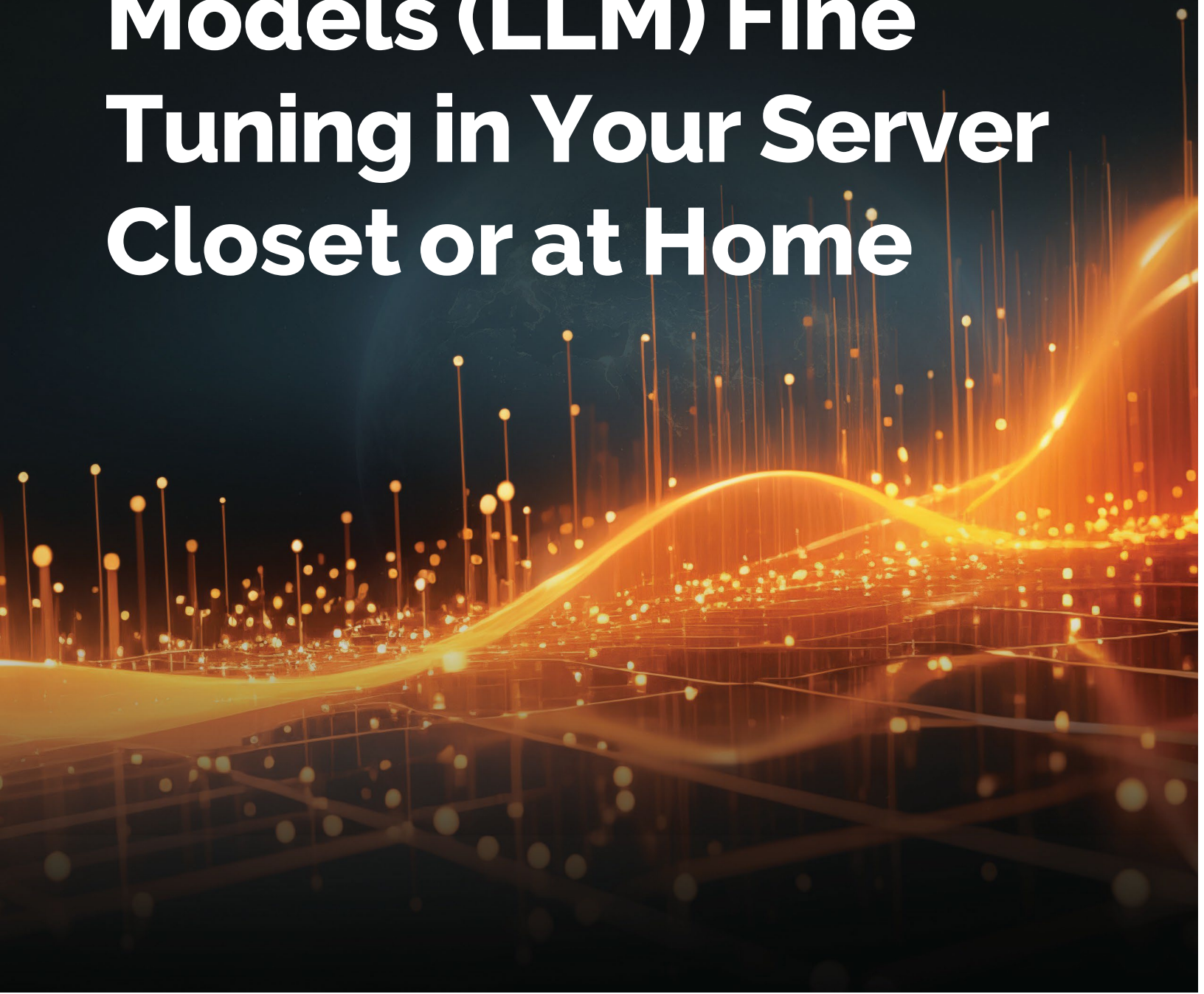


PHISON

**Easy, Cost-Effective
Large Language
Models (LLM) Fine
Tuning in Your Server
Closet or at Home**



Introduction

Small and Medium Businesses (SMBs) face significant challenges when it comes to domain training and fine-tuning Large Language Models (LLMs). The high cost of cloud services, the complexity of setting up AI infrastructure, the necessity for specialized staff training, the power requirements, and the increasing demand for faster, more efficient AI model training solutions are pressing issues. Data privacy, retention, and control from the cloud brings additional uncertainty and stress to the decision-making process. These challenges often put SMBs at a disadvantage, limiting their ability to leverage advanced AI technologies for competitive advantage.

How aiDAPTIV+ addresses the difficulties for SMB fine-tuning LLMs

aiDAPTIV+ emerges as a game-changer for SMBs looking to harness the power of LLMs without the associated hefty price tag and complexity. Designed with cost efficiency, plug-and-play style ease of use, and low power consumption in mind, [aiDAPTIV+](#) empowers businesses to train larger models on workstation-class systems effectively. By eliminating the need for extensive infrastructure or specialized training, SMBs can now focus on innovating and developing AI-driven solutions that were previously out of reach.

How aiDAPTIV+ works

At the core of aiDAPTIV+ is its revolutionary aiDAPTIVCache Family of extreme-endurance SSDs. This ai100 SSD facilitates NVMe offload, allowing for the efficient training of larger models with just 2 to 4 workstation-class GPUs and standard DRAM. This is complemented by aiDAPTIV management software that optimizes data flow between storage and compute resources, significantly enhancing performance, reducing training times, and enabling large-scale model training.

