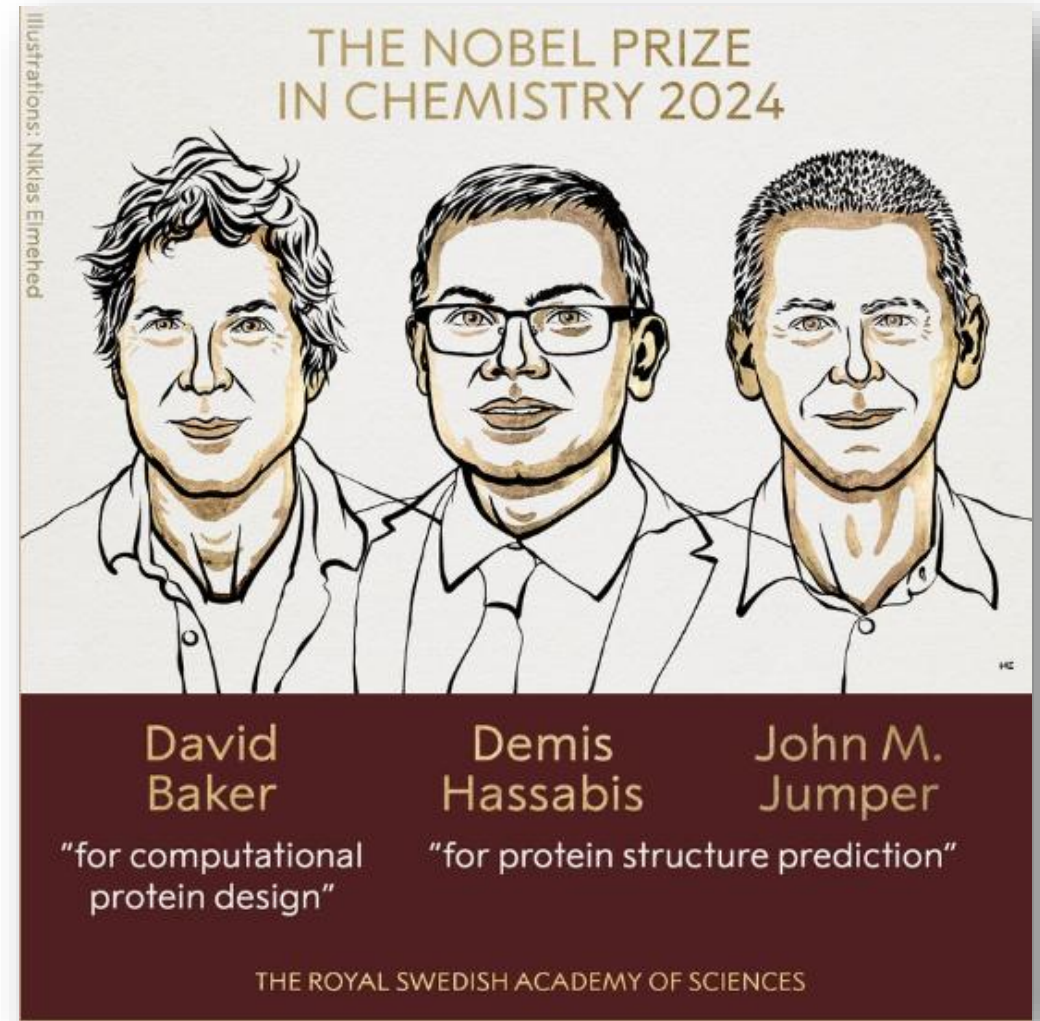
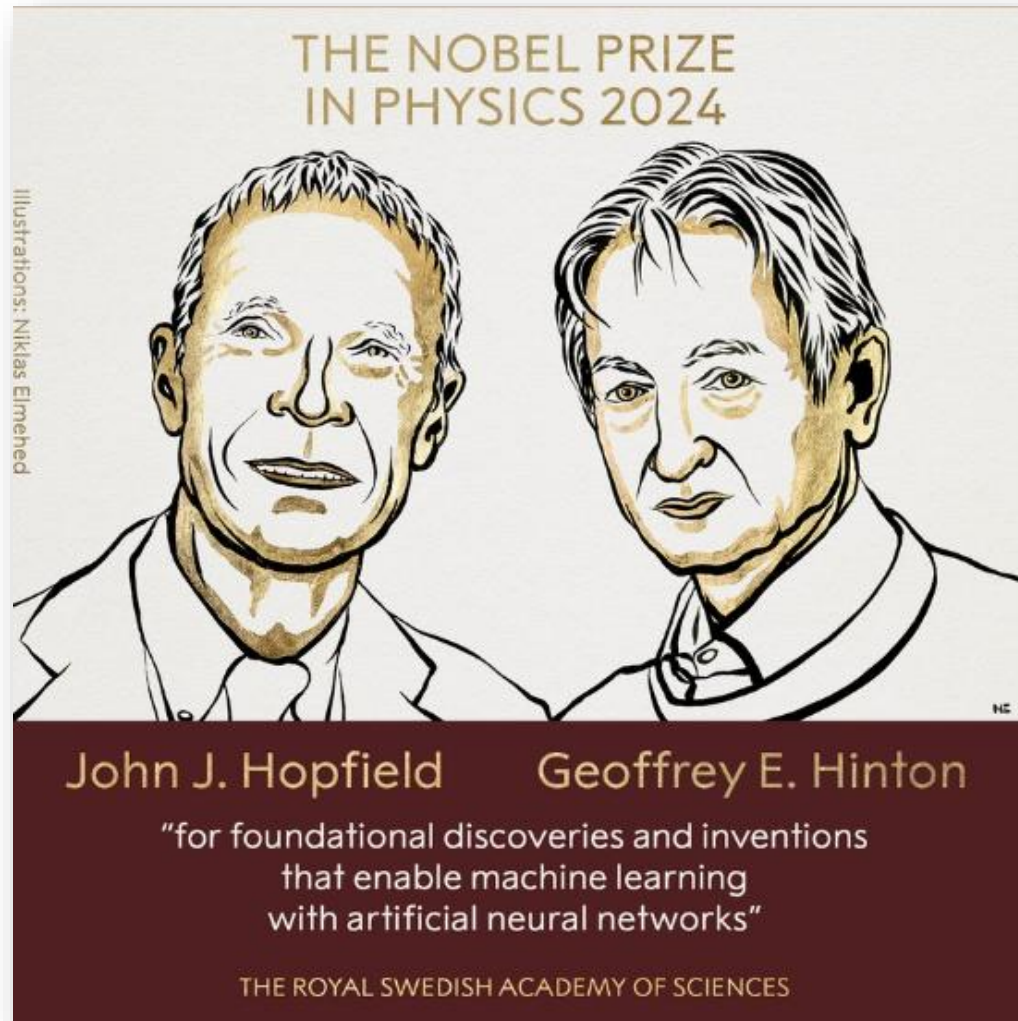


