plugin 2020

Organized by

PIM

Speaker

Steve Helvie | VP of Channel

steve@opencompute.org

Topic

Open source hardware for AI and its impact on the data center

plugin

Today's Topics

Change in Hardware

AI Accelerators

Data Center Density

Advanced Cooling

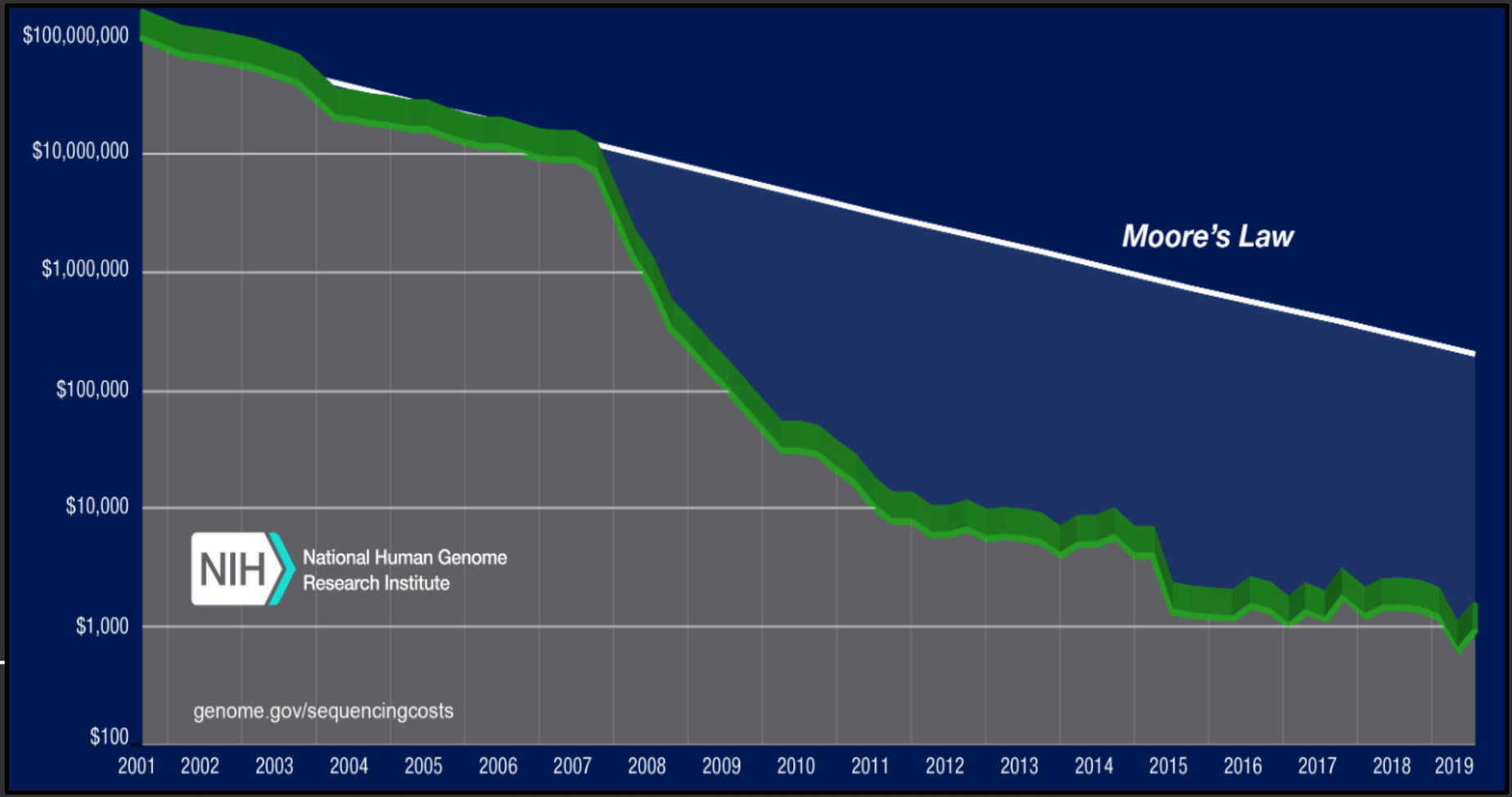
Predictive Server Modelling

Data Center Software

Circular Economy

Reuse at Scale

Cost per Human Genome

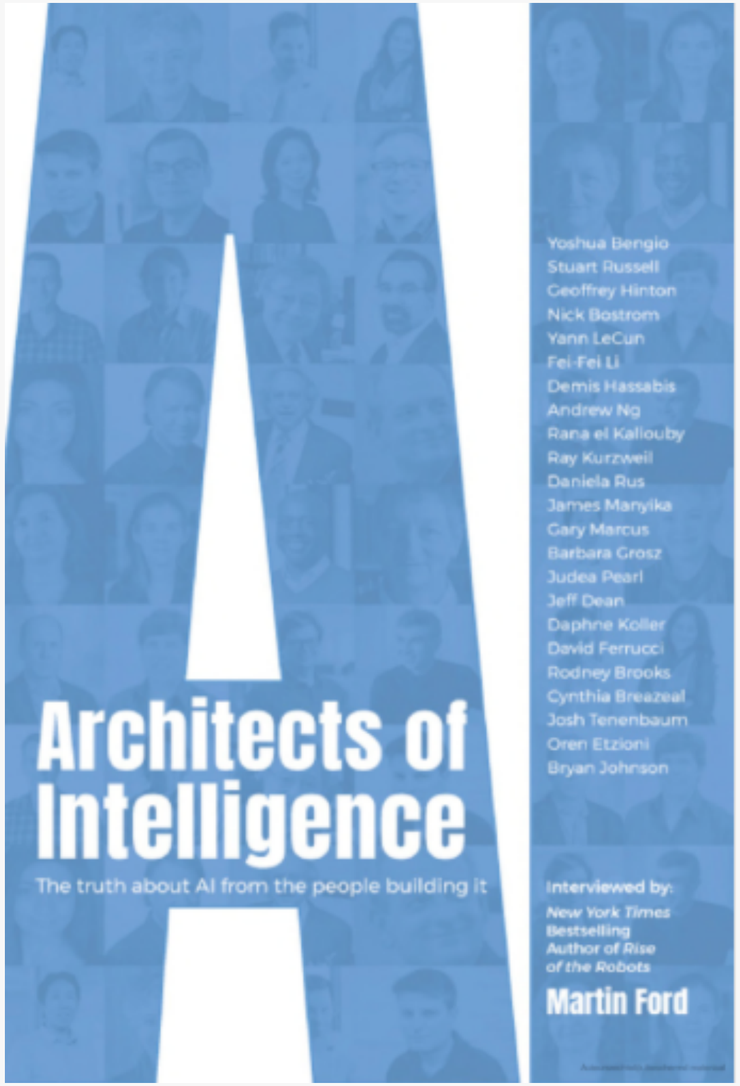


"Full and free exchange of all sequence data"

"Publicly accessible database within 24 hours "

https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost

plugin

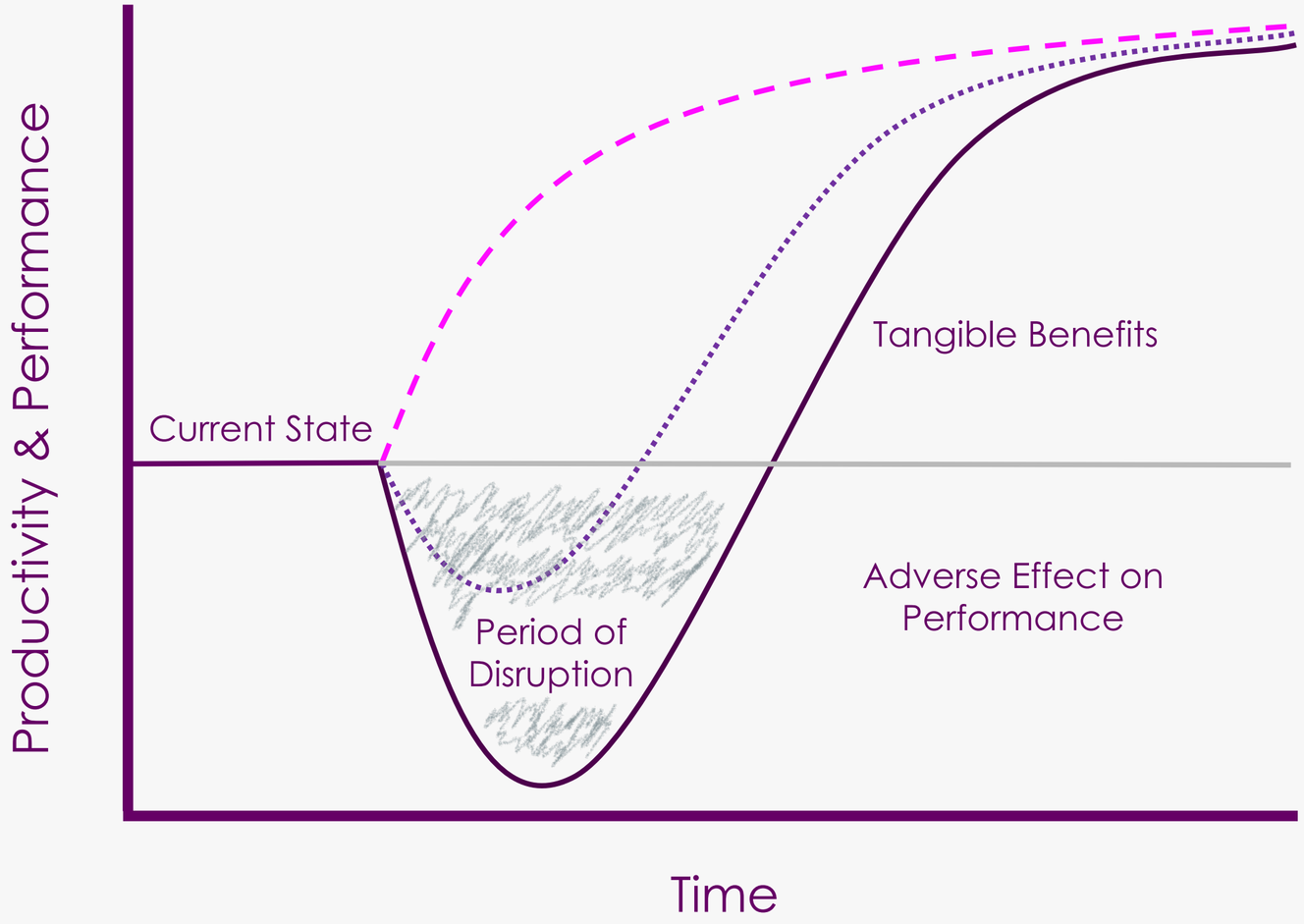


"Architects of Intelligence"

By Martin Ford

plugin

The AI J-Curve



What stakeholders (mistakenly) expect

What stakeholders can expect with proper execution

What actually happens in most cases

https://en.wikipedia.org/wiki/J_curve

OPEN Compute Project®

January 2010

COMPANIES > FACEBOOK

Facebook to Build Its Own Data Centers

Facebook has decided to begin building its own data centers, and may announce its first facility as soon as tomorrow. The fast-growing social network has previously leased server space from wholesale data center providers.

Rich Miller | Jan 20, 2010

plugin

What can we remove from the system?

Can we raise operating temperatures and have the servers survive?

Can we increase relative humidity operational ranges to make the system much more efficient?

Do we need a centralized power supplies?

