg	48	360.0	
power shelf	--	25 kg	1	25.0	
busbar	--	15 kg	1	15.0	TOTAL
rack	--	175 kg	1	175.0	720 kg

REUSE

(MEMORY COMPONENTS)



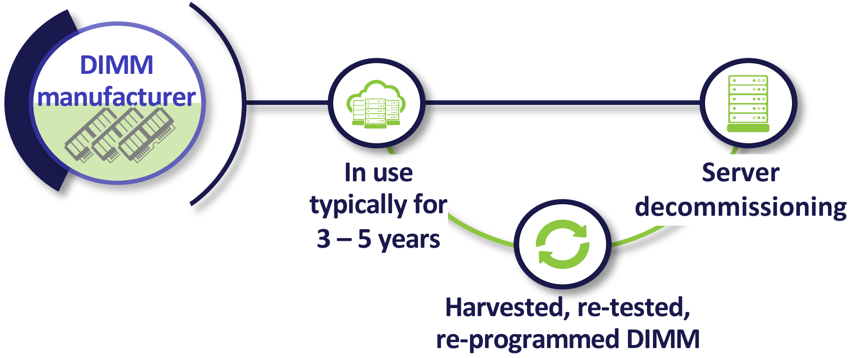
	server	weight	rack	total	
cpus	2	54	96	5.2	1,344 cores
dimms	16	10	768	7.7	12 TB mem

socketed cpu – fits any LGA2011 Haswell/Broadwell server or workstation socket

memory modules fit any DDR4 server ECC socket

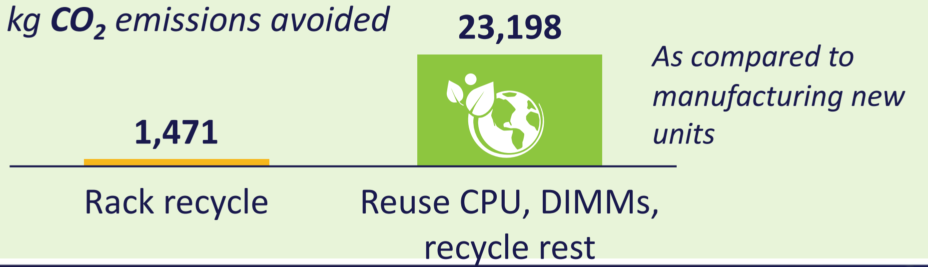
So why does it matter?