MAVERIC: Enabling the Future of AI-Driven Research in Australia

Amr Hassan, Simon Forsayeth, Joseph Pineda, Gin Tan, David Powell



AI as a driving force behind scientific discovery



From Algorithms to Infrastructure: What Powers AI Today?

The success of modern AI is as much about scaling hardware and datasets as it is about smart models.

Computation used to train notable AI systems, by affiliation of researchers

Our World in Data

Computation is measured in total petaFLOP, which is 10^{15} floating-point operations¹ estimated from AI literature, albeit with some uncertainty. Estimates are expected to be accurate within a factor of 2, or a factor of 5 for recent undisclosed models like GPT-4.

Training computation (petaFLOP)



Data source: Epoch (2025)

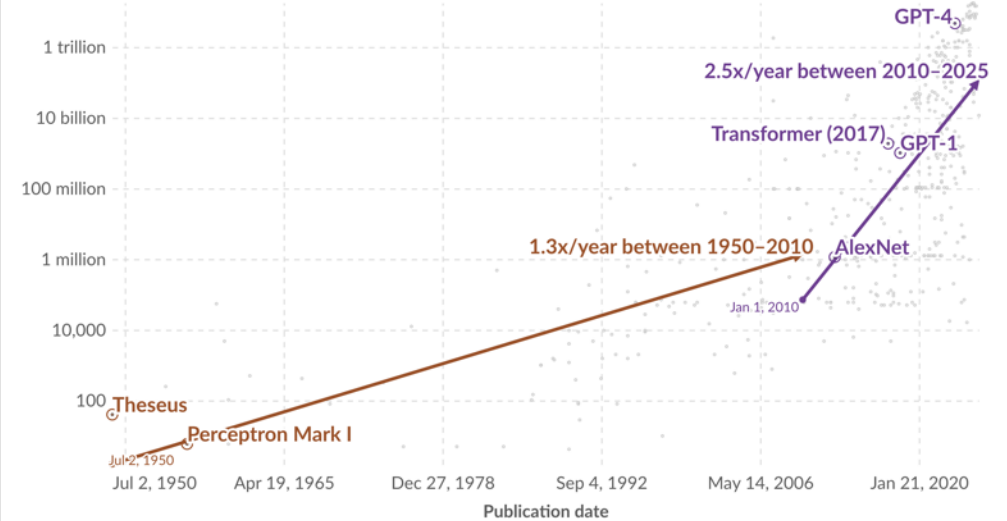
OurWorldinData.org/artificial-intelligence | CC BY

Exponential growth of datapoints used to train notable AI systems

Our World in Data

Each domain has a specific data point unit; for example, for vision it is images, for language it is words, and for games it is timesteps. This means systems can only be compared directly within the same domain.

Training datapoints (datapoints)



