WEKA

CASE STUDY

Accelerating AI Safety Research at Lower Cost in the Cloud

CAIS Background and Business Challenge

The Center for AI Safety (CAIS—pronounced 'case') is a San Francisco-based nonprofit that supports research and field building that promotes safe and responsible artificial intelligence (AI). CAIS believes that while AI has the potential to benefit the world profoundly, many fundamental problems in AI safety have yet to be solved. CAIS's mission is to reduce societal-scale risks associated with AI by conducting and building the field of AI safety research and advocating for safety standards.

Just as the field of AI continues to evolve rapidly, so do the long-term risks. Because future AI systems will likely present different risks than those we face today, CAIS takes a practical approach that encompasses both short and long-term strategies. Through building broad awareness and expertise, CAIS enables AI safety researchers to develop strategies to mitigate the risks we see today. By promoting AI safety research, CAIS encourages researchers to continually highlight new and emerging AI safety risks as the technology develops. Finally, the organization seeks to encourage today's AI developers to incorporate more AI safety into the models and projects they are working on through education and grassroots development programs. "We want our work to increase the number of people working on AI safety or to produce more AI-safety work," says Steven Basart, Research and Development Lead at CAIS.

CAIS-supported research includes topics central to the development of safe AI. Research on the robustness of safety guardrails in large language models (LLMs) highlights the need to limit third-party developers from bypassing safety controls. Research focused on developing technical methods to identify and measure the tendency of LLMs to hallucinate can help make AI systems more truthful and reliable. Work focused on determining the extent to which AI systems act based on reward systems versus ethics will help AI researchers understand how AI systems act according to ethical considerations. A complete list of CAIS-sponsored research is here.

Center for AI Safety

Challenges

- Slow AI model training times
- Performance and scale limits in legacy storage system
- Low utilization of GPU infrastructure storage resources
- High storage costs due to data management/copy requirements

Solution

- WEKA Data Platform Converged Mode

Benefits

- Grew research community from 30 to 200+ active researchers
- Accelerated storage performance by 5x
- Reduced data costs By 90%

The Research Challenge:

Enabling AI Safety Research in an ERA of GPU Scarcity

AI safety researchers who want to experiment on the latest LLMs face a dilemma. Conducting relevant research requires access to the latest GPU infrastructure to run experiments resembling real-world scenarios. However, the cost, complexity, space, and infrastructure skill sets needed to build an AI research cluster create high barriers for most AI-safety researchers. "Doing work on AI safety has recently become very cost prohibitive," explains Basart. For typical researchers, scarcity also comes in the form of space and power available for new on-campus AI research clusters. Along with the skill sets required to build and manage AI research clusters, few researchers can build their own infrastructure. Basart says, "It would be better to pool resources into one computing infrastructure to spread the costs across a large group of researchers, rather than each researcher spending millions to procure their own AI infrastructure to run experiments."

This is precisely what the CAIS Compute Cluster does. It is a dedicated GPU-accelerated cluster that provides AI safety researchers with subsidized, on-demand access to state-of-the-art infrastructure for LLM training and other AI safety projects. The CAIS compute cluster is specifically designed for researchers working on the safety of machine learning systems and supports a diverse range of research interests and collaborators.

> "We can support 6x the amount of research projects and are still growing. WEKA has unlocked a lot of research potential for us."

The Technical Challenge:

Scaling the CAIS Compute Cluster

The CAIS Compute cluster supports a widely diverse set of research interests. "Today, we have over 200 users utilizing the CAIS cluster," explains Basart. "Some researchers are doing reinforcement learning, where each simulation needs a single GPU, and then parallelize many simulations across a pool of GPUs. Then we have LLM research focused on model robustness, transparency, and interpretability." That diversity in research interests leads to challenges around resource management to enable the performance and scale researchers need while controlling costs. To enable the best researcher experience, the CAIS compute cluster runs on Oracle BM.A100 shapes with NVIDIA A100 GPUs, interconnected via a high-speed 1,600 Gbps network in the Oracle Cloud.

To manage job scheduling and efficient use of resources, the team relies on SLURM resource management tools and job scheduling.

The initial deployment relied on Oracle File Storage Service, which, as the native storage service in Oracle was easy to get started. However, as the team scaled up the cluster, they quickly faced challenges around performance, data management, and cost controls. "AI model training and tuning is heavy on metadata operations," says Basart. "We were bottlenecked on storage IOPs because the original system constantly did metadata lookups." Fast metadata processing has emerged as a critical infrastructure challenge to support

AI projects like LLM training and tuning. Model training occurs by reading and loading trillions of parameters onto one or more GPUs, followed by reading millions of individual files in a typical LLM training data set. The model constantly looks up the data it needs, finds the right file(s), reads, and then moves on. This emerging data analysis scenario often breaks legacy data systems designed for more continuous and predictable reads and writes, dramatically slowing down training times.