1. Internal / external reuse

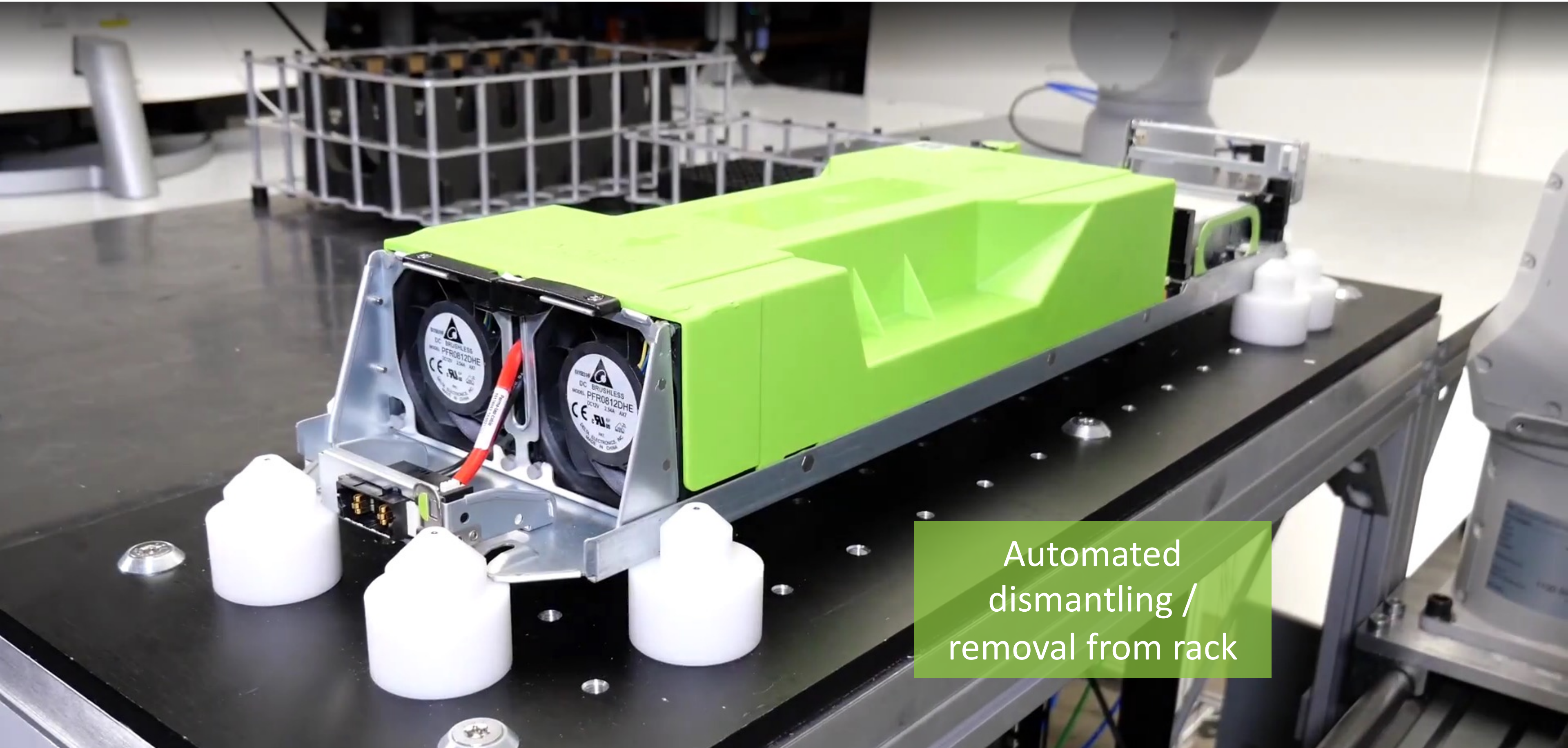


2. Economics – market value

3. Avoided emissions per rack

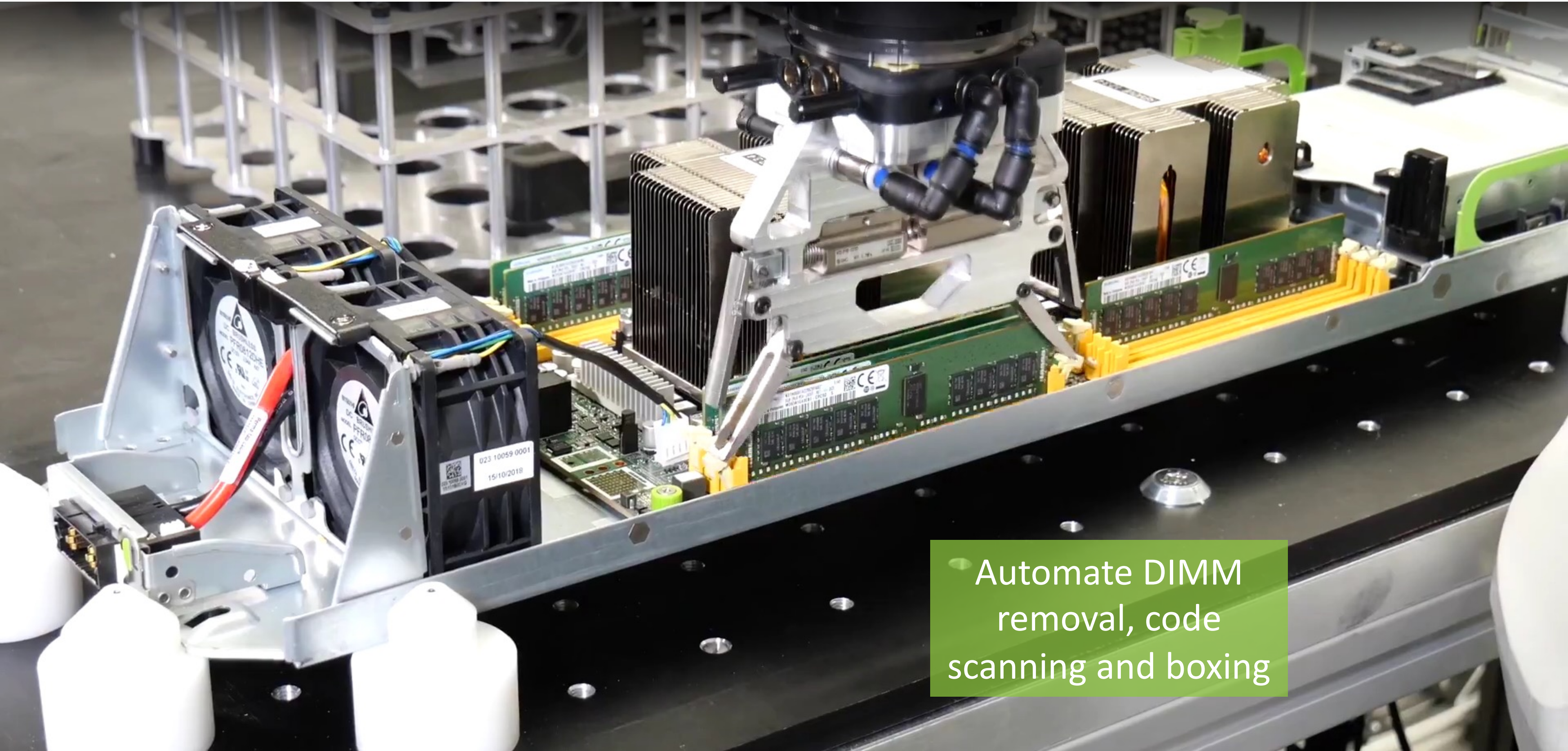


Automation is key for reuse of components



Automated dismantling / removal from rack

Automation is key for reuse of components



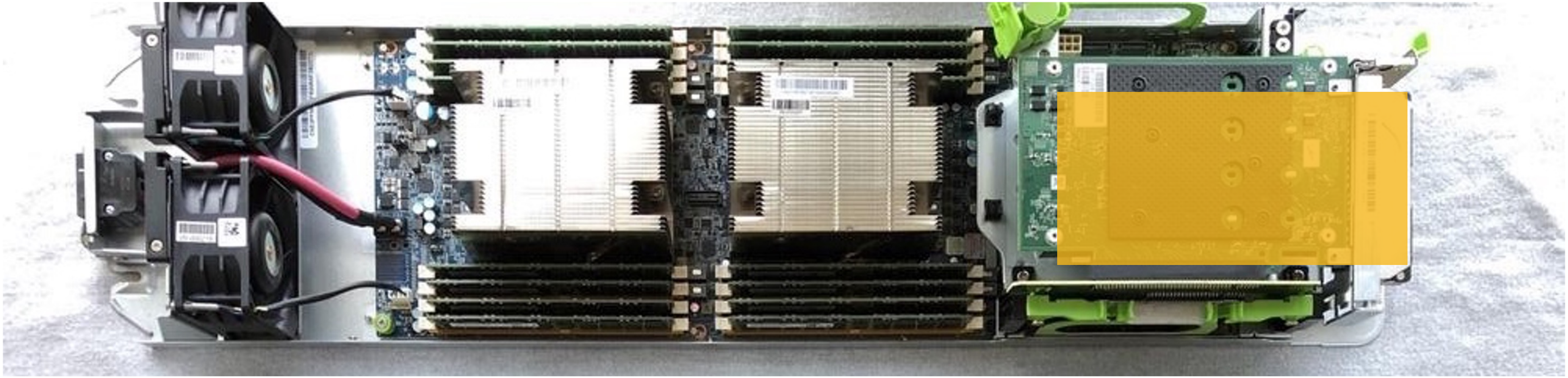
Automate DIMM removal, code scanning and boxing