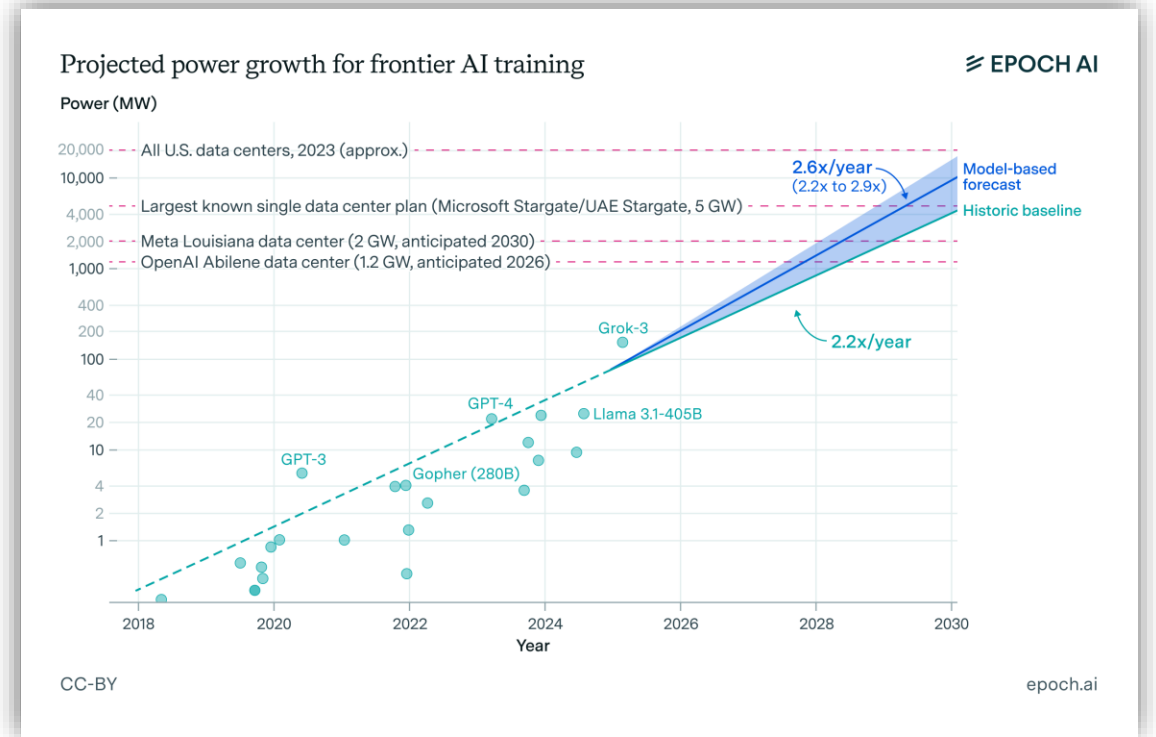
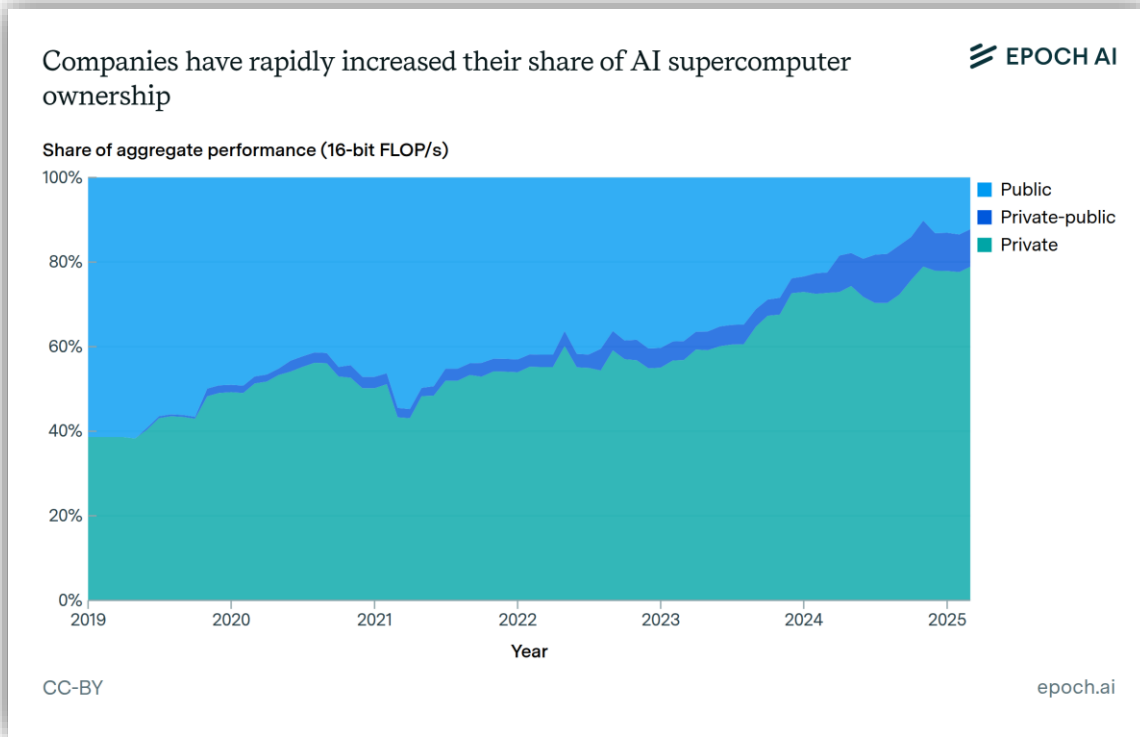
Data source: Epoch (2025)

OurWorldinData.org/artificial-intelligence | CC BY

Note: The regression lines show a sharp rise in data used to train AI systems since 2010, driven by the success of deep learning methods that leverage neural networks and massive datasets.

From Algorithms to Infrastructure: What Powers AI Today?