Open Compute Project

A collaborative community focused on redesigning hardware technology to efficiently support the growing demands on compute infrastructure.

plugin

150+ companies
189 contributions
6K engineers

Our Projects

Networking

Server

Storage

Rack & Power

Open Accelerator Infra

Advanced Cooling

Data Center

Modular DC

Telco

openEDGE

HW Mgmt

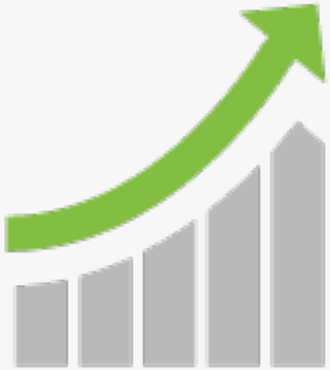
Open System Firmware

HPC

Security

https://www.opencompute.org/projects

Efficiency



Scale

Impact

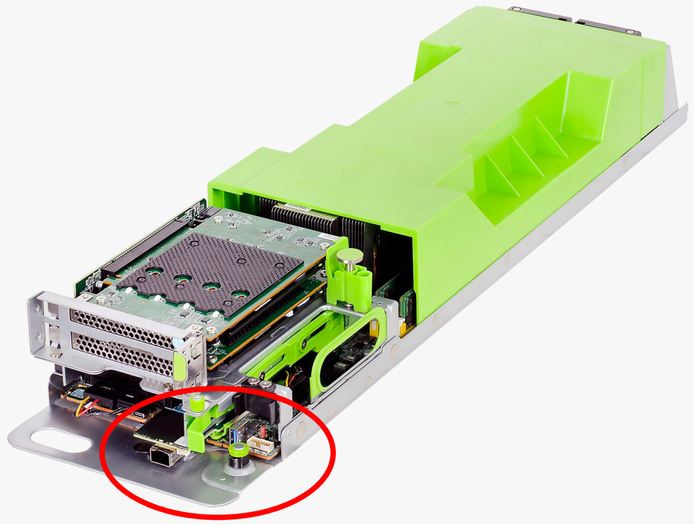
Openness

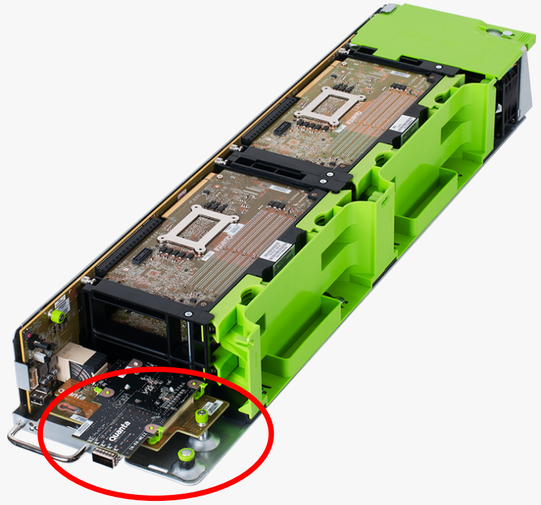
Server bezels

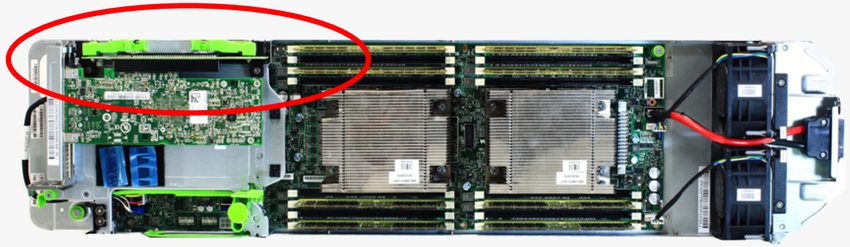
….gratuitous differentiation

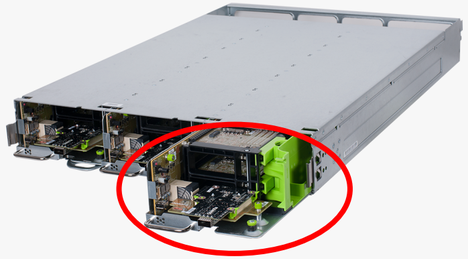
plugin

Tool less OCP vanity free open source cubby servers



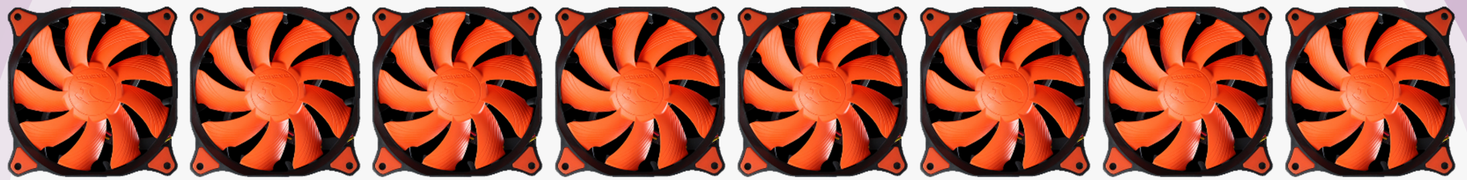






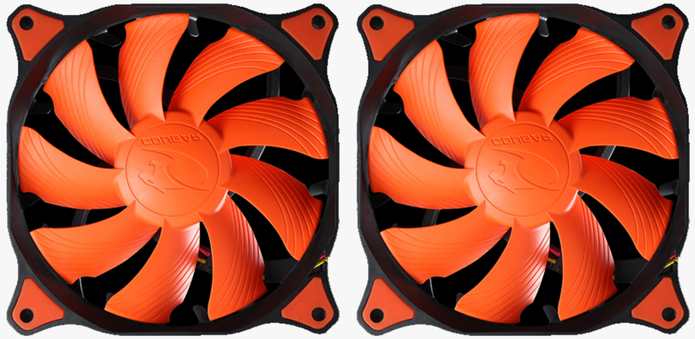
Open Rack Unit Innovations
Server Fans “Cube Law”

40 mm
Fans



EIA
“U”