Automation is key for reuse of components

Automated heat sink and CPU removal

~~REUSE~~ SHRED

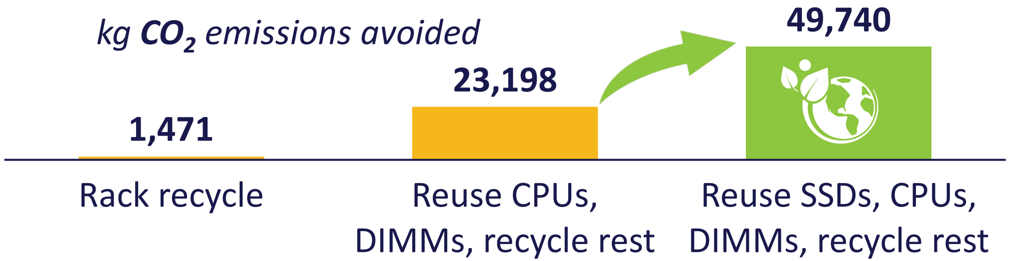
(STORAGE COMPONENTS)



	server	weight	rack	total
M.2 NVMe	6	75	288	21.6

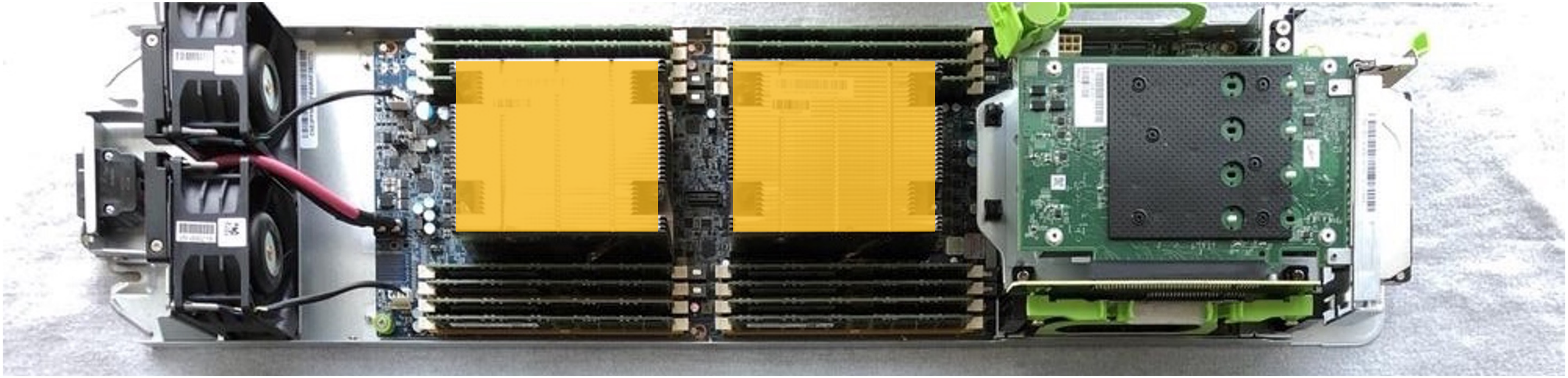
flash drives fit any m.2/nvme server, workstation, or desktop

but…usually need to be shredded.