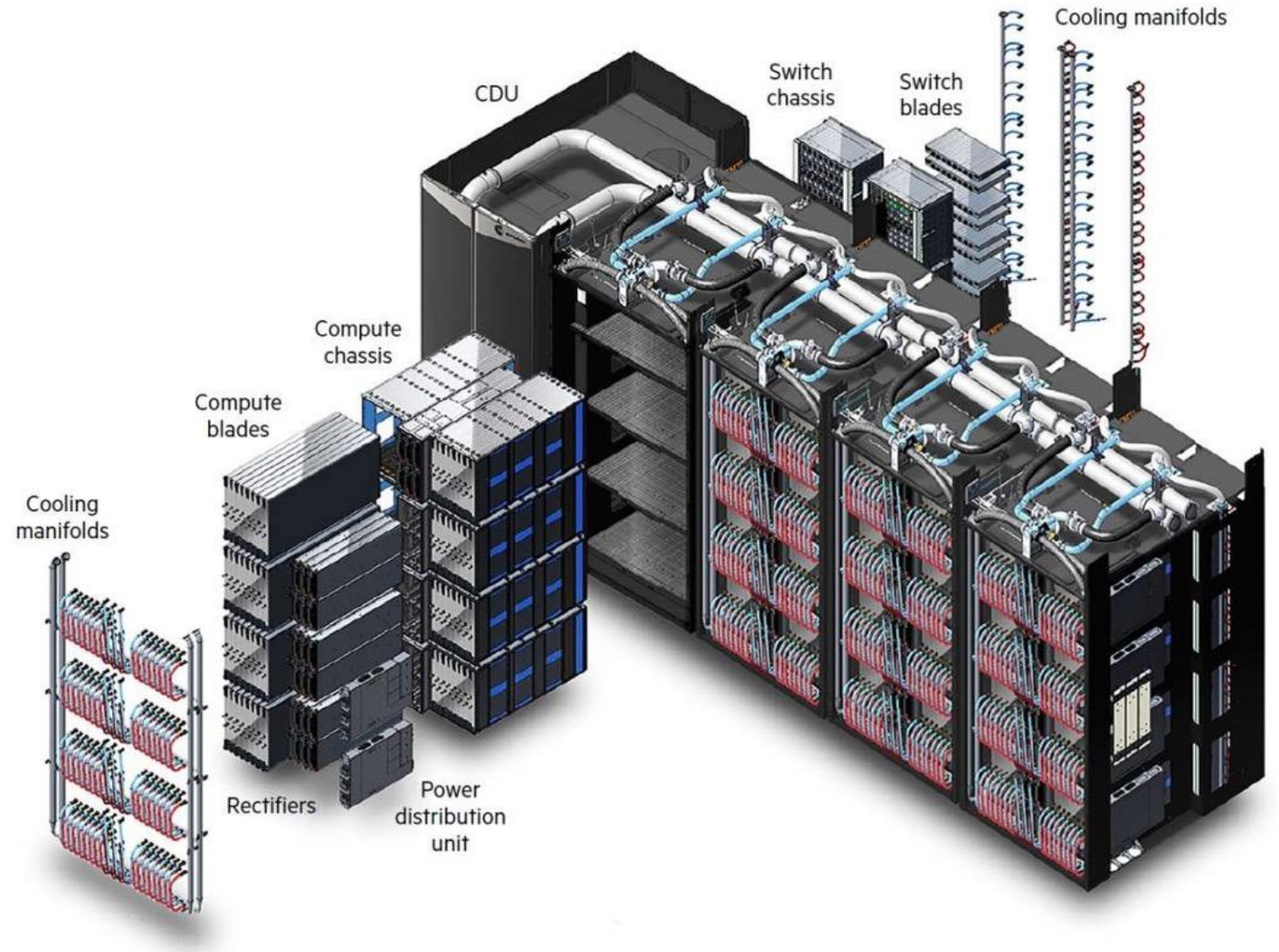
The success of modern AI is as much about scaling hardware and datasets as it is about smart models.



As of May 2025, the largest known public AI supercomputer, Lawrence Livermore’s El Capitan, achieves less than a quarter of the computational performance of the largest known industry AI supercomputer, xAI’s Colossus.

Power demands for frontier training runs have historically grown at a rate of 2.2x per year, with the largest runs now exceeding 100 MW.

MAVERIC Hardware: The Concept of AI Supercomputers



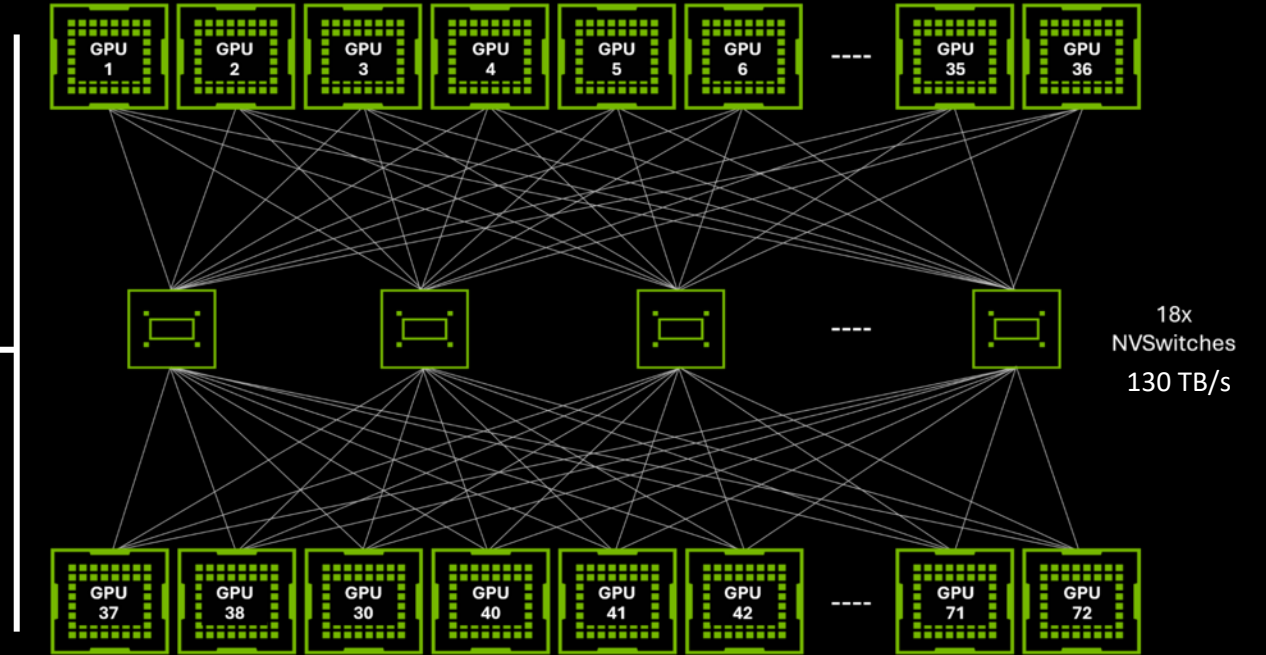
Rack-Scale Architecture

NVIDIA GB200 NVL72



NVIDIA GB200 NVL72

- 36 GRACE CPU
- 72 BLACKWELL GPUs
- 13.4 TB GPU Memory
- 130 TB/sec
- 360 PFLOPS (FP16)