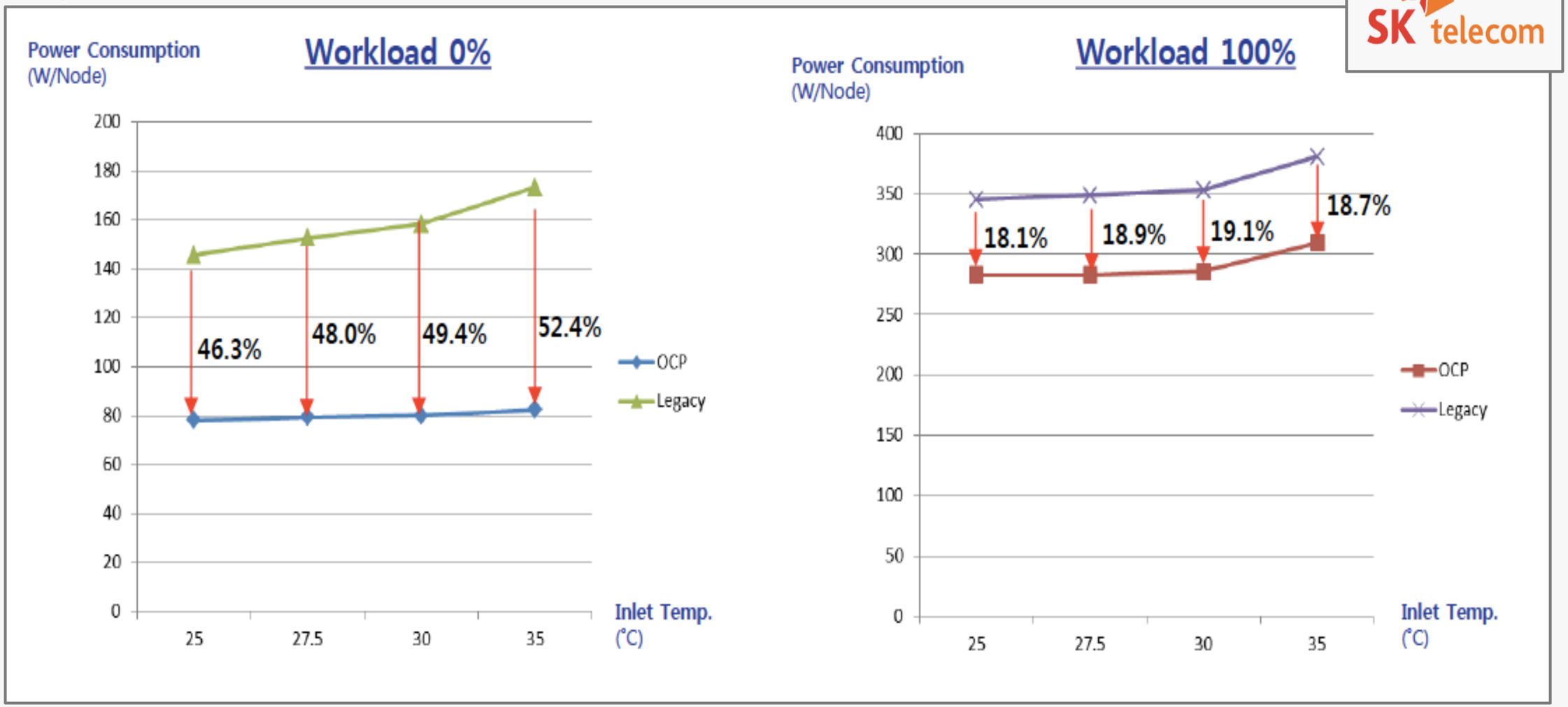
½ the speed
7/8th energy reduction



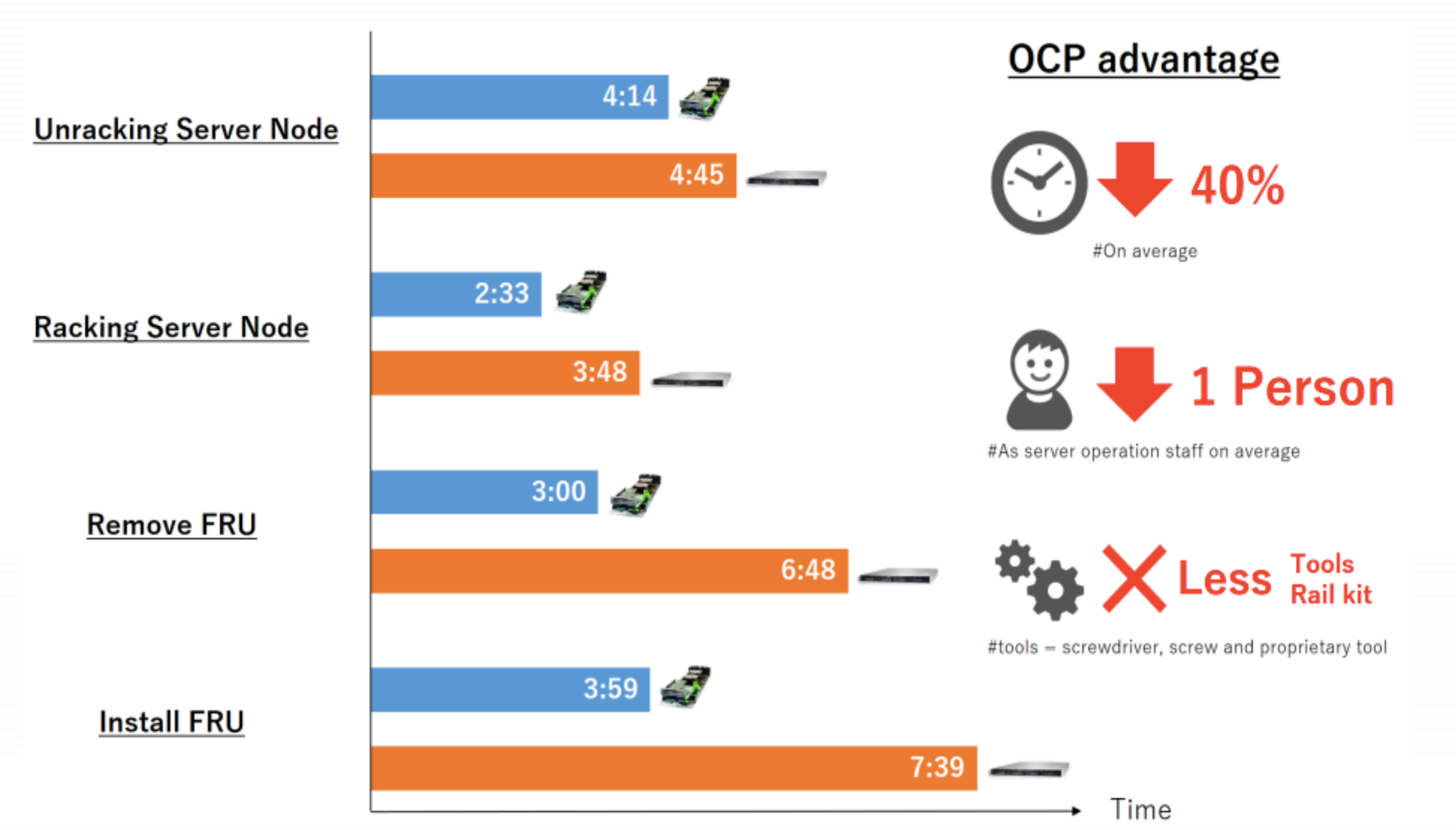
80 mm
Fans

plugin

OCP Power Consumption



https://www.opencompute.org/events/past-summits



Traditional

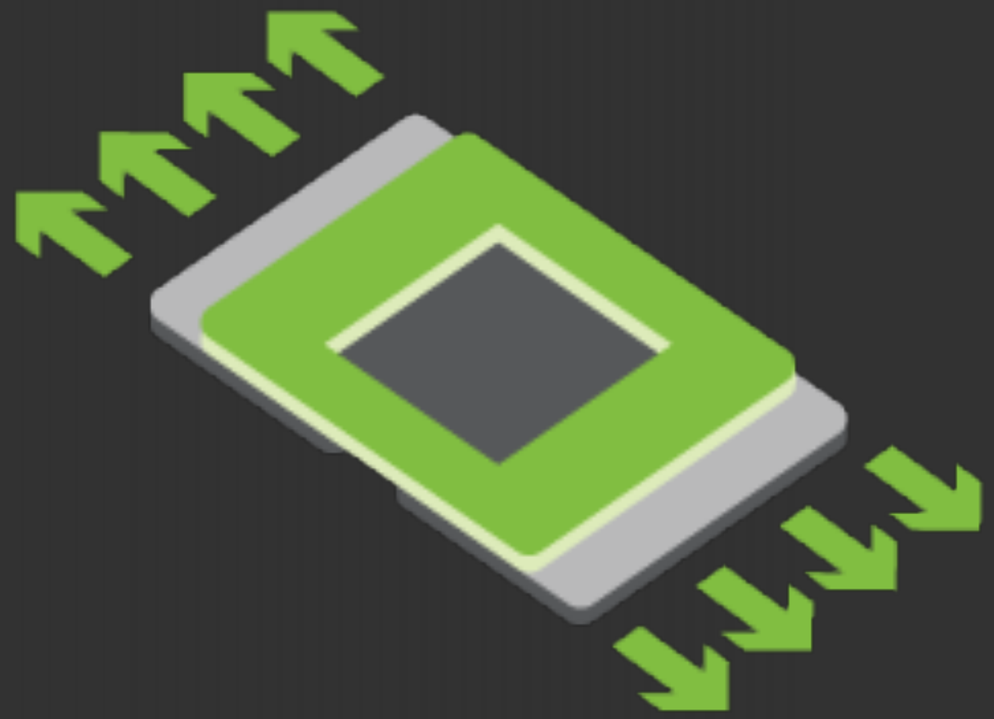
OCP

YAHOO! JAPAN

https://www.opencompute.org/events/past-summits

plugin

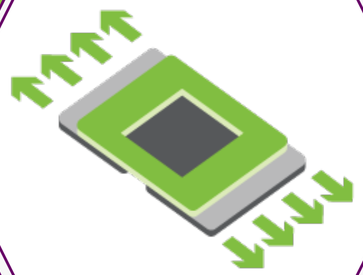
Open Accelerator Infrastructure Project (OAI)



OPEN Compute Project®

facebook

Open ACCELERATOR Infrastructure



https://www.opencompute.org/blog/new-open-accelerator-infrastructure-oai-sub-project-to-launch-within-the-ocp-server-project

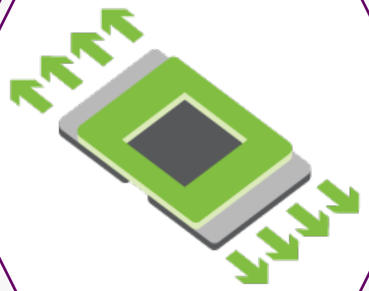
plugin

OPEN Compute Project®

Open ACCELERATOR Infrastructure

facebook

Microsoft Azure



Baidu 百度

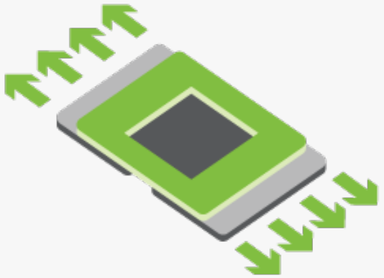
plugin

OPEN Compute Project®

Open Accelerator Infrastructure

facebook
Microsoft Azure
intel
Qualcomm
XILINX
habana
NVIDIA
BittWare a molex company
Baidu 百度
Google
inspur
Lenovo
AMD
IBM
wiwynn
Alibaba Group
Tencent
molex
PENGUIN COMPUTING
A subsidiary of SMART Global Holdings, Inc.
GRAPHCORE
HUAWEI
QCT

plugin



Open Accelerator Infrastructure (OAI)

01 OAM: Open Accelerator Module

02 UBB: Universal Baseboard

03 PSB: PCIe Switch Board

04 SCM: Secure Control Module

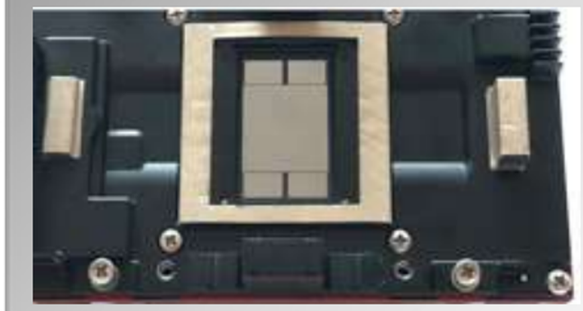
05 Tray

06 Chassis

https://www.opencompute.org/blog/new-open-accelerator-infrastructure-oai-sub-project-to-launch-within-the-ocp-server-project

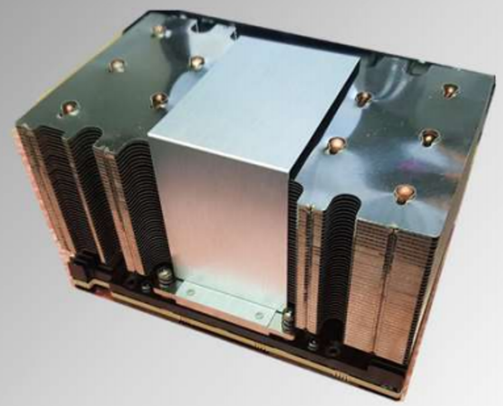
OPEN Compute Project ®

Open Accelerator Module (OAM)