RECYCLE

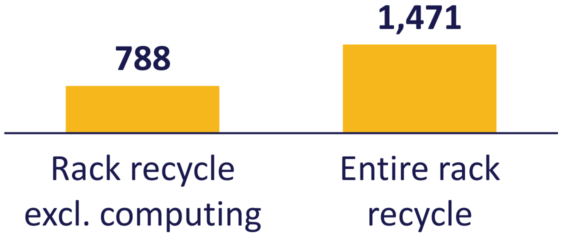
(METAL COMPONENTS)



	server	weight	rack	total
heatsinks	2	185	96	17.8
chassis	1	7.5 kg	48	360.0
busbar	--	15 kg	1	15.0
rack	--	175 kg	1	175.0

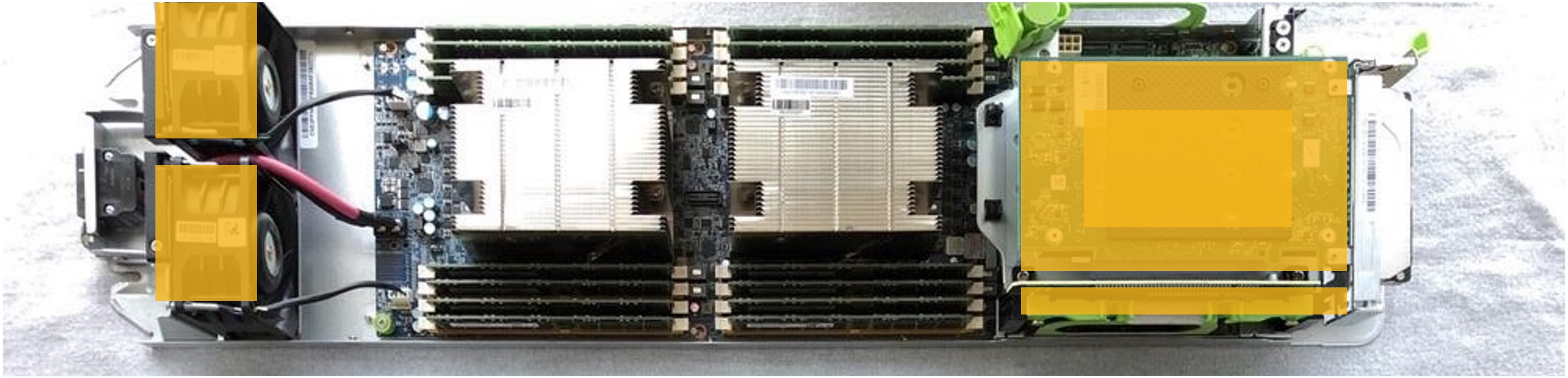
➔ *metal extracted for direct recycling*

kg CO_2 emissions avoided



~~RECYCLE~~ REUSE

(SYSTEMS COMPONENTS)



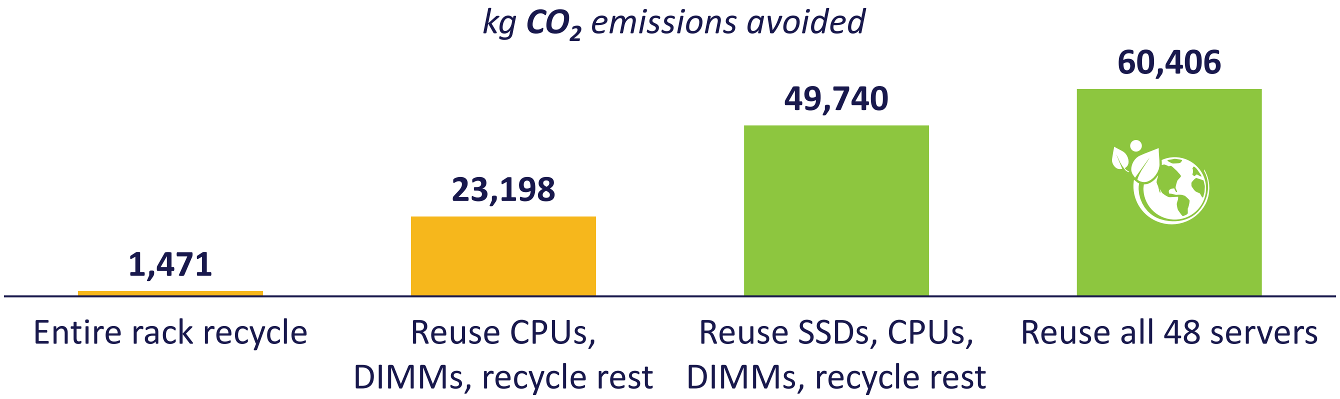
	server	weight	rack	total
motherboard	1	325	48	15.6
fans	2	30	96	2.9
AVA board	2	307	96	29.5
NIC 2x 25G mezz	1	154	48	7.4
PCI riser (x16 + x8)	1	92	48	4.4