Standardized performance

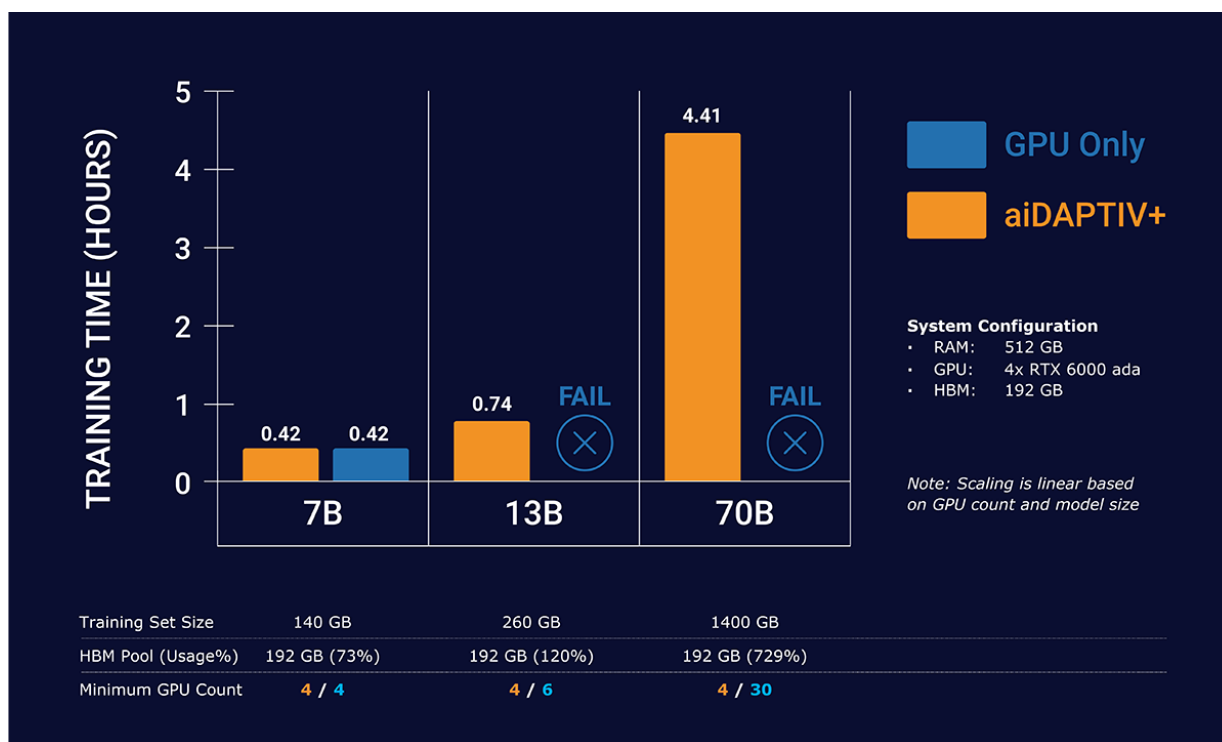
“The one single important event last year, and how it has activated AI researchers here in this region, is actually Llama-2. It’s an open-source model.”

Jensen Huang, President and CEO of NVIDIA

The performance of aiDAPTIV+ with Llama-2 models is a testament to its efficiency and effectiveness. Training times for a single epoch have been significantly reduced across various model sizes.

aiDAPTIV+ trains larger models with linear scaling

Single node 4x GPU configuration comparing GPU and GPU + aiDAPTIV+



These figures [showcase aiDAPTIV+'s current capability](#)* to handle substantial computational tasks swiftly, making it an invaluable asset for SMBs looking to fine-tune LLMs efficiently and affordably.

In the chart on the preceding page, you see that GPUs alone are not able to complete the Llama-2 13b and 70b training. The dataset is too large to fit into the GPU memory alone. This causes a fault that crashes the training. aiDAPTIV+ manages the data in slices to feed the GPUs in manageable pieces, then recompiles the data on high-speed flash to finish the model.

*Test system specifications

Cloud vs. aiDAPTIV+ technology comparison

When compared to traditional cloud-based AI training solutions, [aiDAPTIV+](#) stands out in several key areas:

- **Cost:** aiDAPTIV+ is significantly more cost-effective, eliminating the high expenses associated with cloud services.
- **Data Privacy:** By enabling training on-premises, aiDAPTIV+ ensures that sensitive data remains within the secure confines of a company's own infrastructure, addressing critical data privacy concerns.
- **Ownership:** aiDAPTIV+ offers full control and flexibility, with direct hardware access for customization and upgrades, reducing dependence on third-party services and enhancing ROI compared to recurring cloud costs.

These advantages make aiDAPTIV+ not only a more affordable solution but also a safer, more reliable choice for businesses keen on protecting their valuable data.

aiDAPTIV+ represents a significant leap forward for SMBs striving to stay competitive in the rapidly evolving landscape of AI and machine learning. By offering a low-cost, easy-to-use, and efficient solution for training LLMs, aiDAPTIV+ enables businesses of all sizes to leverage advanced AI technologies without the need for extensive resources or infrastructure. With aiDAPTIV+, the power of AI is now more accessible than ever, allowing SMBs to unlock new opportunities and drive innovation from the comfort of their server closets.

PHISON

phison.com

phisonblog.com

[aiDAPTIV+ solution](#)

YouTube Video:

[aiDAPTIV+ video](#)

Sales inquiries:

sales@phison.com