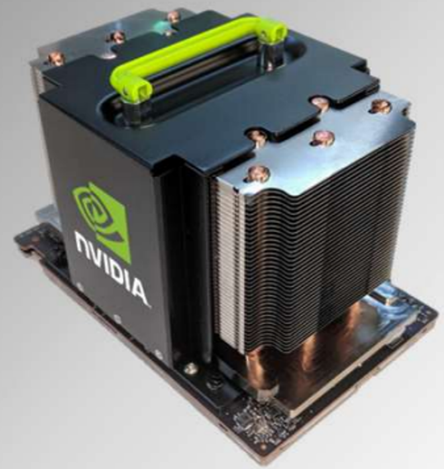


Nervana™ NNP-T OAM
Intel

Gaudi OAM
Habana

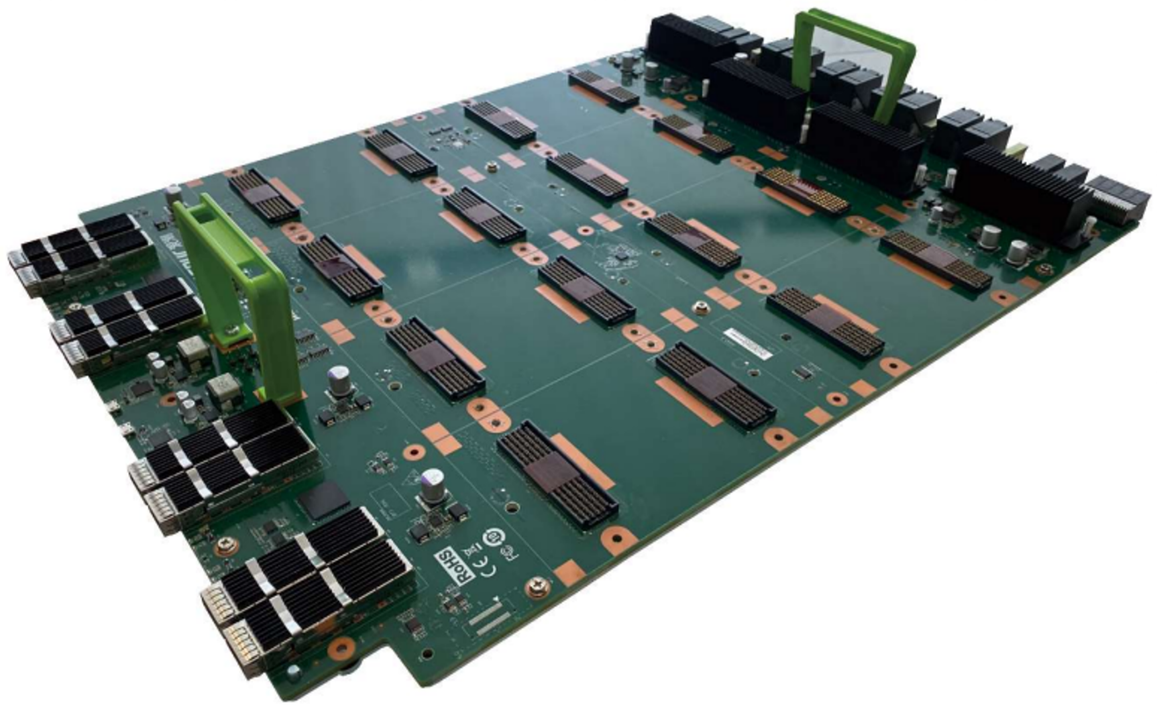


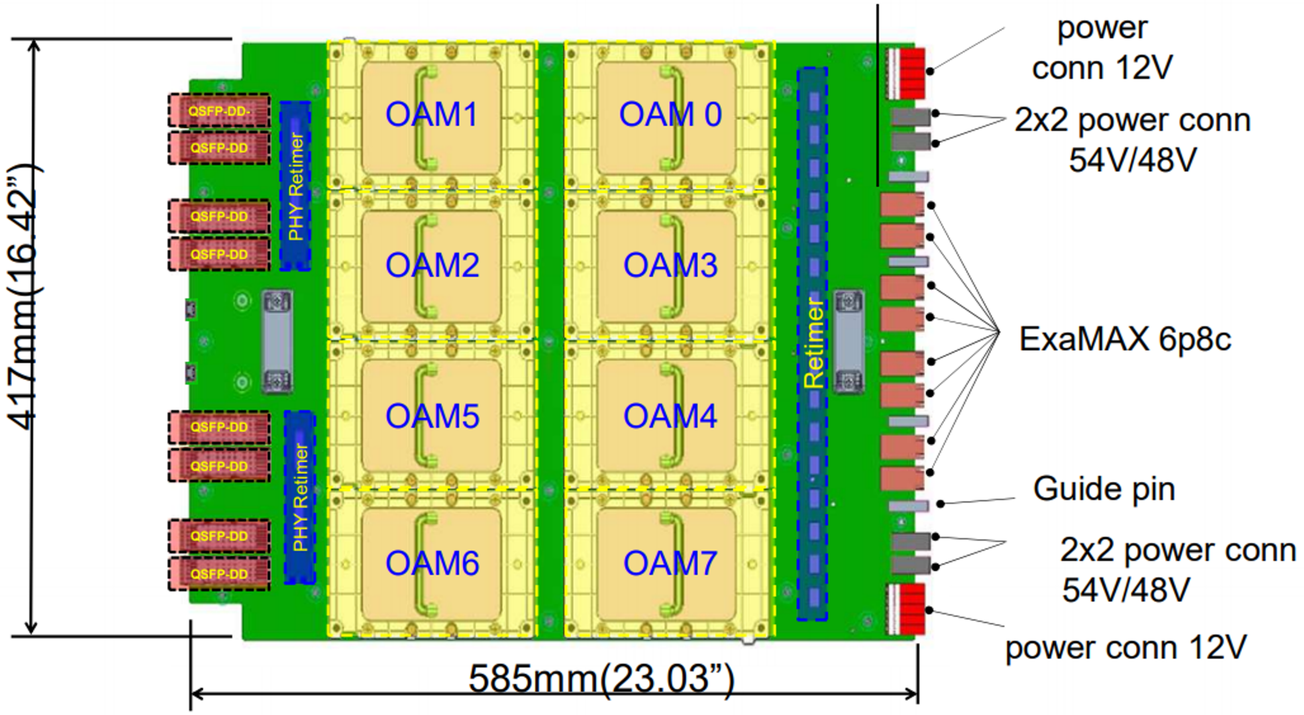
POC OAM
AMD



V100 POC OAM
Nvidia

UBB: Universal Baseboard





UBB/ Common Tray

Stacked System
Disaggregate from Host

https://onlinexperiences.com/Launch/Event.htm?ShowKey=86206&DisplayItem=E369184

Rack Level Solution

48VDC Open Rack

1 pairs 48V Bus Bar

1 shelf per Rack

System	Power	Subtotal
Inspur OAI system	20kW (5kW x4)	22.8kW
Inspur Compute Node	2.8kW (700W x4)	
Total Power/Rack		33kW

Power Shelf

33KW(12xPSU)

40V-58V

93mm (H, 2OU) x 537mm (W, 21") x 586 (D) mm

System Devices

Inspur 3OU OAI systems x4

Inspur 2OU compute node x4

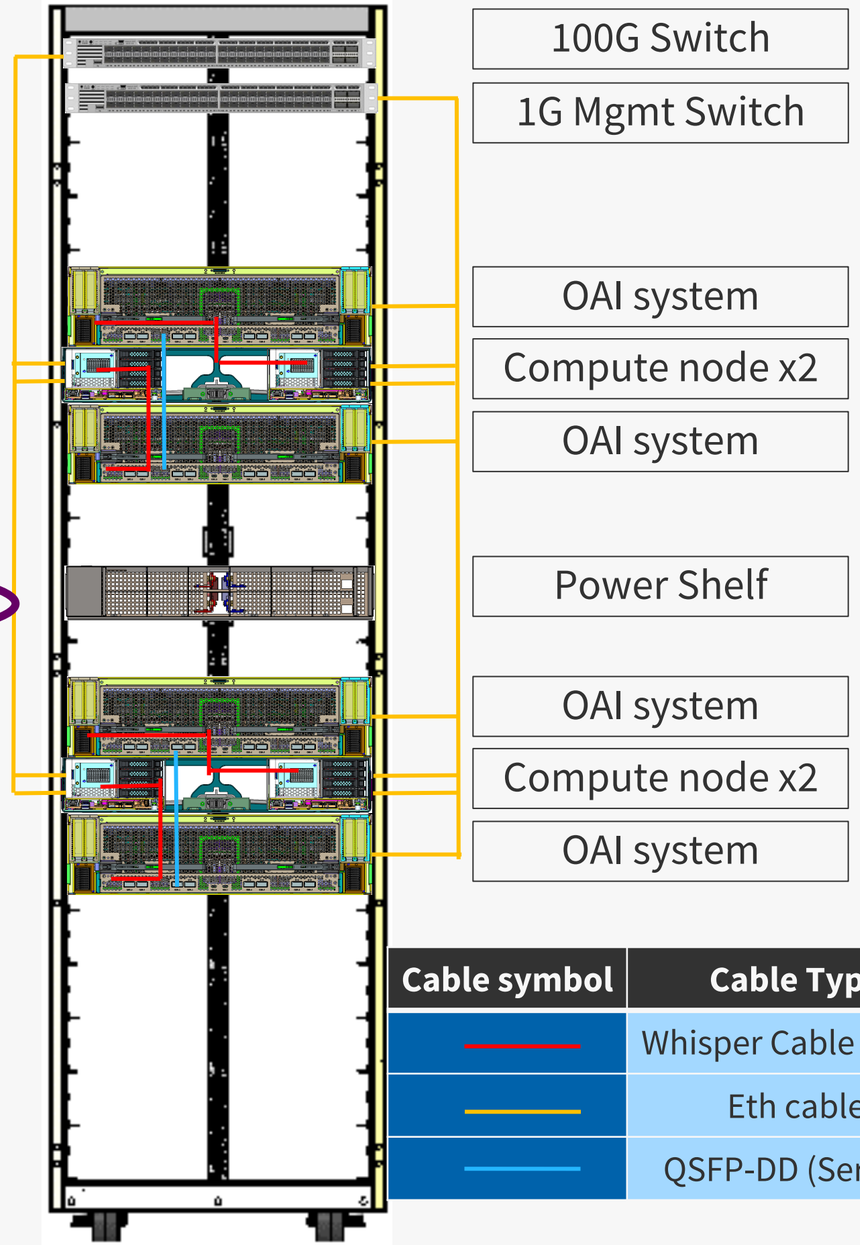
OAM Vendor

Habana Gaudi

Nvidia

AMD

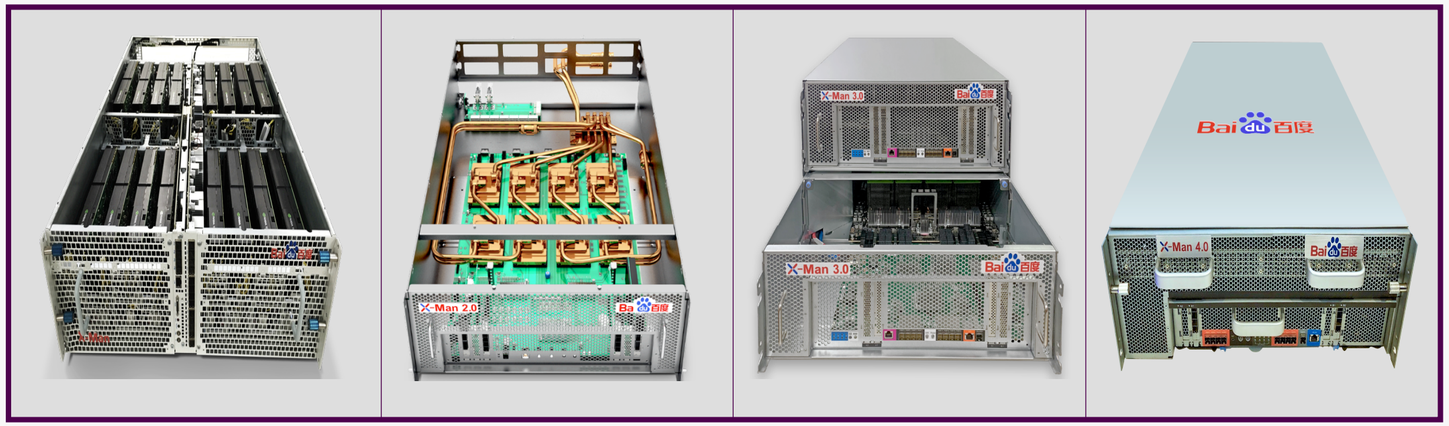
..etc



Cable symbol	Cable Type
	Whisper Cable (PCIe)
	Eth cable
	QSFP-DD (Serdes)

https://www.opencompute.org/solutions/4/open-cloud-ai-rack

2016Q2 X-MAN 1.0	2017Q4 X-MAN 2.0	2018Q4 X-MAN 3.0	2019Q4 X-MAN 4.0
Highest Density	Liquid Cooling	Modular Design	OAI Compatible



Baidu is actively participating in OAI project and will build X-Man4.0 based on OAI

Baidu AIBox is able to train 10TB sparse models with single X-Man, and also scale to more nodes for bigger models and more samples.
Papers @ MLSys 2020
https://proceedings.mlsys.org/papers/2020/173