LLM Inference

30X vs. H100

LLM Training

4X vs. H100

Energy Efficiency

25X vs. H100

Rack-Scale Architecture

NVIDIA GB200 NVL72



GB200 Compute Tray

- 4 x B200 GPUs
- 2 x 72 Core CPU
- 192 GB Memory per GPU
- 768 GB GPU Memory
- 960 GB CPU Memory
- 4 x 400 Gb/s InfiniBand
- 4 x 200 Gb/s Ethernet



NVLink Switch Tray

- 1.8 TB/s GPU-to-GPU
- 144 Ports with 14.4 TB Switching Capacity

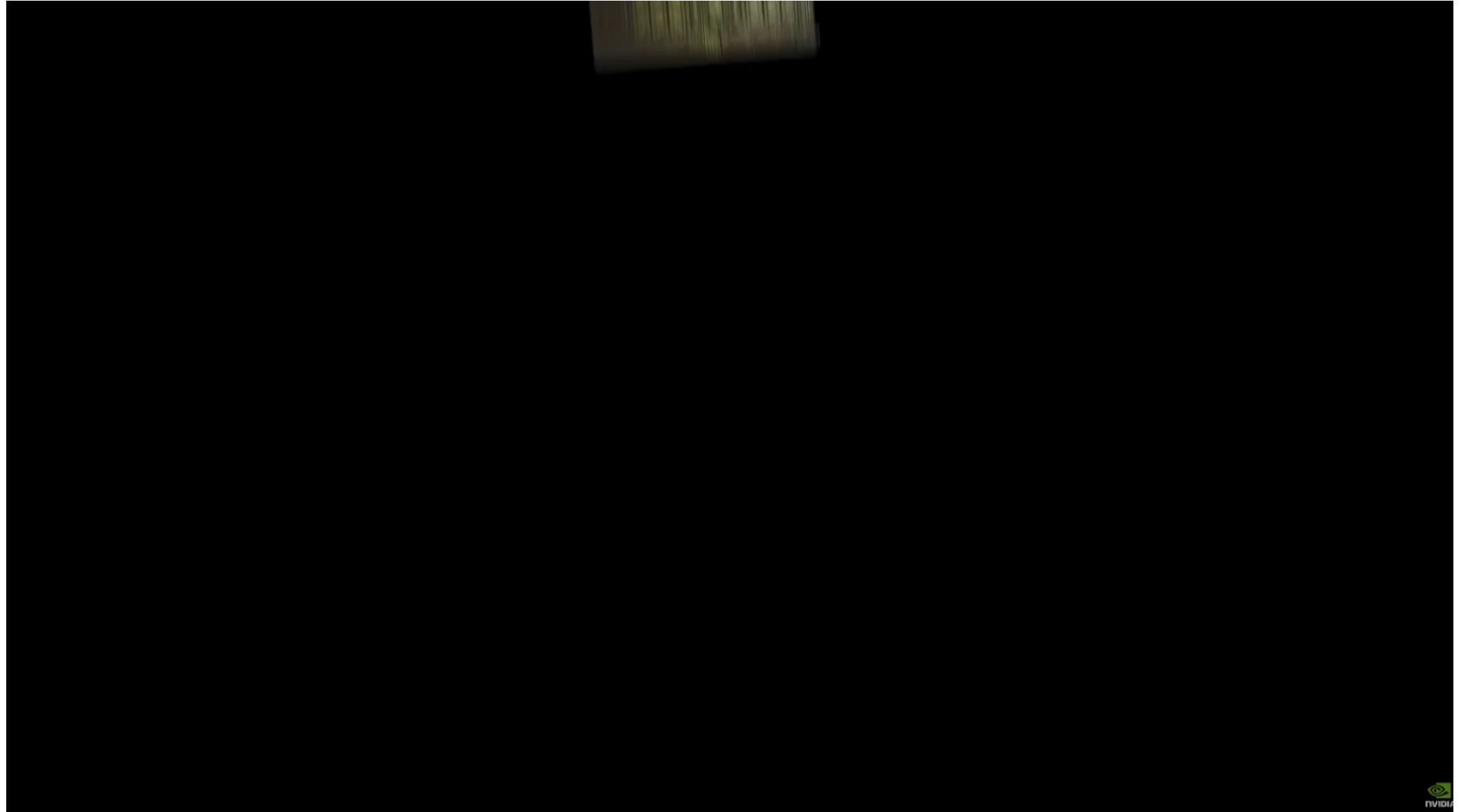
Rack-Scale Architecture

NVIDIA GB200 NVL72

	GB200 NVL72	GB200 Grace Blackwell Superchip	H200 NVL
Configuration	36 Grace CPU : 72 Blackwell GPUs	1 Grace CPU : 2 Blackwell GPU	Hopper GPU
FP4 Tensor Core	1,440 PFLOPS	40 PFLOPS	
FP8/FP6 Tensor Core	720 PFLOPS	20 PFLOPS	3.3 PFLOPS
INT8 Tensor Core	720 POPS	20 POPS	3.3 POPS
FP16/BF16 Tensor Core	360 PFLOPS	10 PFLOPS	1.67 PFLOPS
TF32 Tensor Core	180 PFLOPS	5 PFLOPS	0.84 PFLOPS
FP32	5,760 TFLOPS	160 TFLOPS	60 TFLOPS
FP64	2,880 TFLOPS	80 TFLOPS	30 TFLOPS
FP64 Tensor Core	2,880 TFLOPS	80 TFLOPS	60 TFLOPS
GPU Memory Bandwidth	Up to 13.4 TB HBM3e 576 TB/s	Up to 372 GB HBM3e 16 TB/s	141 GB 4.8 TB/s
NVLink Bandwidth	130 TB/s	3.6 TB/s	900 GB/s GPU-to-GPU
CPU Core Count	2,592 Arm Neoverse V2 cores	72 Arm Neoverse V2 cores	
CPU Memory Bandwidth	Up to 17 TB LPDDR5X Up to 18.4 TB/s	Up to 480 GB LPDDR5X Up to 512 GB/s	

