HPC & AI Infrastructure Convergence

Olivier Philippe
Chief R&D Officer
Oct 2026

© Bull SAS

Our heritage

Building trusted technology through years of innovation

From the start, Bull explored how computers could assist experts and solve complex problems through pioneering projects

1919



Fredrik Rosing Bull creates the first tabulating machine, the ancestor of supercomputers

1931

Creation of the Bull company

1933



Industrial production of tabulators as the Compagnie des Machines Bull

New Factory in Angers

1966



Launch of the "Plan Calcul" by the French State with the creation of CII, and new factory in Angers

1981

Nationalisation of CII-Honeywell- Bull. First AI department

New Factory (2x) in Angers

2026

Rebirth of Bull as a standalone company, now separated from the Atos Group, with the French State joining as a shareholder

2025



Launch of BullSequana XH3500, converged supercomputing platform for the AI-HPC era

2023

Carve out from Atos with the Eviden brand

2021

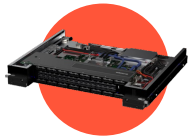
Acquisition of Nimbix strengthening HPC aaS and of DataSentics for AI/ML solutions

2017



Launch of the QLM first quantum simulator

2016



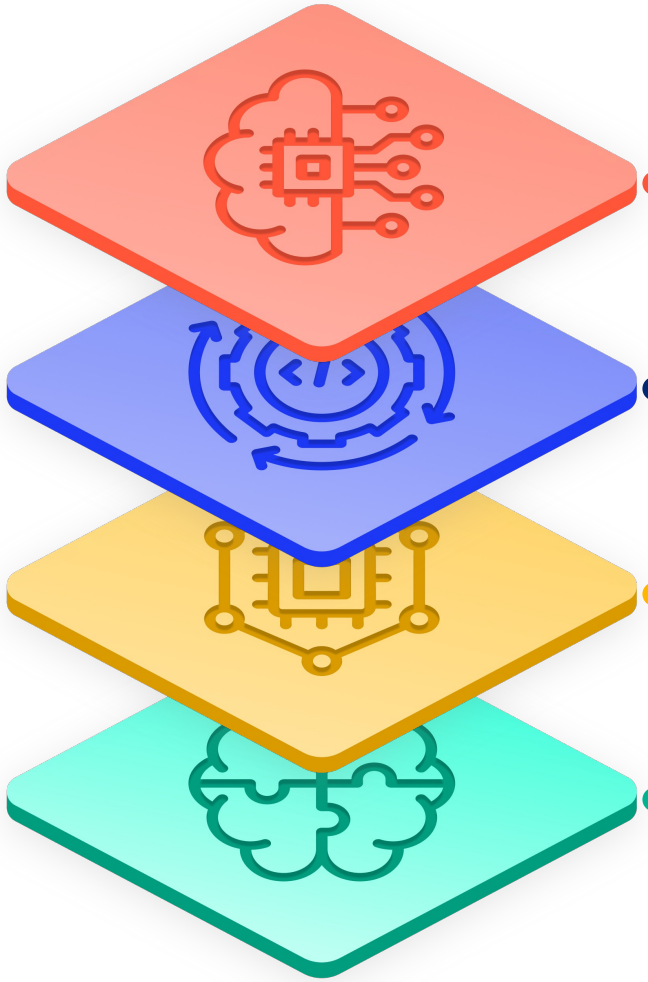
Launch of BXI, proprietary highspeed interconnect for HPC

2014

Acquisition of Bull by Atos

© Bull SAS

And it's no longer only about systems



AI use cases

Service, models and business applications where AI is applied to deliver value and business outcomes

Data + AI platforms

Software environments that develop, govern and operate AI solutions including model training and inference

Systems

Integrated technical stacks that combine hardware (chips, servers, networking), system software, data infrastructure into a fully operational environment