➡ *difficult materials extraction*

extract precious metals such as Au, Ag, Pd, and Cu from PCBs

…can we do better ?

(more reuse, repurpose, redeploy)

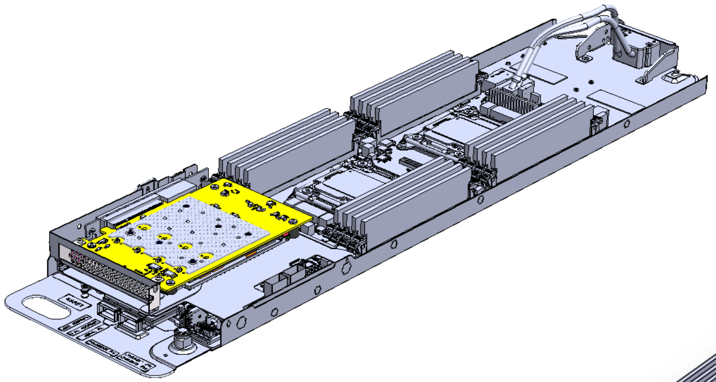


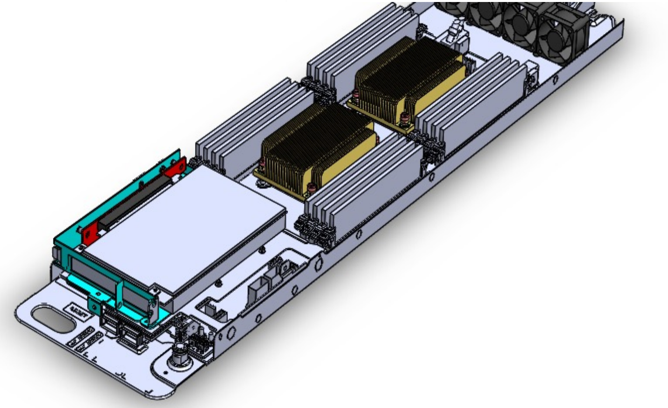
…additional details…

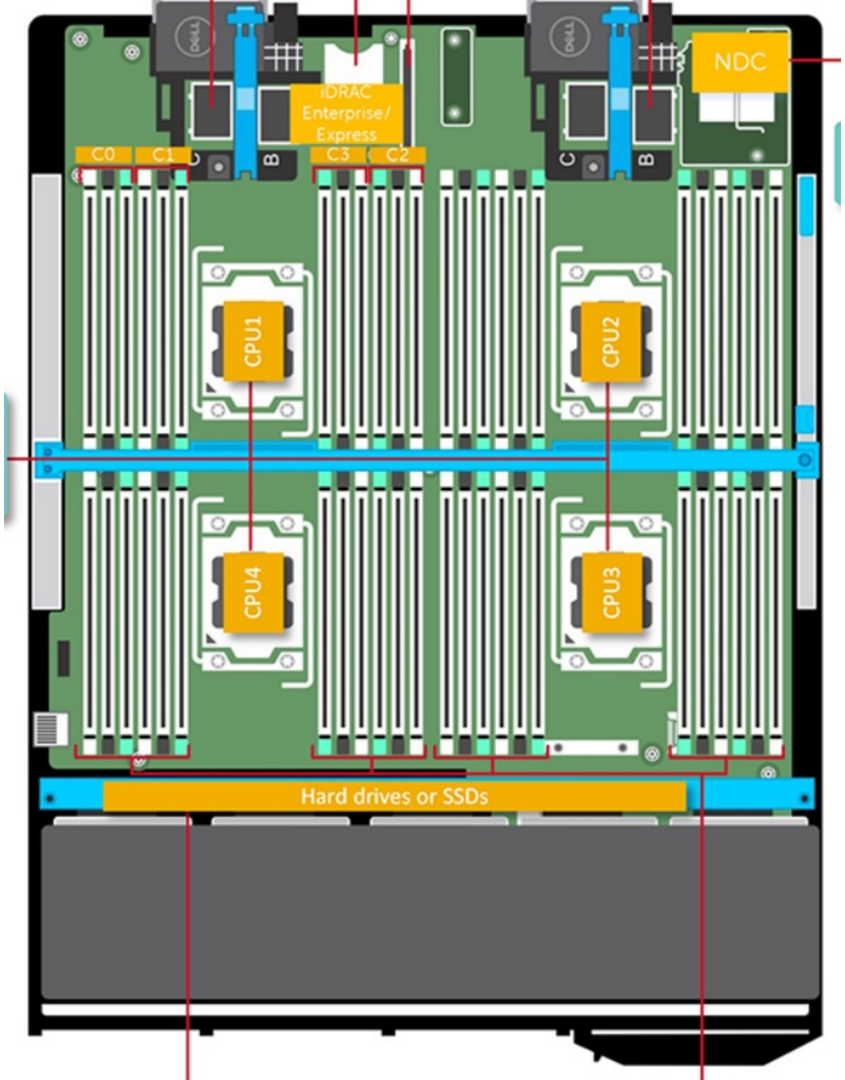
(higher density design optimization)