Rack-Scale Architecture

NVIDIA GB200 NVL72

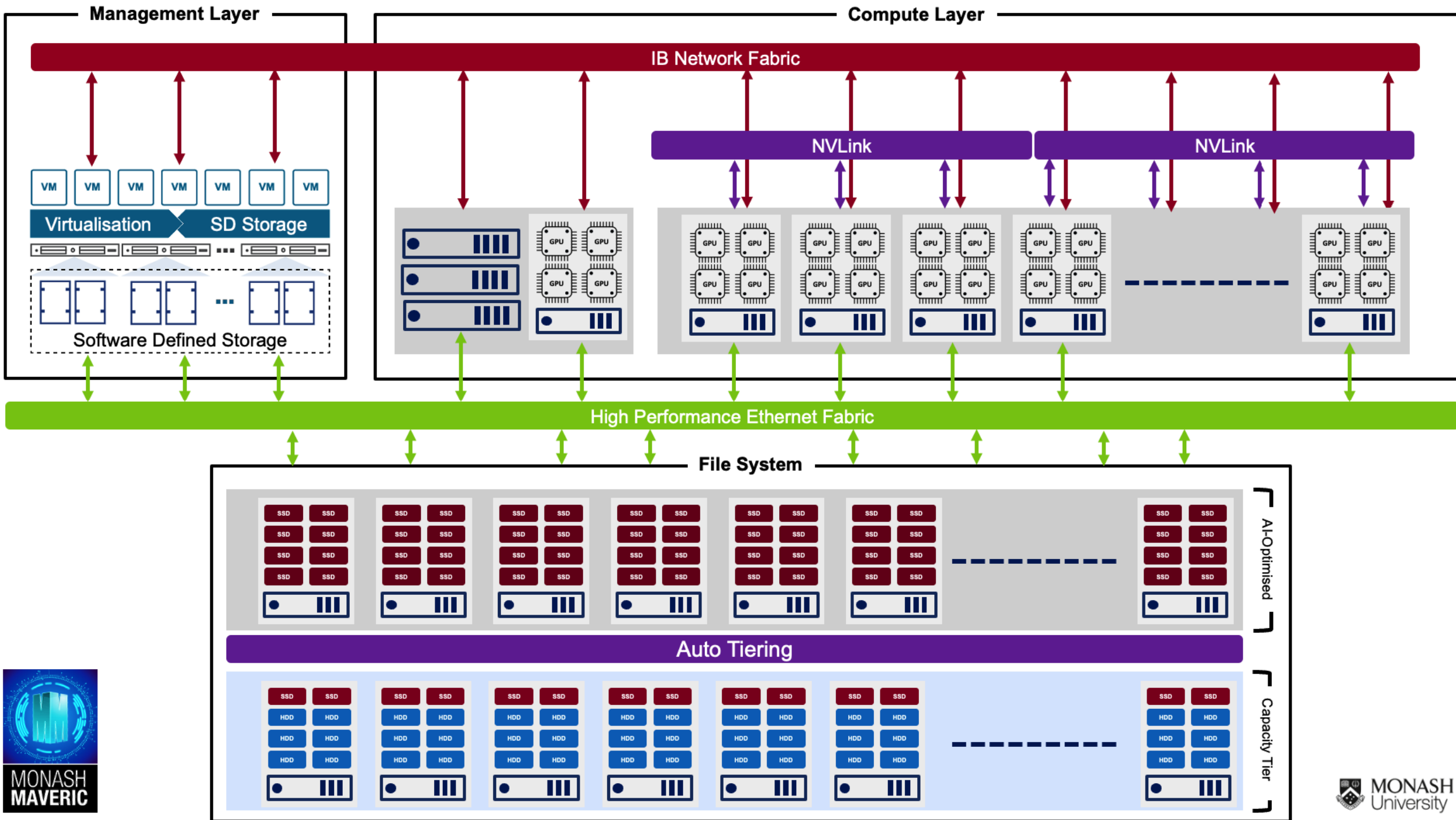


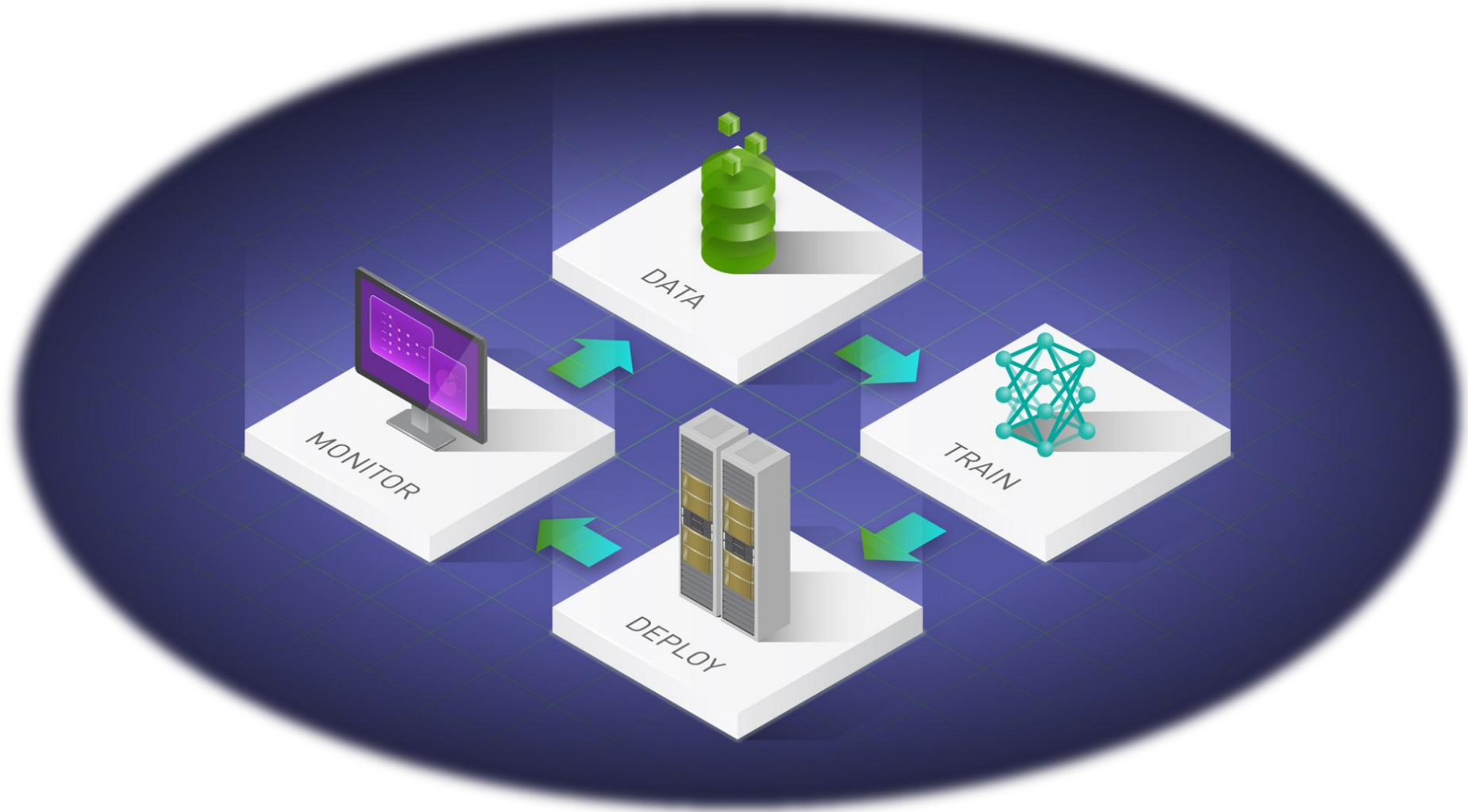
Rack-Scale Architecture

NVIDIA GB200 Technical Challenge

- ❑ Single Massive GPU
- ❑ Power Usage: 132-195 KW
- ❑ Weight: 1.3T
- ❑ 80% DLC
- ❑ Rack density:
 - ❑ 1.2m Components
 - ❑ ~ 3km copper cables.
- ❑ 3 Different Network Layers