Data management became more challenging as more researchers were accepted into the cluster. "We had one episode where someone unexpectedly generated 50 TB of synthetic data during model training," observes Basart. As the team dug in, they found a data copy sprawl problem born from a lack of data management controls or quotas. Multiple researchers each manage model data in their own way, often storing multiple copies of the same training data set. This drives rapid data growth, resulting in spiraling storage costs.

90%
COST DROP

"We immediately saw our storage costs drop by 90% when we switched to WEKA."

The Solution:
WEKA Converged Mode

While the CAIS team investigated a few open-source storage options, such as Lustre, concerns around operational complexity and management overhead quickly led the team to WEKA. CAIS became interested in WEKA Converged Mode to reduce costs through more efficient resource utilization. "We noticed the storage on the compute nodes in our previous solution was being woefully underutilized," says Basart. "We had 70% of the local NVMe resources in the cluster going unused."

In Converged Mode (see Figure 1), the data storage environment resides on the same infrastructure resources as the model training environment. Contrast this with a traditional architecture where the model training (or applications) and data storage reside on separate silos of infrastructure. The goal is to drive significant cost savings across the data infrastructure, increase resource utilization in the GPU cluster, and be more efficient in data management. "We have some datasets that are commonly used across multiple experiments," explains Basart. "With the legacy system, we had to copy these data sets to all the cluster nodes to eliminate the network traffic and get the performance we needed."

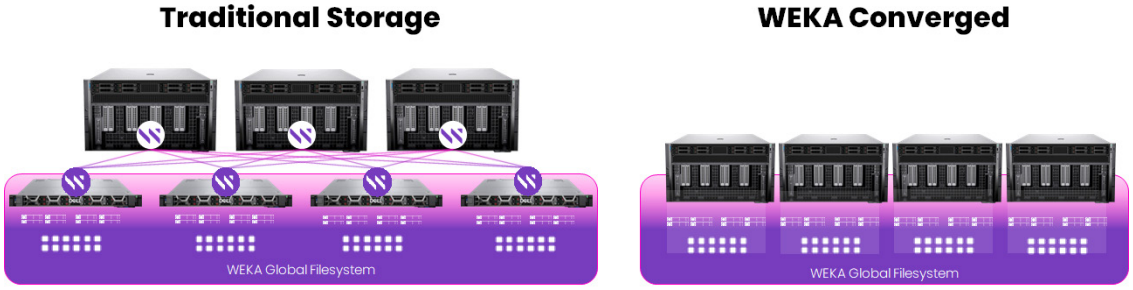

FIG. 1 WEKA Converged Mode versus a Traditional Deployment

The Benefits:

Higher Research Productivity at Lower Costs

Once WEKA Converged Mode was implemented, CAIS immediately reduced its cloud storage costs by 90%. Converged Mode enabled CAIS to support high-performance data operations using resources already available in the CAIS Cluster itself, with only the added cost of the WEKA software license. The new architecture takes advantage of under-used NVMe storage built into the nodes in the training cluster. In this case, resource utilization went from less than 20% to 100%. "It's weird to pay for something and not use it," says Basart. "Which is what was happening in the old architecture. We can now fully utilize the resources we have available." Adding in data management and setting storage quotas for researchers enabled further cost savings by eliminating unnecessary copies of research data.

CAIS was also able to dramatically improve the performance of the research cluster with WEKA. "The previous bottlenecks are gone. WEKA is able to fully saturate the network, " says Basart. "We would need a faster network to stress the WEKA environment." With the new WEKA deployment delivering 1.7 Million IOPs across an initial 10-node cluster, CAIS estimates they see a 5x improvement in storage performance versus the prior solution.

That performance is also translating to increased research productivity. When the CAIS team started looking for a new solution, they had an average of 30 researchers using the cluster at any given time. They were already experiencing delays as a result of storage bottlenecks. Today, over 200 researchers use the CAIS cluster on average, and there is no end in sight to the number of users the new environment can support. "We can support six times the number of research projects, and we are still growing. WEKA has unlocked much research potential for us," says Basart.

About the WEKA Data Platform

The WEKA® Data Platform removes the barriers to data-driven innovation through its advanced software architecture optimized to solve complex data challenges and streamline the data pipelines that fuel AI, ML, and other modern performance-intensive workloads.

The design philosophy behind the WEKA® Data Platform was to create a single architecture that runs on-premises or in the public cloud with the performance of all-flash arrays, the simplicity and feature set of network-attached storage (NAS), and the scalability and economics of the cloud. Whether on-premises, in the cloud, at the edge, or bursting between platforms, WEKA accelerates every step of the enterprise AI data pipeline – from data ingestion, cleansing, and modeling to training validation or inference.

Mind-bendingly fast. Seductively simple. Infinitely scalable. Sustainable. Spanning edge, core, hybrid, and cloud. The WEKA Data Platform helps to overcome complex data challenges and accelerate next-generation workloads to unleash your organization's imagination, creativity, and potential.

??????-01 04/24