1OU Tioga Pass

- OCP Tioga Pass
- in 1 OU form factor
- low-profile heat sinks
- low-profile 4x NVMe M.2
- dual SkyLake-SP, CascadeLake-SP
- 768GB memory (12x 64GB)
- 15 TB NVMe (4x 3.84TB)
- dual 25G net + host mgmt
- optional dual 100G net







8 sockets,
96 dimms

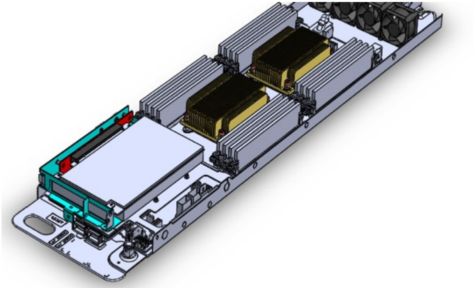
2 servers / 2U
- quad CPUs
- 48x memory sockets
- four 1U heatsinks
- twenty 1U fans
- eight SAS/SATA drives 2.5”
- two PSUs

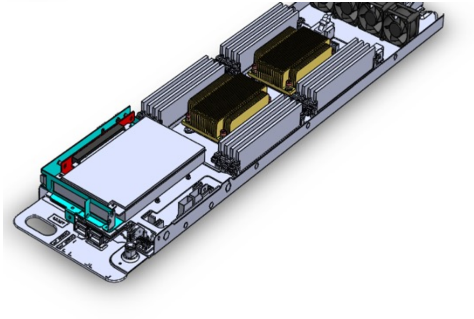
(PowerEdge FC830)

(OCP Tioga Pass)

3 servers / 1 OU
- dual CPUs
- 12x memory sockets
- two 2U heatsinks
- two 2U fans
- eight NVMe drives M.2
- shared PSUs

12 sockets,
72 dimms



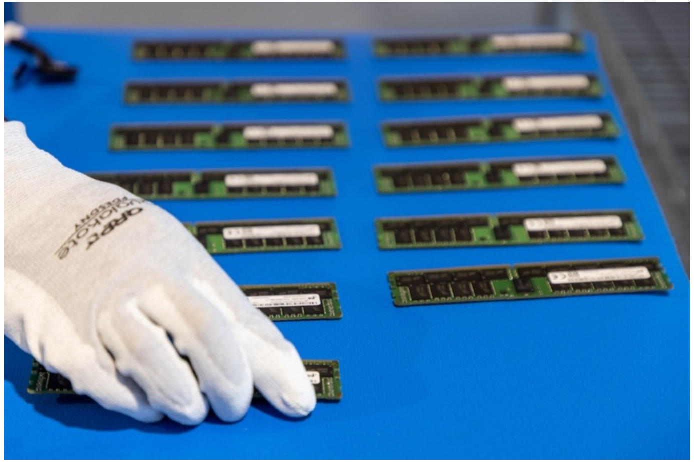


	server	weight	rack	total	
cpus	2	54	192	10.4	3,840
dimms	16	10	1536	15.4	cores
heatsinks	2	185	192	35.5	
motherboard	1	325	96	31.2	
fans	2	30	192	5.8	
M.2 NVMe	6	75	576	43.2	over 3
AVA board	2	307	192	58.9	cores
NIC 2x 25G mezz	1	154	96	14.8	per kg
PCI riser (x16 + x8)	1	92	96	8.8	
--	--	0	48	0.0	
psus	--	5.5 kg	10	55.0	
chassis	1	75 kg	96	720.0	
power shelf	--	25 kg	1	25.0	
busbar	--	15 kg	1	15.0	TOTAL
rack	--	175 kg	1	175.0	1215 kg