Components

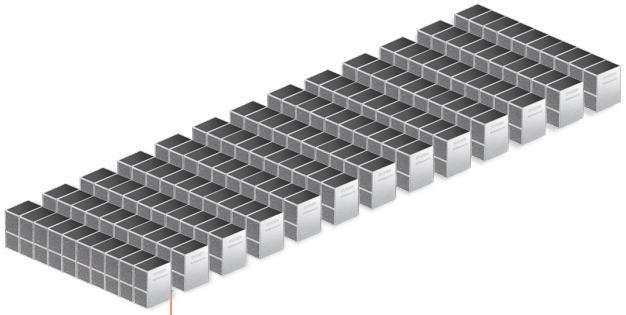
Core physical and technical building blocks, such as GPUs, semiconductors, memory, storage systems, and high-speed interconnect networks

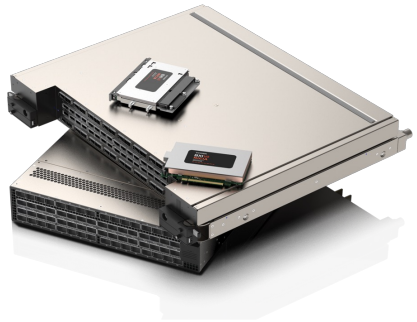
Bull

© Bull SAS

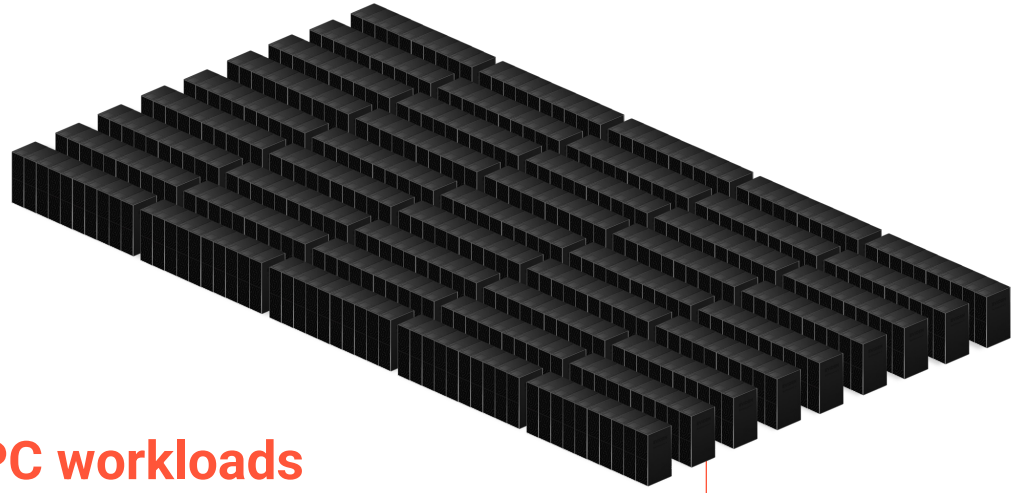
Bull eXascale Interconnect (BXI) unifies best of both worlds & accelerates HPC & AI workloads