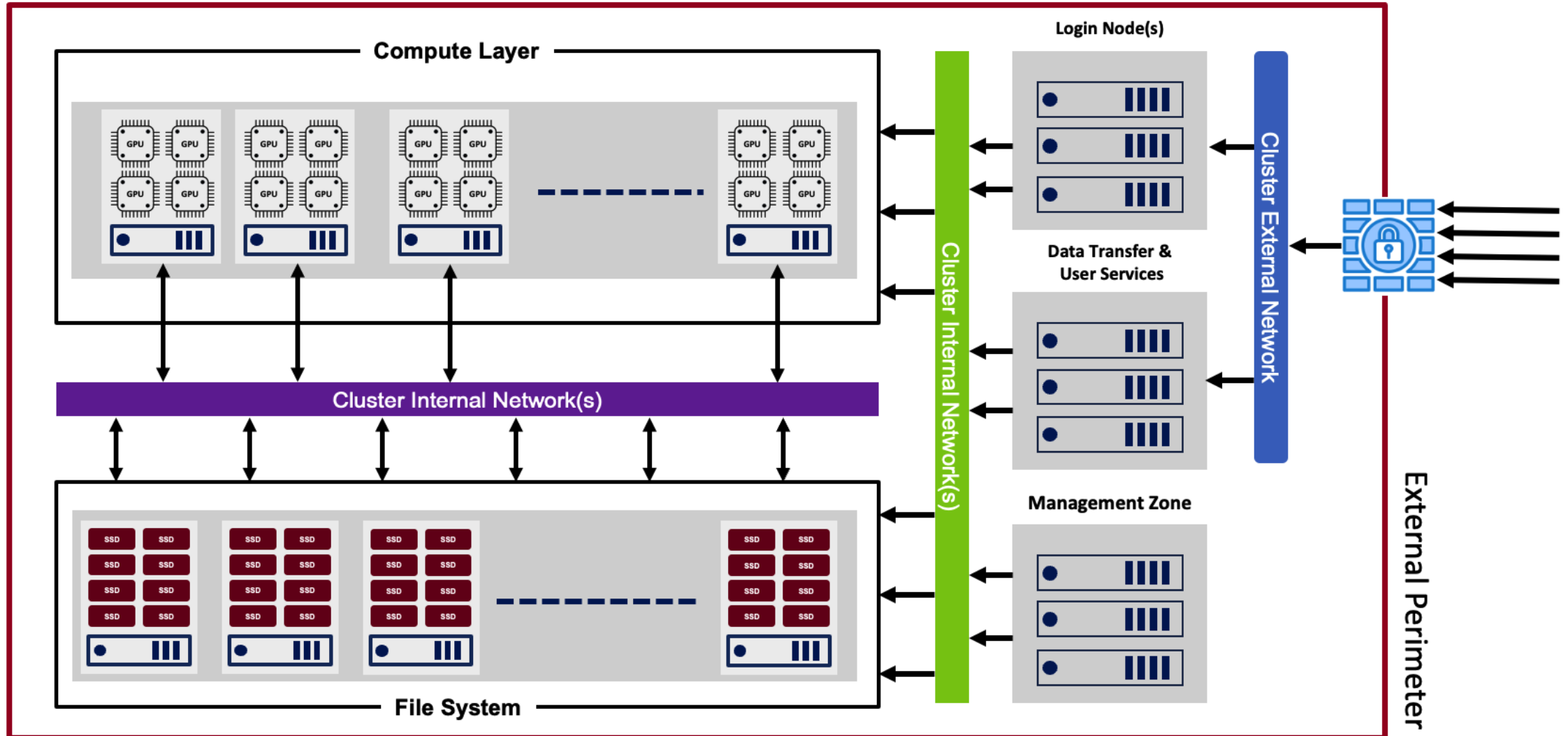




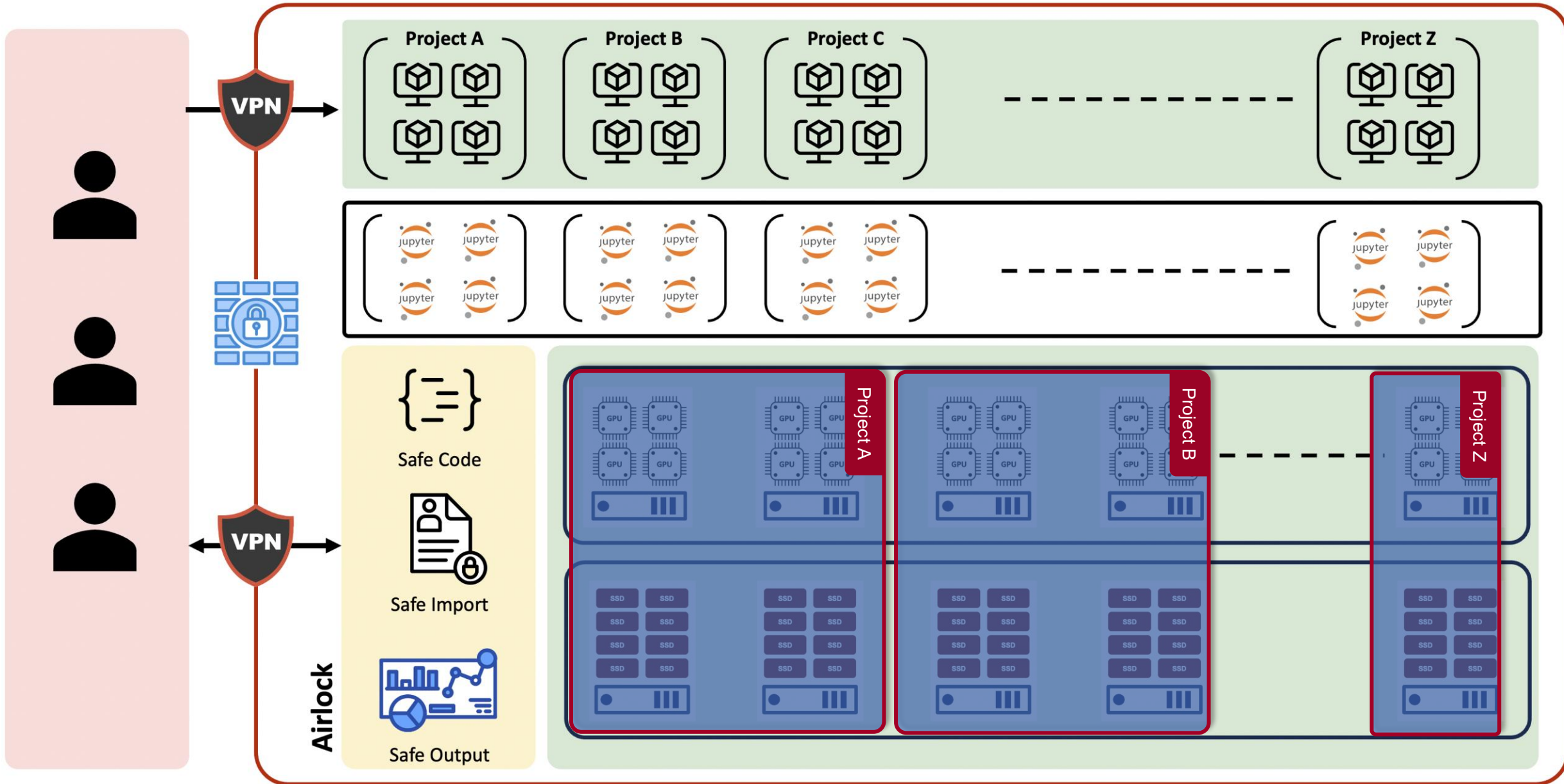


AI Software Stack for Very Sensitive Data

HPC System Reference Architecture (NIST SP 800-234)



MAVERIC Trusted Research Environment



What MAVERIC will introduce to our Ecosystem?



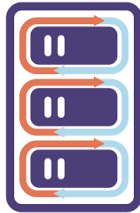
ISO 27001
Certified Infrastructure



ITIL Compliant
Service Management



AI Optimised Storage with
Automated Tiering



Monash-First Direct Liquid
Cooling Infrastructure



Multi-Tenancy with full Project
isolation



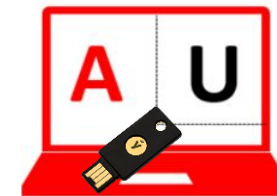
Storage Isolation



Researcher-Managed Software
Stack with Containers



Managed Data Egress and Ingress
with approval workflow



Privileged Access Workstation
(Trusted Administration Layer)



Contact us:
maveric@monash.edu