workload / carbon

Example – DIMMs

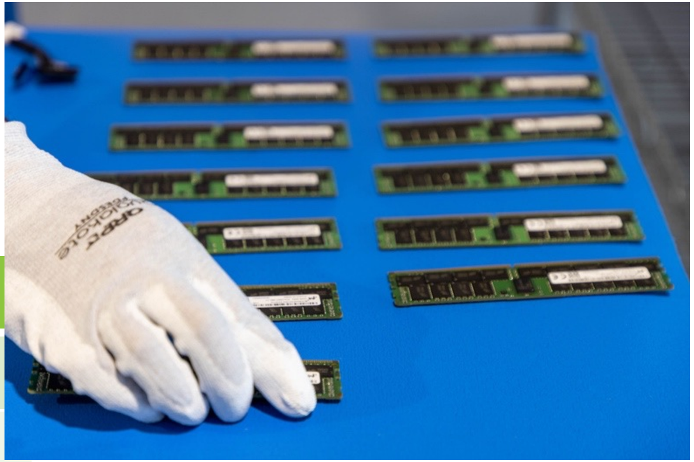
- Dell R630
  - 24x DDR4 modules, 2x cpu Broadwell
- OCP Leopard/v2
  - 16x DDR4 modules, 2x cpu Broadwell
- OCP Tioga Pass
  - 12x DDR4 modules , 2x cpu SkyLake
- Dell FC830
  - 32x DDR modules , 4x cpu Broadwell



Examples – DIMMs by Size

per server GB

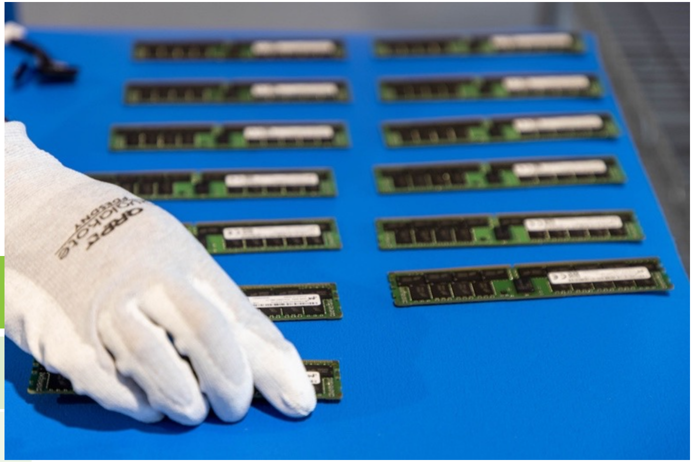
	R630	Leopard	Tioga Pass	FC830
8 GB DDR4	192	128	96	256
16 GB DDR4	**384**	256	192	**512**
32 GB DDR4	**768**	**512**	**384**	1024
64 GB DDR4	1536	1024	**768**	2048



Examples – DIMMs by Size

per rack GB

	R630	Leopard	Tioga Pass	FC830
8 GB DDR4	7,680	6,144	9,216	10,240
16 GB DDR4	15,360	12,288	18,432	20,480
32 GB DDR4	30,720	24,576	36,864	40,960
64 GB DDR4	61,440	49,152	73,728	81,920



See It For Yourself

OCP Experience Center - Nautilus (Stockton, CA) - hosted by Flax Computing

OCP Experience Center - Nautilus (Stockton, CA) - hosted by Flax Computing The Sustainable Server Lab (SSL) center for open hardware i...

Solution Provider: Flax Computing

Model #: OCP Experience Center - Nautilus (Stockton, CA)

OCP Experience Center

OCP Experience Center - MGHPCC (Holyoke, MA) - hosted by Flax Computing

OCP Experience Center - MGHPCC (Holyoke, MA) - hosted by Flax Computing The Sustainable Server Lab (SSL) center for open hardware i...

Solution Provider: Flax Computing

Model #: OCP Experience Center - MGHPCC (Holyoke, MA)

OCP Experience Center

Efficient Computing and Energy Reduction Test Center - Hosted by Flax Computing

The Recertification for Efficient Computing and Energy Reduction Test (RECERT) design and manufacturing center hosts a range of activities related to advancing ...

Solution Provider: Flax Computing

Model #: N/A

Carbon Footprint Analysis and Reduction (CFAR) Center - Hosted by Flax Computing

The Carbon Footprint Analysis and Reduction (CFAR) analysis and design process allows everyone to succinctly and accurately measure the carbon footprint of thei...