An **Exascale HPC System**





Trillion Parameters AI Factory



BXI V3

SmartNIC acceleration for AI & HPC workloads

HPC 1 Quintillion FLOPS, 1 Trillion parameters AI
Dozens of thousands endpoints

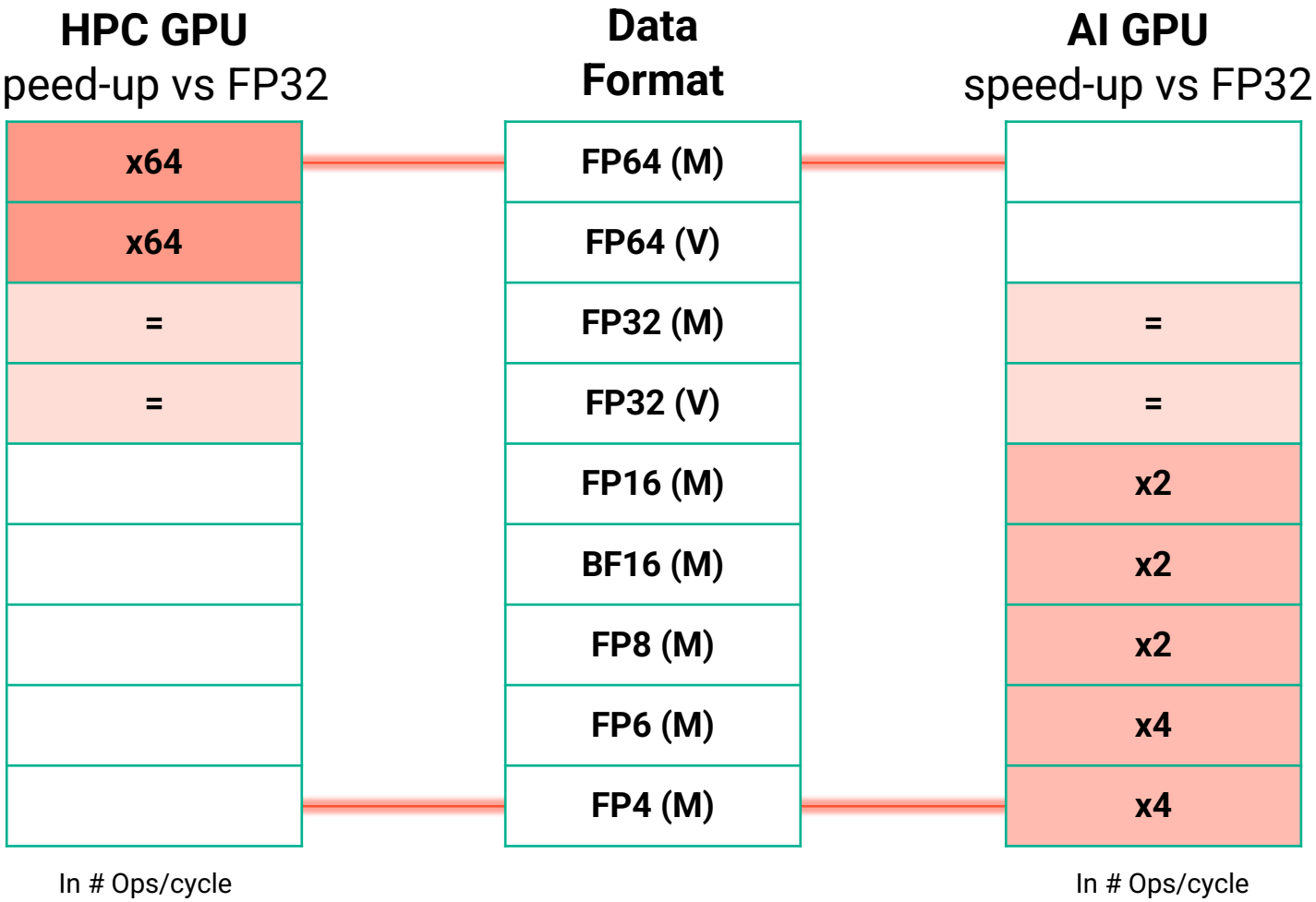
Accelerated MPI & NCCL, RCCL based Scientific & AI workloads
FatTree, DragonFly+ , Rail, … optimized topologies

© Bull SAS

HPC & AI : the same peak dense compute capacity, different precisions

Same architecture, same TDP, same HBM size, different focus

Scientific Computing



Generative AI

© Bull SAS

Summary of HPC & AI Infrastructure commonalities & differences

Common - mostly technology bricks and Data Center site implications:

- GPU & CPU platforms
  - ❑ Processing density per rack,
  - ❑ Power density per rack,
  - ❑ Weight density per rack,
  - ❑ Data Center energy,
  - ❑ Cooling needs.
- Scale-out network requirements & technologies

