aiDAPTIV+

Scaling Boundaries, Not Budgets

Unlock Large Model Training

Until aiDAPTIV+, small and medium-sized businesses have been limited to small, imprecise training models lacking the ability to scale beyond Llama-2 13B. Phison's aiDAPTIV+ solution enables significantly larger training models, giving you the opportunity to run workloads previously reserved for data-centers.

Hybrid Solution Boosts LLM Training Efficiency

Phison's aiDAPTIV+ is a hybrid software / hardware solution for today's biggest challenges in LLM training. A single local workstation or server from one of our partners provides a cost-effective approach to LLM training, up to Llama-3 70B and Falcon 180B.

Product Features

Ease of Use and Deployment

Spend your time training your data, not yourself or your engineers to set up and familiarize with an aiDAPTIV+ system.

Cost and Accessibility

aiDAPTIV+ leverages cost-effective NAND flash to increase access to large-language model (LLM) training with commodity workstation hardware.

Security and Data Privacy

aiDAPTIV+ workstations will help you retain control of your data and keep it on premises.

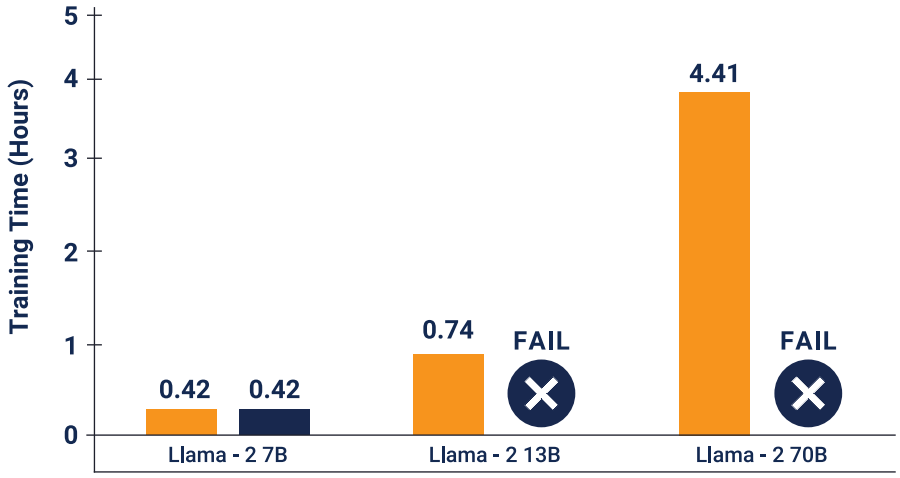
Supported Models

- Llama, Llama-2, Llama-3, CodeLlama
- Vicuna, Falcon, Whisper, Clip Large
- Metaformer, Resnet, Deit base, Mistral, TAIDE
- And Many More Coming Soon

PHISON

Workstation

Single node 4x GPU configuration comparing GPU and GPU +



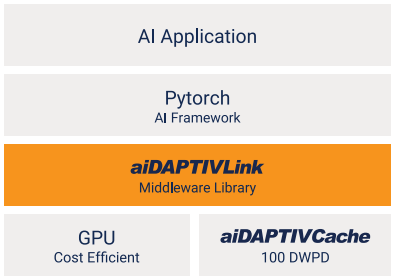
Model Size (Training)	140 GB	260 GB	1400 GB
HBM Pool (Usage%)	192 GB (73%)	192 GB (120%)	192 GB (729%)
Minimum GPU Count	4 / 4	4 / 6	4 / 30

System Configuration

- RAM: 512 GB
- GPU: 4x RTX 6000 ada
- GDDR: 192 GB

Note:
Scaling is linear based on GPU count and model size

aiDAPTIVLink



Drop-in solution for PyTorch

Experience seamless integration with the benefits offered by this system. It features a transparent drop-in function that eliminates the need to modify your AI application. You can effortlessly reuse existing hardware or add nodes as needed. System integrators have access to AI100E SSD, middleware library licenses, and full Phison support to facilitate smooth system integration.

Seamless Integration with GPU Memory

The optimized middleware extends GPU memory capacity by utilizing 2x 2TB aiDAPTIVCache to support a 70B model with low latency. Additionally, the high endurance feature offers an industry-leading 100 DWPD over 3 years, utilizing SLC NAND with an advanced NAND correction algorithm.

aiDAPTIVCache Family

AI100E M.2 SSD

PHISON

Find more information and resources at: phisonenterprise.com