Solution Provider: Flax Computing

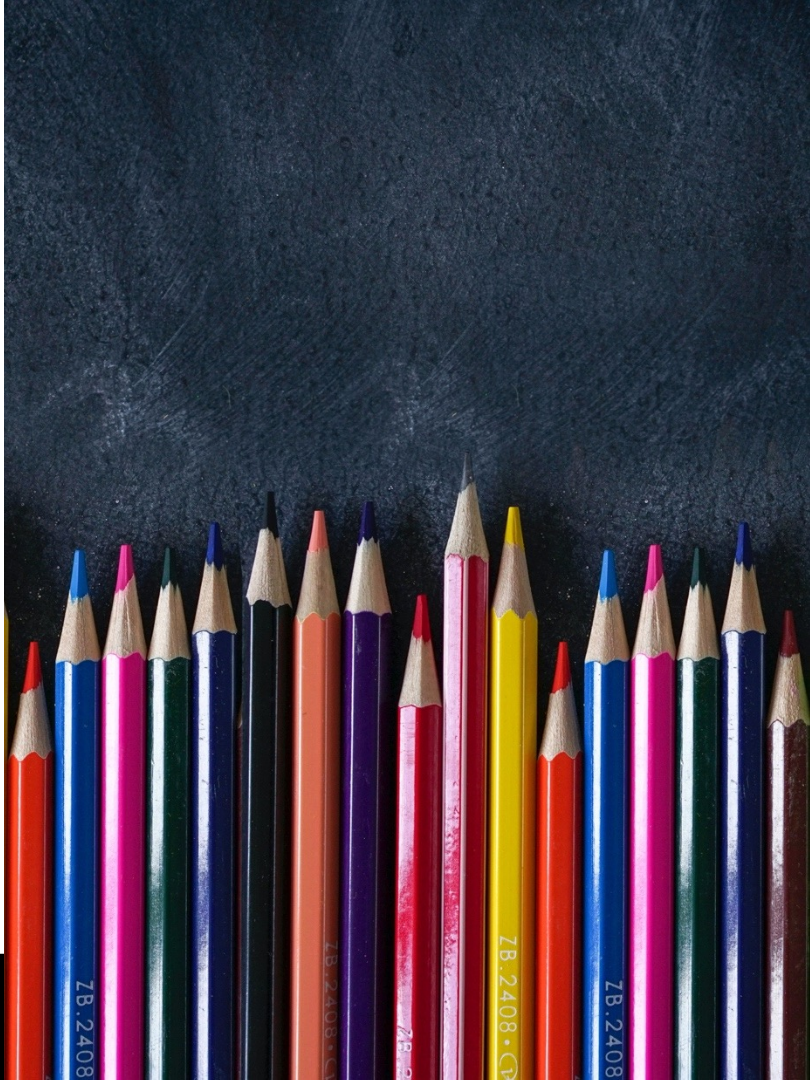
Model #: Carbon Footprint Analysis and Reduction Center

Call to Action

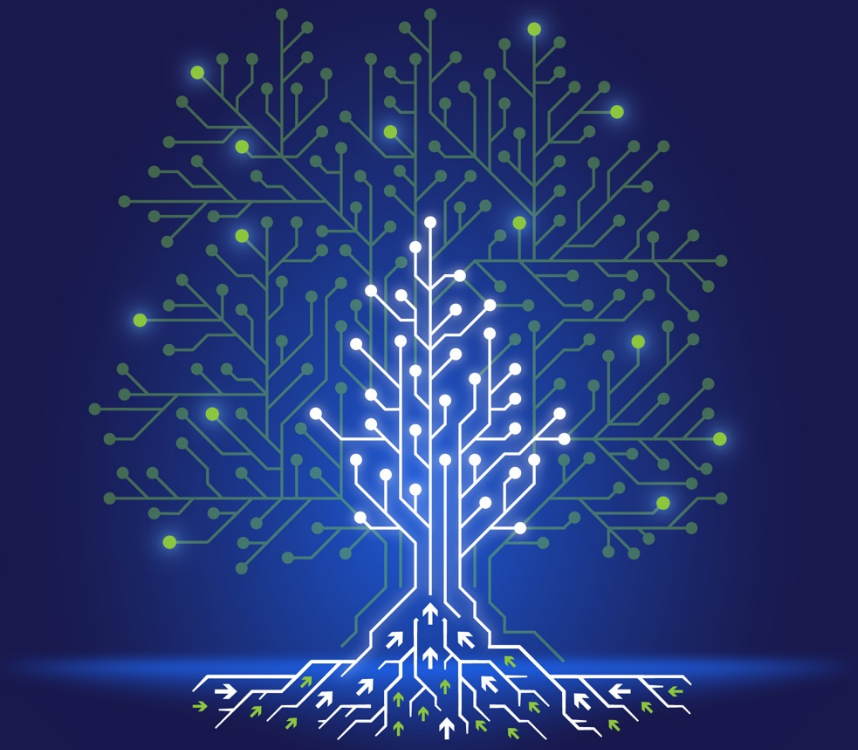
- Reach out to us to get involved
- Engage us to evaluate / quantify your server carbon footprints
  - www.flaxcomputing.com
  - www.simslifecycle.com
- Evaluate your own servers, share the results with us
  report @ flaxcomputing.com
- Contribute measurements and component details (reuse stories, workload information, etc.) and share what components you would be most interested in

  data @ flaxcomputing.com

  sls.sustainability@simsmm.com



Thank you!



Scaling Innovation Through Collaboration

OCP REGIONAL SUMMIT

24-25 APRIL 2024
LISBON, PORTUGAL