AIBox is good show case for OAI Fully Connected topology

https://www.opencompute.org/summit/virtual-summit

https://onlinexperiences.com/Launch/Event.htm?ShowKey=86206&DisplayItem=E369188

Rack & Power Project

OPEN Compute Project ®

https://engineering.fb.com/data-center-engineering/open-rack/

Converged rack frame

Flexible power shelf

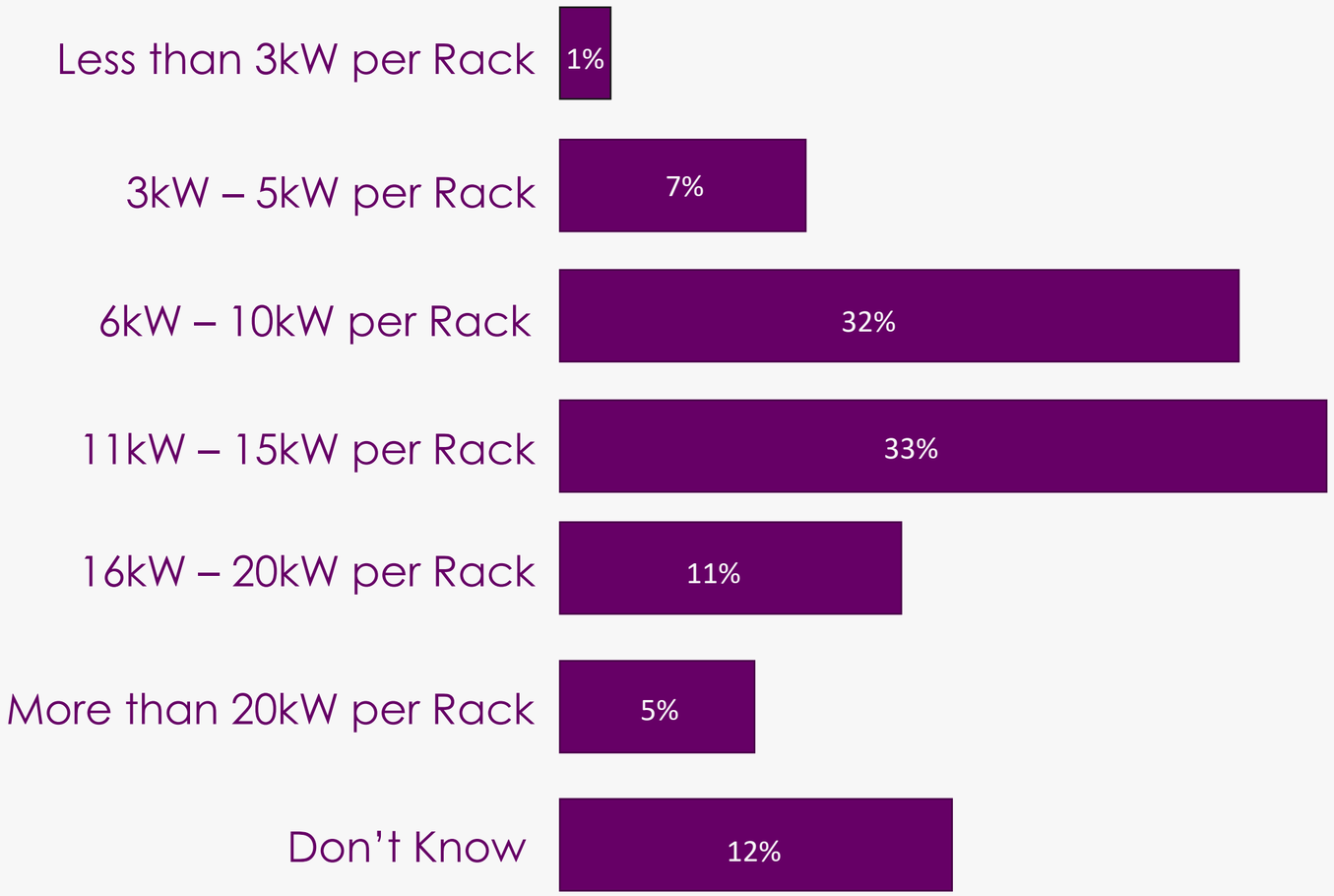
Universal AC power interconnect

Pluggable DC power shelf output interconnect

Battery backup systems

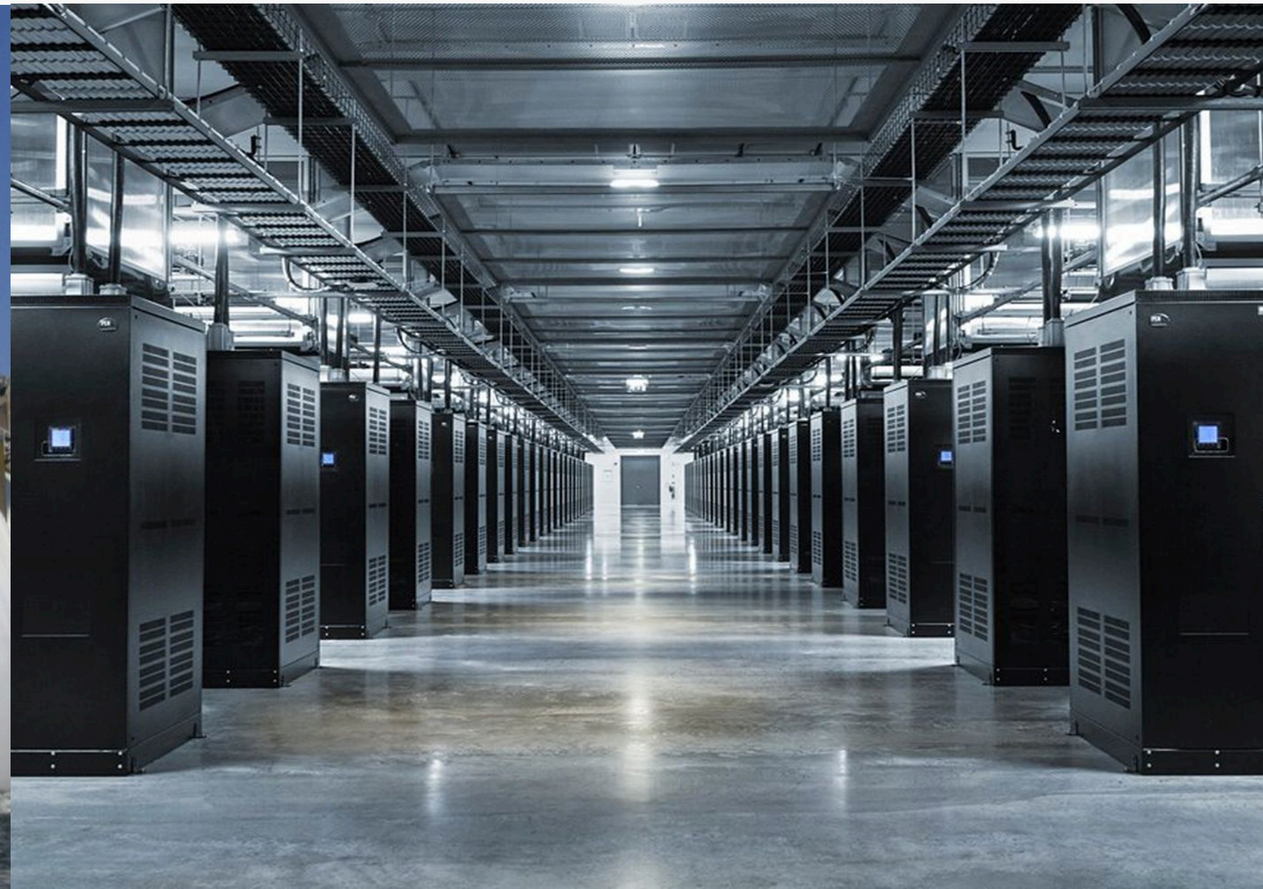
plugin

The Impact of AI on the average Density per Rack



Source: 451 Research 2019

Gartner, Inc., predicts that more than 30 % of data centers that fail to sufficiently prepare for AI will no longer be operationally or economically viable by 2020.



Advanced Cooling Solutions(ACS) Sub-Project

Advanced Cooling (ACS) Sub-Project

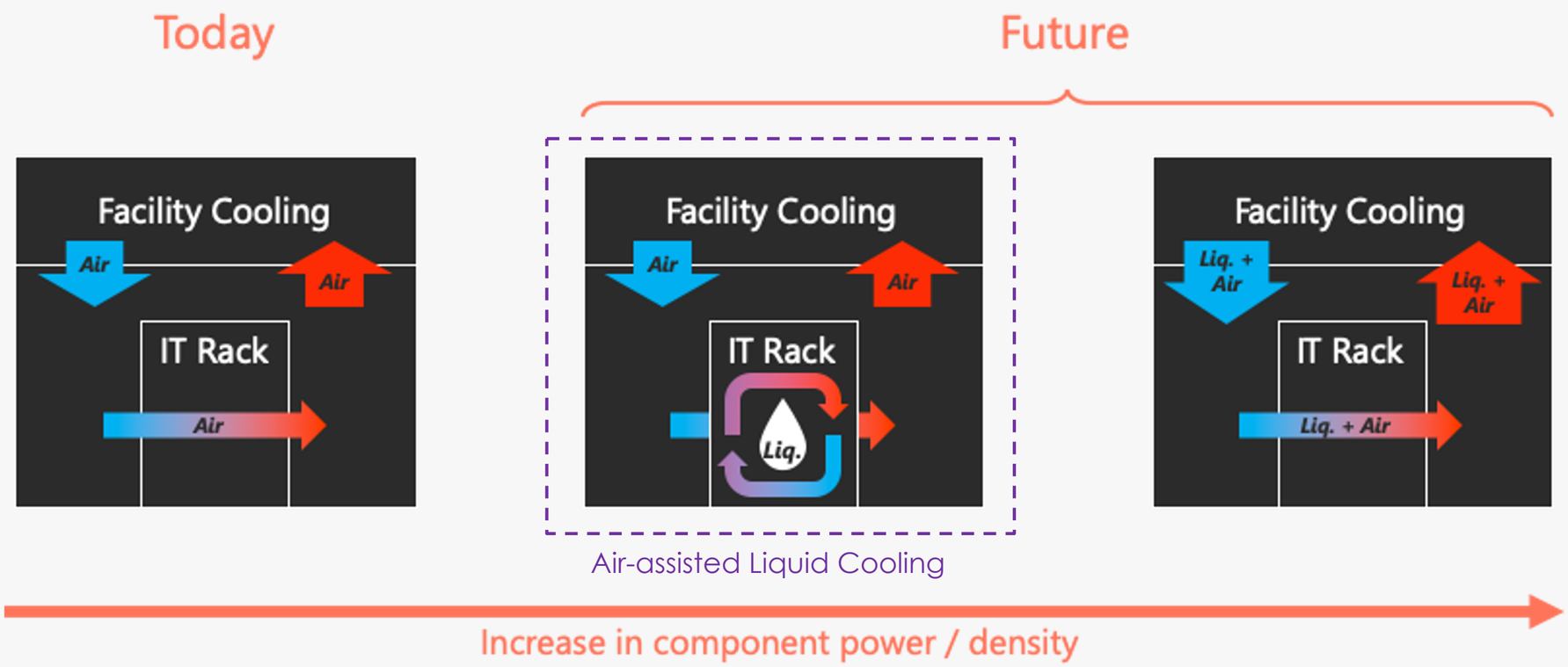
01 Fully Immersed Systems

02 Cold Plate Direct Contact Liquid Loop

03 Door-based Heat Exchangers

Stretching Limits

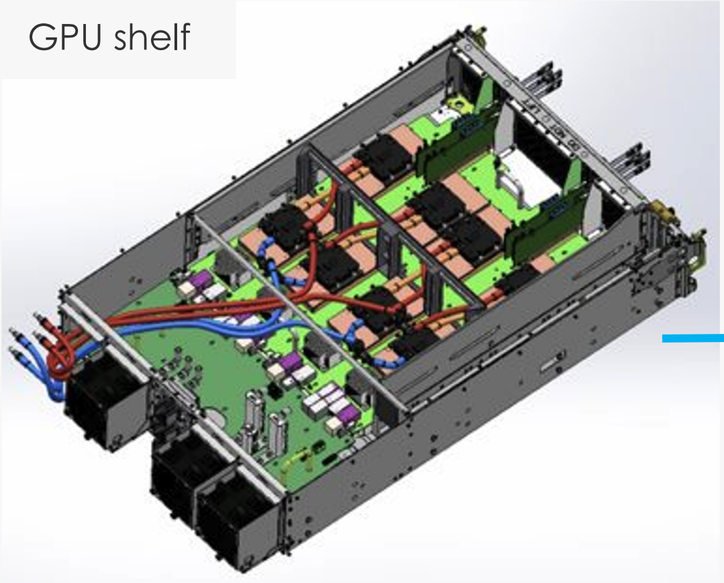
Extend life of air-cooled facilities & permit smooth transition



https://onlinexperiences.com/Launch/Event.htm?ShowKey=86206&DisplayItem=E369065

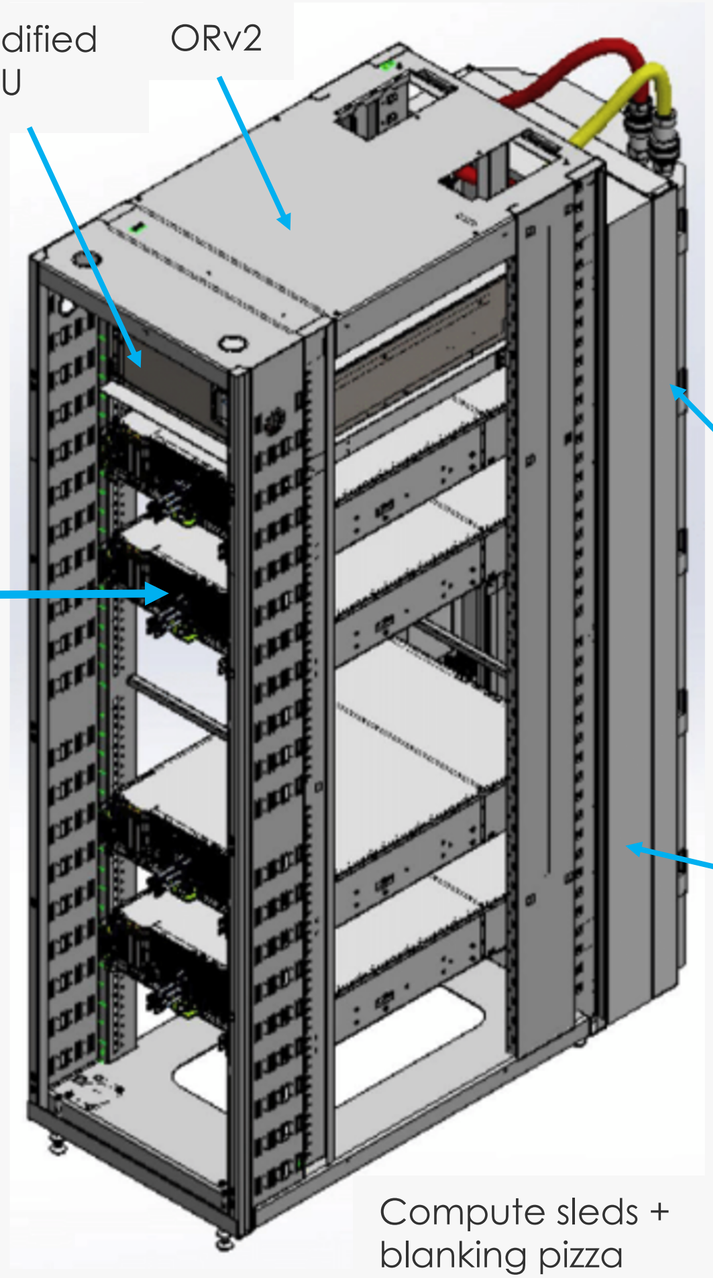
Case Study

GPU shelf



Modified CDU

ORv2



Rear door HX

Adapter frame + manifolds (hidden)