Different - mostly architectural at cluster level :

- Use Cases
  - ❑ Simulations vs Model inference/training,
  - ❑ Versatility of compute use cases,
- CPU / GPU mix
- Storage types and hierarchies
- Network topologies
- Scale-up requirements
- Software orchestration (e.g. SLURM vs Kubernetes)

© Bull SAS

Power and cooling evolutions at rack level

Now 20+ years of pioneering innovations at Bull, including DLC

Top 3 Systems in Green500 !



Novascale
1st supercomputer



Bullx B700 DLC
1st gen DLC



BullSequana X1000
2nd gen DLC



BullSequana XH2000
All-In-One
3rd gen DLC



BullSequana XH3000
All-In-One
4th gen DLC platform



BullSequana XH3500
All-In-One
5th gen DLC platform
Maximum Flexibility

NextGen

	2005	2012	2015	2018	2022	2026	2028-2029
	1100kg/m²	1700kg/m²	1900kg/m²	2200kg/m²	2600kg/m²	**2900kg/m²**	***>>3000kg/m²***
	20kW	70kW	84kW	87kW	147kW	**264kW**	***up to 800kW***
	100% air	70% DLC	85% DLC	**100% DLC**	**100% DLC**	**100% DLC**	***100% DLC***

Gen N+2 :

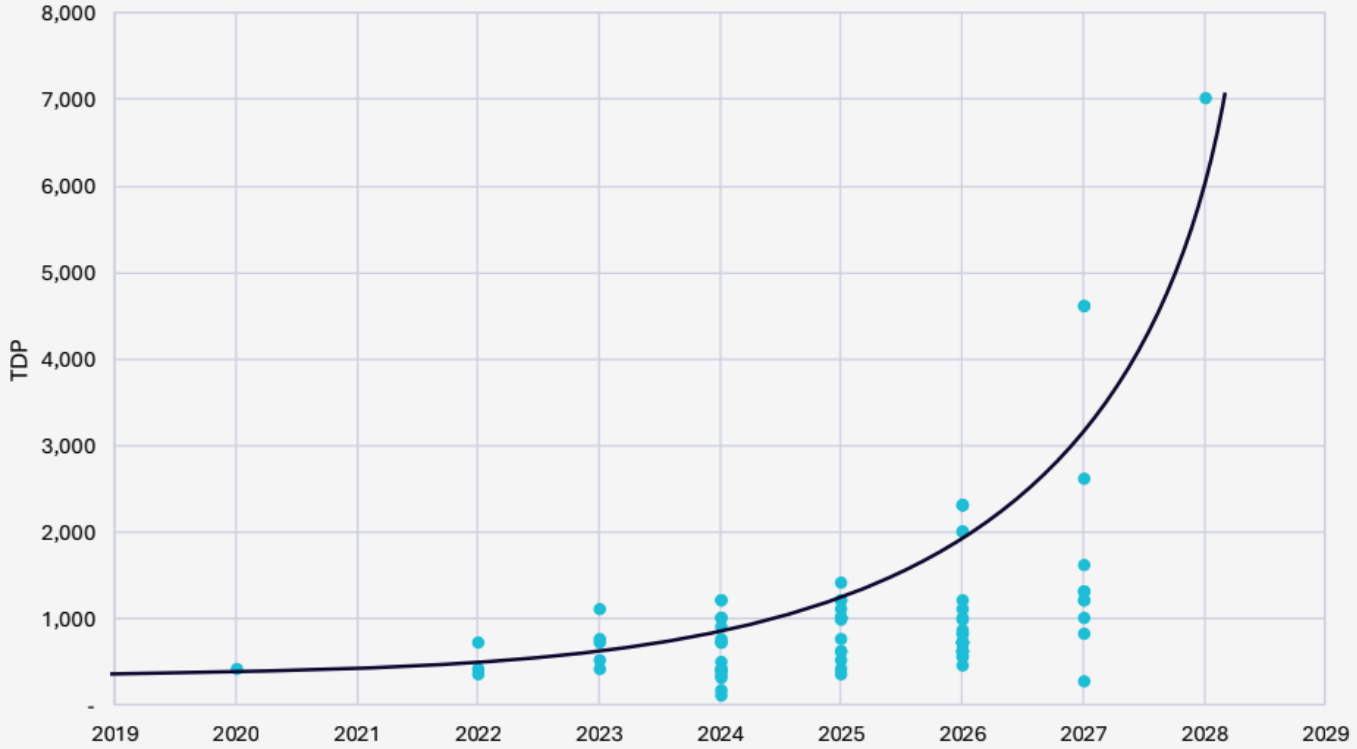
54V AC 1MW rack: 200 Kg copper -> 1 GW site = 500 T copper
800V DC 1MW rack: 13.5 Kg copper -> 1GW site = 34 T copper