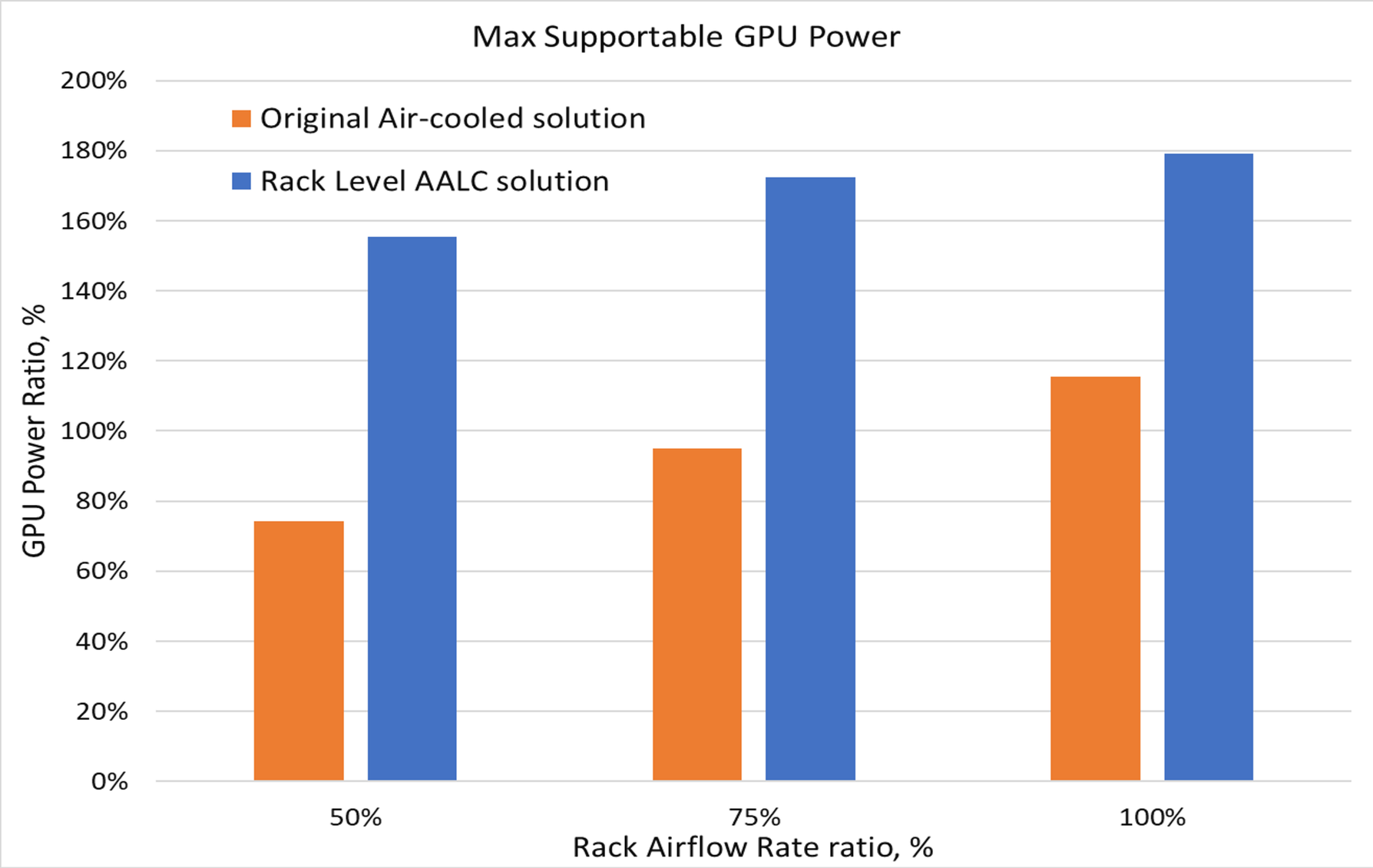
Compute sleds + blanking pizza boxes not shown

Heat flow: GPU $\xrightarrow{\text{Coldplate}}$ Coolant $\xrightarrow{\text{RDHx}}$ Air

Without facility water supply

https://onlinexperiences.com/Launch/Event.htm?ShowKey=86206&DisplayItem=E369065

Promising: +50% power compared to original air-cooled solution



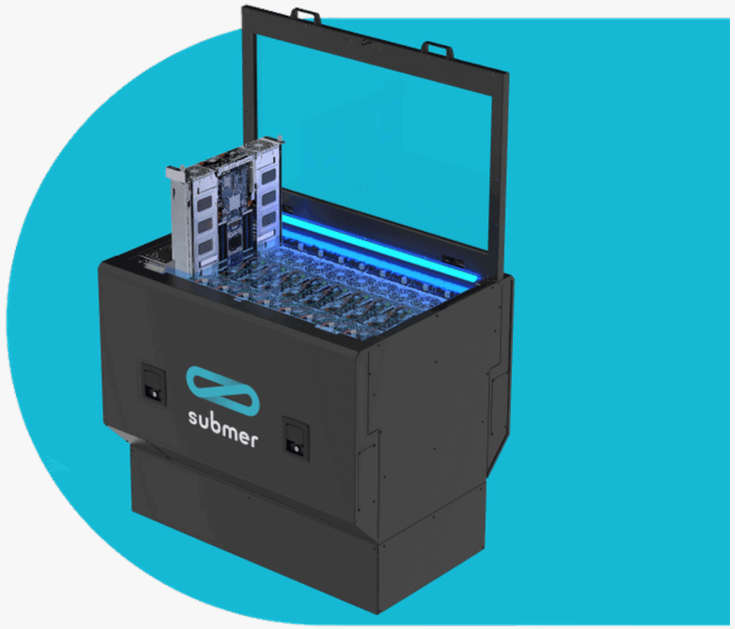
Hyper… Aggregated… Empowered… Edge

https://techwireasia.com/2020/01/walmart-takes-on-amazon-with-edge-computing-and-5g/

OPEN Compute Project ®

submer

https://submer.com/



https://dcx.eu/

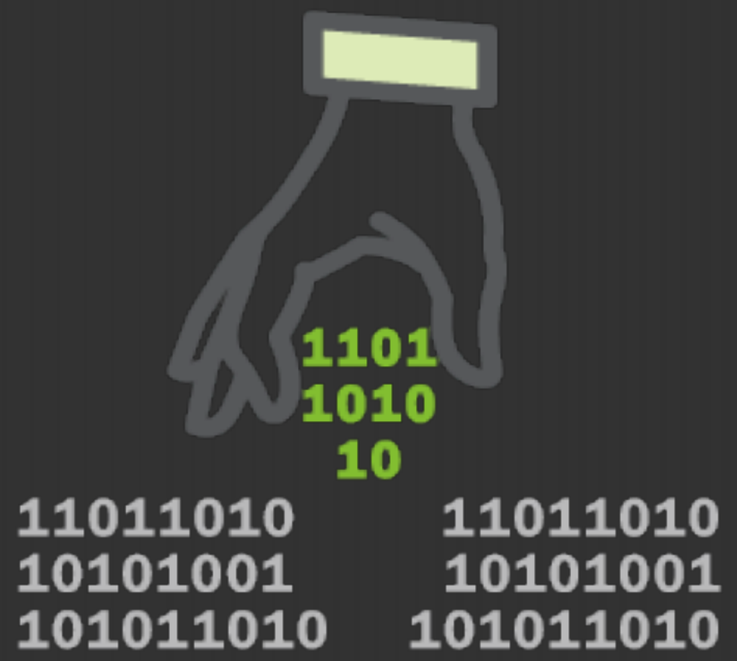
Asperitas IMMERSED COMPUTING®

https://www.asperitas.com/

https://www.opencompute.org/projects/acs-immersion-cooling-sub-project

plugin

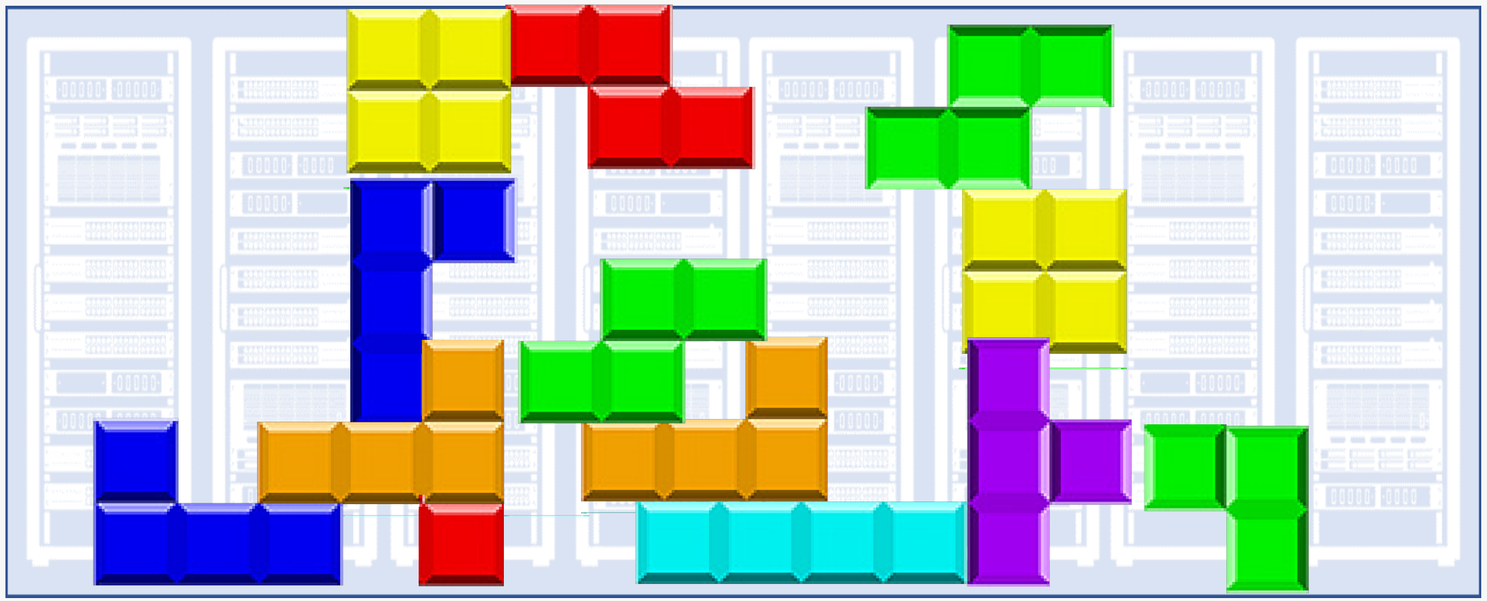
Predictive Server Modelling



https://www.opencompute.org/circular-economy

Workloads normally placed ……..