GPU generations imply exponentially growing TDPs

Monophasic DLC remains possible, but still not a commodity

THERMAL DESIGN POWER (TDP)

TDP is increasing but remains within a reasonable range that a single phase cooling can manage.



Source: **SemiAnalysis & CoolIT**

CASE TEMPERATURE

- Processor case temperatures remain steady (80-90C°)
- Peripheral components are often more sensitive (<65C°).

HEAT FLUX

- Hotspot and average heat flux is rising steadily, typically between 100-200W/cm^2.
- Not rising at the same rate as TDP

© Bull SAS

HPC & AI convergence in Use Cases

Use Cases are rapidly converging, but…

Convergence categories

- Model-aware and workflow-centric simulation,
- AI-augmented digital twins,
- Scientific foundation models (LLM or WM ?),
- Surrogate AI models

Example 1: Augmented simulation.
An AI model trained on past runs—allows for the exploration of a vastly greater number of scenarios before selecting and launching full-scale, computationally expensive simulations.
→ AI makes simulations/calculations more efficient and faster.

Example 2: Surrogate model
In some cases, an AI model can replace a traditional algorithm.
→ simulations neededto train the model and improve it.

So, what ?

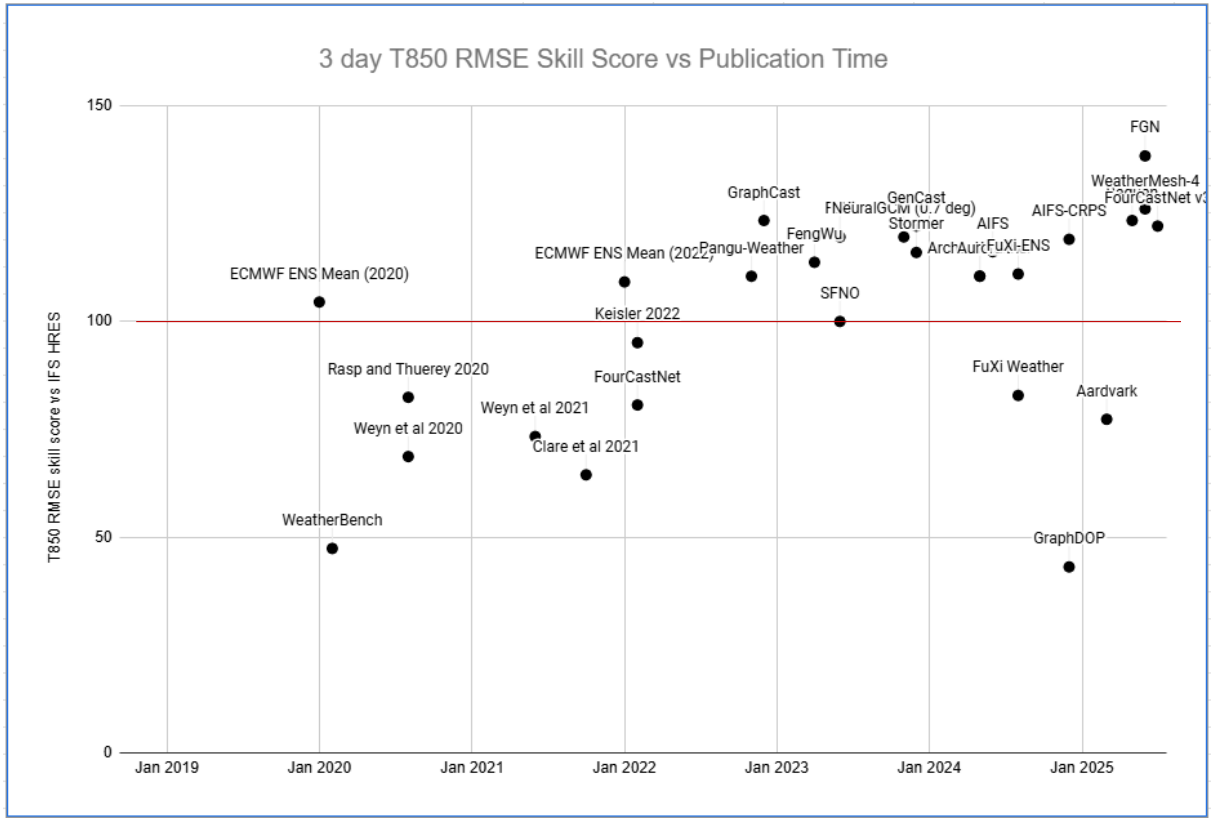
- **It is a symbiosis, not a replacement**
  - simulations feed the model,
  - expert systems needed to harness pure AI models.
- What is the optimal infrastructure for this mix of HPC and AI workloads?
- **Many moving targets** : power, cooling, density placement, xPU mix, memory & storage placements…
- **Flexibility matters** as well as performance & efficiency.
  A fixed AI+HPC combination can end up being a mistake worth 100s m€.