Energy & Cost



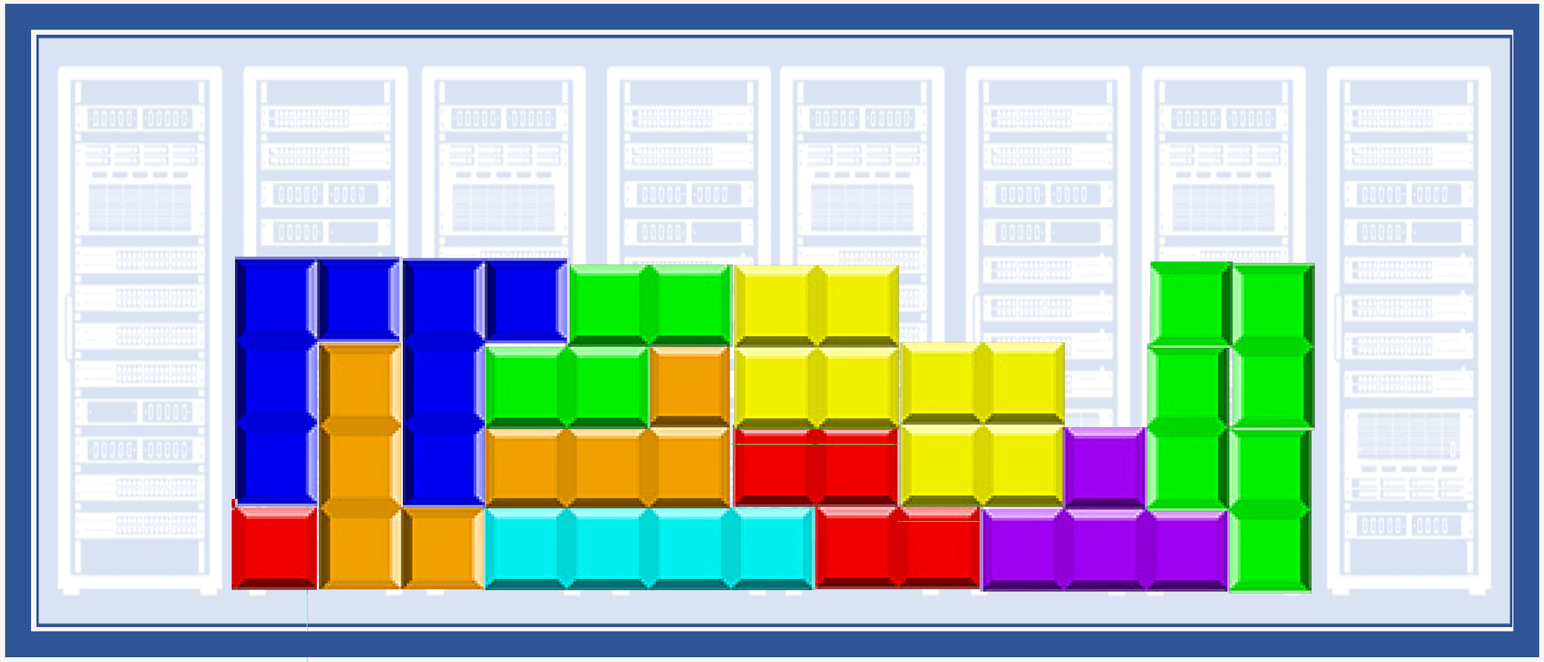
Efficiency & Throughput

https://www.edgetic.com/

plugin

Intelligent ML optimisation reduces energy by > 20%

Energy & Cost



Efficiency & Throughput

https://www.edgetic.com/

plugin

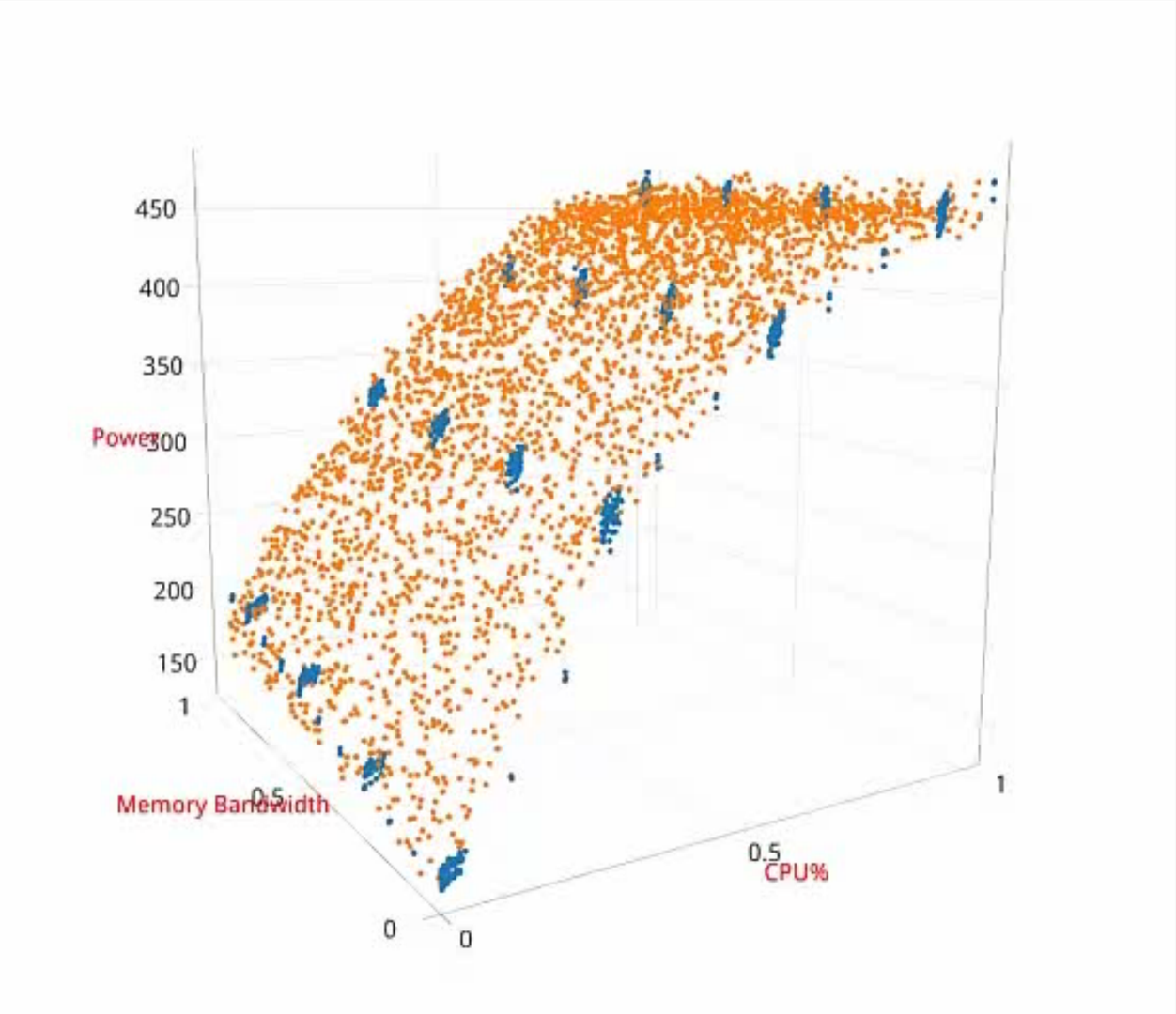
Actual & Predictive Server Modelling

Actual readings (in blue)

Predicted values (in orange)

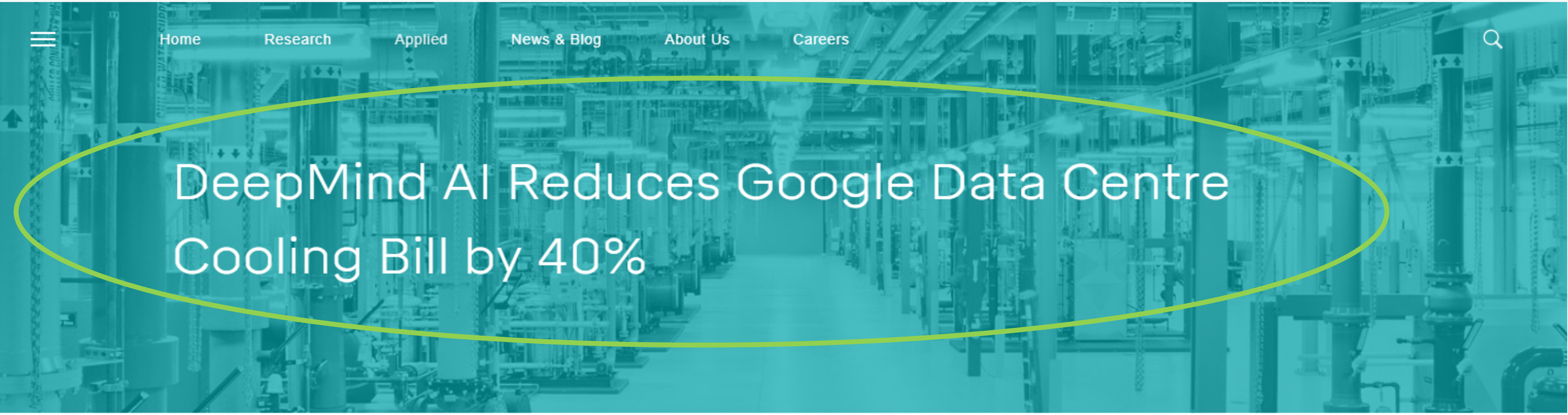
For a given software loading, one can predict over 100 attributes to 97% accuracy

This includes heat output & power consumption



https://www.edgetic.com/

plugin



https://deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-40/

plugin

Circular Economy

https://www.opencompute.org/circular-economy

There is no such thing as "away"….when we throw things away it must go somewhere.

Annie Leonard

plugin

The vision is a **data center** industry where no **heat** is wasted….

DATA | ECONOMY NEWS ▾ ECONOMY ▾ BUSINESS ▾ MARKETS ▾ LEADERSHIP ▾ INDUSTRY ▾ LIFE & ART

Around $10bn is set to be invested in India's data centre market over the next six years, with reports predicting that $3bn will be spent building data centres in the country.

A recent research report predicts that the data centre market in India will grow at a CAGR of around 5% during the period of 2019 to 2025, according to Arizton.

The report also highlighted that data explosion and favourable government policies will aid the establishment of data centre parks across India.

As of October 2019, there were over ten data centre projects, which are expected to be operational between June 2020 and December 2022.

Adani Group, Bridge Data Centers, Colt DCS, and Yotta Infrastructure are among the new entrants to the market, targeting hyperscale investments with multiple facilities covering over 40 MW of power capacity.