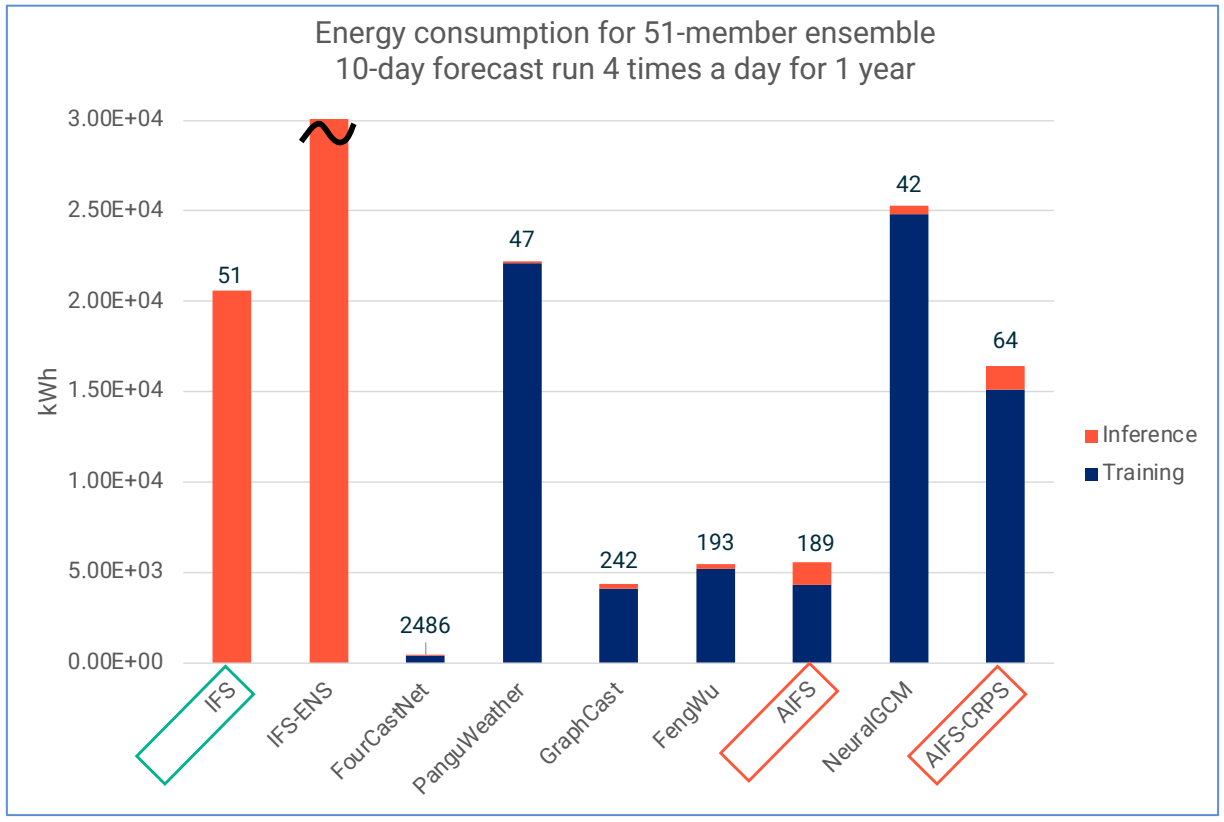
Bull

© Bull SAS

Example : AI for Weather Forecast Landscape

AI weather forecasting models are progressing steadily, especially model-aware and workflow-centric.
Low AI inference costs are promising, but models remain dependant on simulation data and training





Rasp, Stephan (2024). AI-Weather SotA vs Time. figshare. Dataset. https://doi.org/10.6084/m9.figshare.28083515.v1

Source: Rieutord et al. (2026)

© Bull SAS

What about standardisation ?

Progressing, but a complex matrix and still in storming & forming phase. It's a journey.

Many stakes at play at rack level

- Multiple xPU vendors sometimes with their own standards
  - Multiple processing density strategies
  - Multiple power supply strategies
  - Multiple reference architectures
  - Multiple mechanical formats
- Multiple xPU and networking ASIC generations
  - All evolving physical constraints impacting components and system architecture levels, hence standards.

x

A complex standardisation journey

- OCP leading multiple efforts
  - OCP OpenRack (ORv3 today),
  - OAM, OAC (in development)
  - Reference Architectures (topologies),
  - Open DC (Energy and Grid),
  - Open-pods (Inference, OPG-M),
  - Open-clusters (Training, XOC-N)
- Scale-out
  - Ultra Ethernet vs proprietary
- Scale-up
  - E-SUN/Ultra Ethernet vs UALink vs proprietary

© Bull SAS

Conclusions

- HPC & AI infrastructure technical constraints are largely the same
- Innovation pace accelerates (HPC & /AI) : 1 new GPU per year, 1 new switch ASIC per year, racks per 2 years
- Compute density accelerates, but multiple component vendor strategies exist :
  - Density spread across chip / rack / site levels differ
  - Mix of GPUs, CPUs, TPUs/NPUs,
  - QPUs
- Major evolutions also accelerate : power, cooling, fabric, memory & storage hierarchy
- Mix of specific use cases + AI training & inference use cases, but simulation data remains needed
- Standardisation progresses but will remain immature for a while and is a moving target

> **Multi-vendor, multi-architecture HPC & AI systems and tailor-made solutions are required**

© Bull SAS



Confidential information owned by Bull SAS, to be used by the recipient only. This document, or any part of it, may not be reproduced, copied, circulated and/or distributed nor quoted without prior written approval from Bull SAS.

© Bull SAS - Confidential