Adani Group recently signed a Memorandum of Understanding with Digital Realty with the aim of enabling both companies to jointly evaluate developing and operating data centres, data centre parks and cultivating undersea cable provider communities of interest across India.

plugin

46M servers to be “End of Life” (EOL’ed)” between 2019 and 2023

ELLEN MACARTHUR FOUNDATION

https://www.ellenmacarthurfoundation.org/

CEDaCI - Circular Economy for the Data Centre Industry

https://www.nweurope.eu/projects/project-search/cedaci-circular-economy-for-the-data-centre-industry/

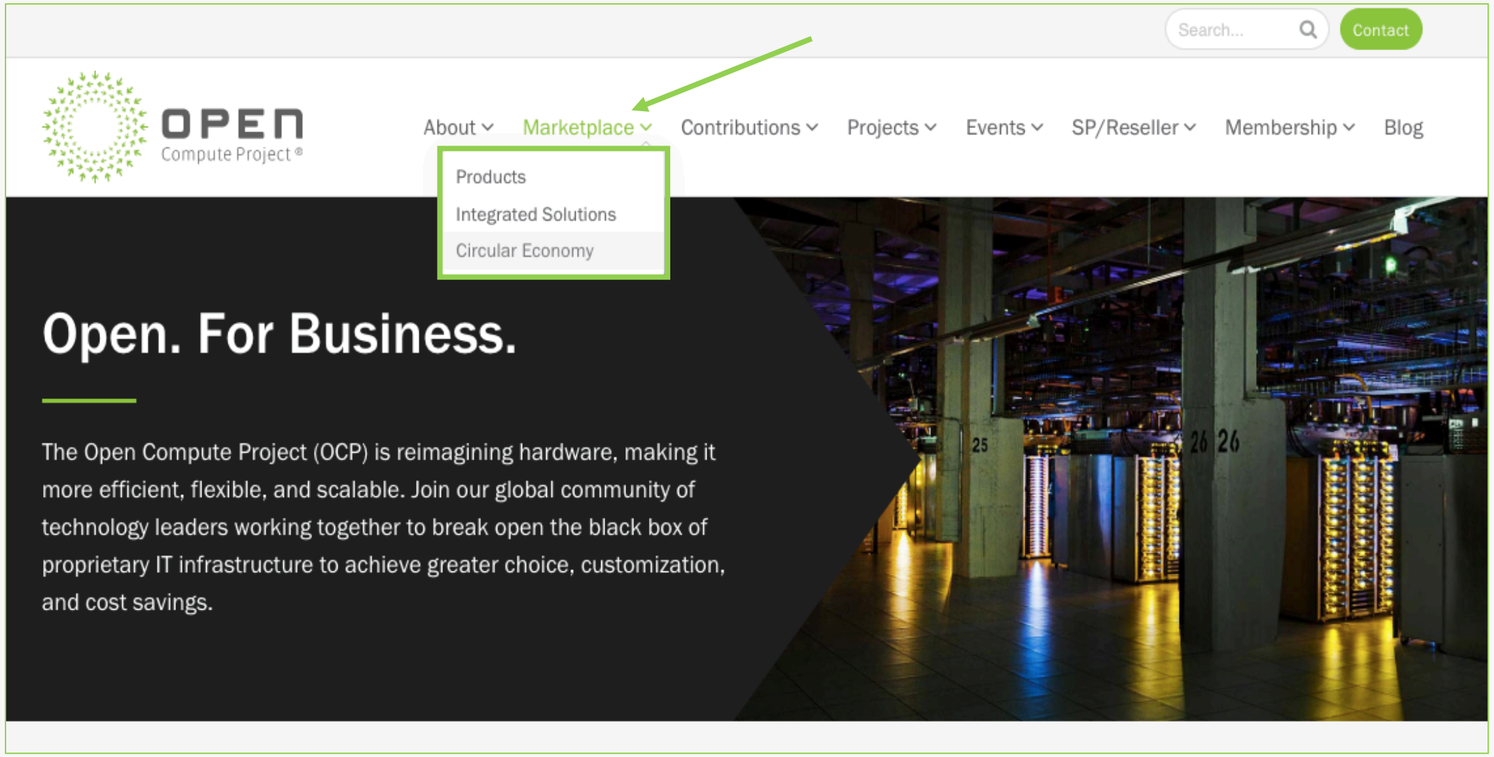
Ali Fenn – President of ITRenew

https://www.itrenew.com/

https://www.youtube.com/watch?v=PuzEMYCGeVk

ITRENEW

OPEN PLATINUM™



SESAME BY ITRENEW

https://www.opencompute.org/circular-economy

Sesame for AI/ML provides the high bandwidth, low latency performance you need to manage the huge volumes of data in your demanding learning environment.

https://onlinexperiences.com/Launch/Event.htm?ShowKey=86206&DisplayItem=E369316

plugin

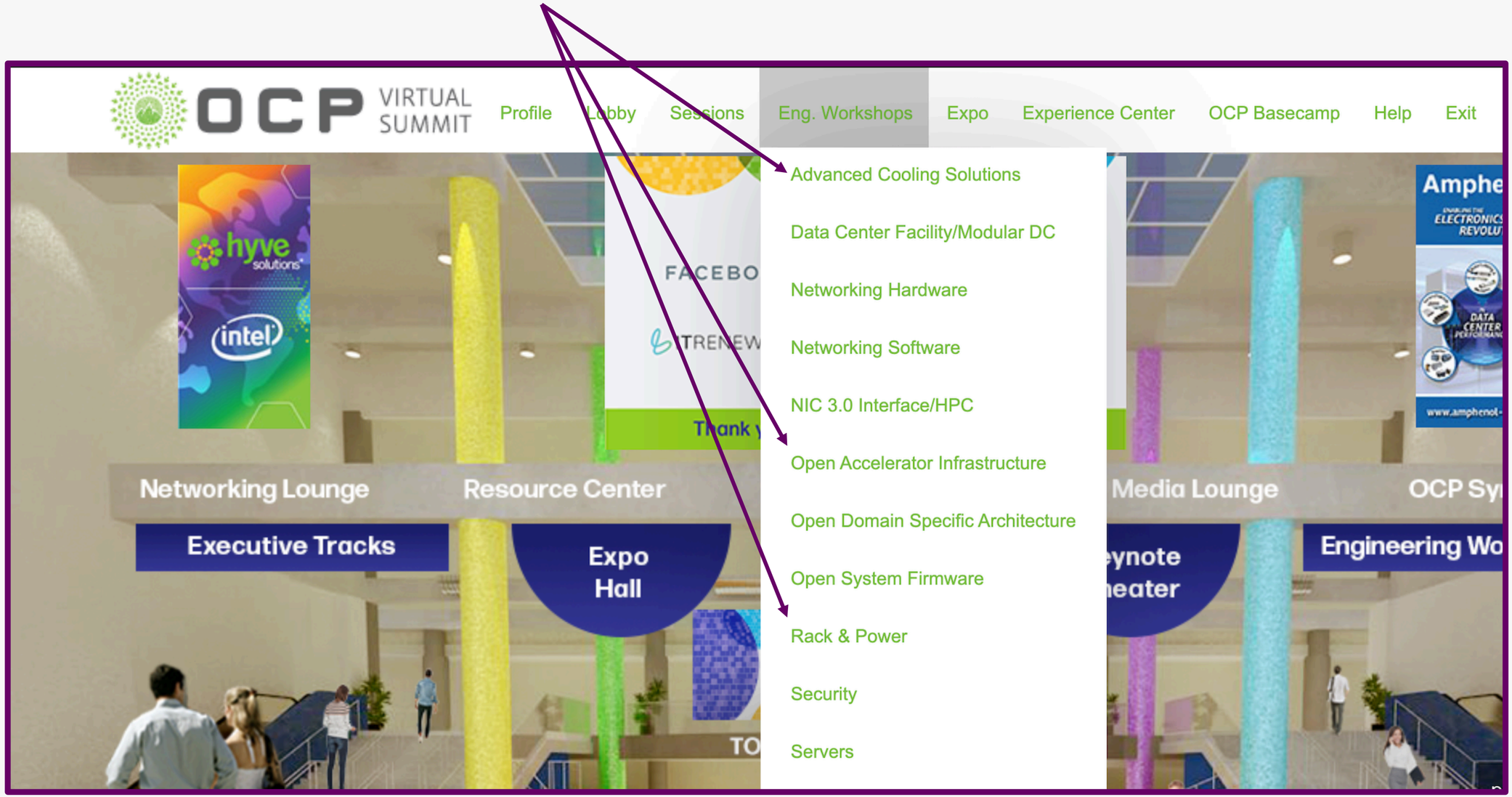
Open Cloud AI Rack

Customize your solution with a variety of compute nodes, JBOG, storage, FPGA, and networking components for applications ranging from high-density deep learning training to data mining and data analysis.

TensorFlow kubernetes openstack

https://www.opencompute.org/solutions

plugin



https://www.opencompute.org/summit/global-summit

plugin

How to get involved?

steve@opencompute.org

OPEN Compute Project ®

OCP Marketplace
https://www.opencompute.org/products

Past Events (recordings and slides)
https://www.opencompute.org/events/past-summits

https://www.opencompute.org/events/past-events

Social
https://www.youtube.com/user/OpenComputeProject

https://www.linkedin.com/groups/4152886/

@OpenComputePrj

https://www.facebook.com/groups/opencompute/

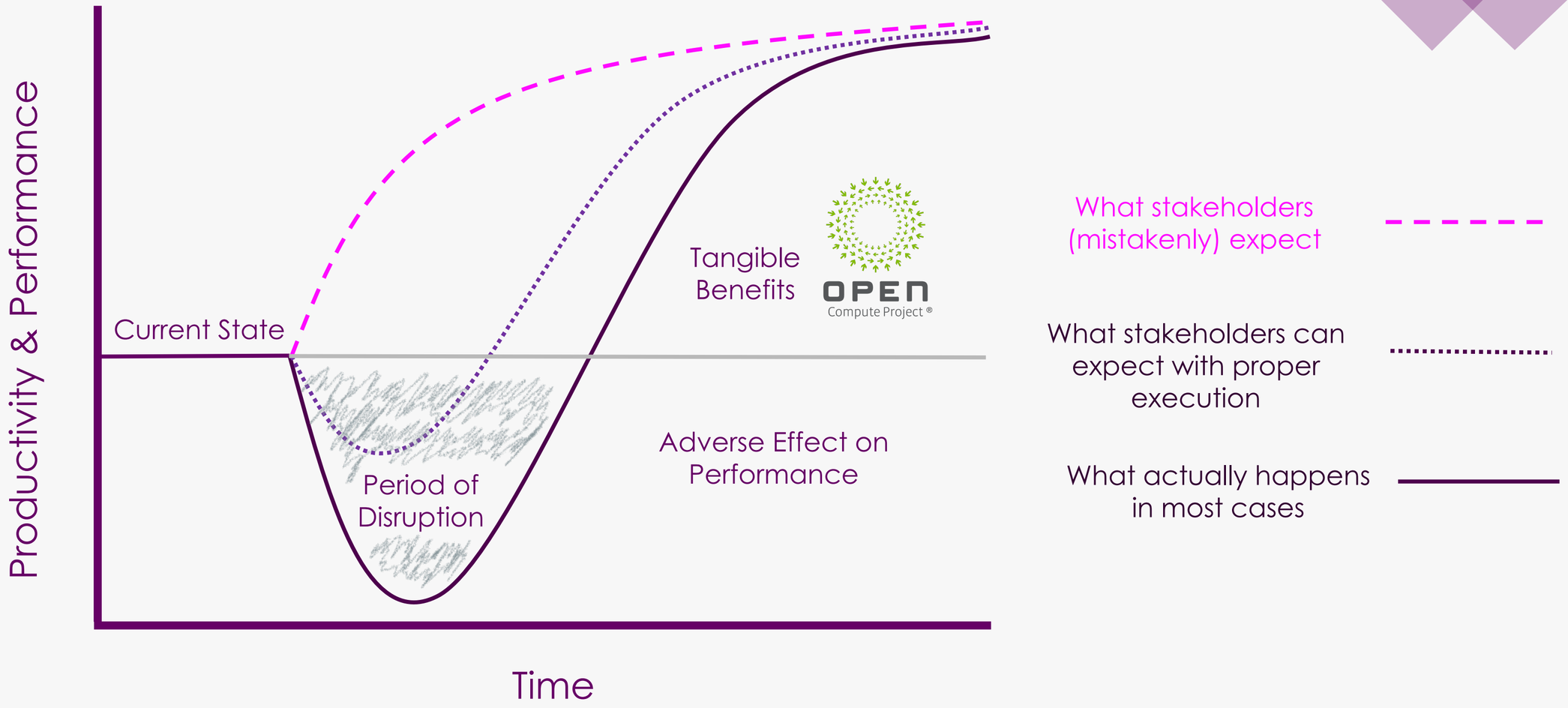
OCP Projects
https://www.opencompute.org/projects

OAI Project Mailing List
https://ocp-all.groups.io/g/OCP-OAI

OAI Project Wiki with latest specification
https://www.opencompute.org/wiki/Server/OAI

Advanced Cooling
https://www.opencompute.org/projects/acs-immersion-cooling-sub-project

The AI J-Curve



https://en.wikipedia.org/wiki/J_curve

Thank you

plugin

Questions?

plugin 2020