Success Story

Enhancing Performance and Driving Groundbreaking Research

VAST | NCSA

Overview

The National Center for Supercomputing Applications (NCSA) at the University of Illinois Urbana-Champaign is a premier research institution dedicated to accelerating scientific research and discovery through cutting-edge HPC infrastructure. NCSA supports a wide range of users, including National Science Foundation (NSF) researchers, University of Illinois scientists, and Fortune 100 companies. To meet the complex and evolving demands of its users, NCSA needed a modern, reliable data platform capable of handling high I/O workloads and large-scale data processing.

Challenge

NCSA's infrastructure comprises multiple HPC clusters, storage systems, and secure environments that serve both academic and industry users. Managing this environment, which totals 60 petabytes of storage across 18 file systems, was becoming increasingly complex and resource-intensive. The ICI Directorate, the team responsible for deploying the cyber infrastructure to support the research at NCSA, needed to simplify operations while improving performance.

As J.D. Maloney, Lead HPC Storage Engineer at NCSA, explained: "Our challenge was to find a solution that could handle the read-heavy nature of home and software directories while reducing the overall footprint and complexity of our data environment." Additionally, with NCSA expanding into GPU-heavy research utilizing its 1,400 GPUs, the need for an optimized data platform that could keep up with demanding interactive and batch workloads became critical.

Solution

After evaluating several options, NCSA chose the VAST Data Platform for its home and software directories under its Harbor system. VAST's architecture offered the read performance, compression, and data reduction capabilities NCSA needed to streamline its operations. "The reason we went with VAST is that it's a very read-optimized file system. For areas like home and software directories, which are hit heavily by I/O, it was a perfect match," said Maloney.

One of the key benefits of VAST's system is its ability to deduplicate and compress data

About VAST

VAST Data is the data platform software company for the AI era. Accelerating time-to-insight for workload-intensive applications, the VAST Data Platform delivers scalable performance, radically simple data management and enhanced productivity. Launched in 2019, VAST is the fastest-growing data infrastructure company in history.

Looking to simplify your data center and unlock insights from all of your data? **Contact us at** hello@vastdata.com

Scan Here learn more about how VAST and Dremio are accelerating and streamlining modern analytics at enterprise scale.

efficiently. This allowed NCSA to provide its users with larger home directory quotas while keeping the physical footprint of the system manageable. “We’ve obtained a 3.5:1 data reduction ratio with the potential to hit 4:1 as we onboard more data,” Maloney noted. “This allowed us to support more systems with a smaller footprint and provide larger home directory quotas for our users.”

Results

Since deploying VAST Data, NCSA has seen significant improvements in its operations:

1. **Improved Performance**
   Users experienced faster load times for software and home directories, which is critical for interactive HPC workloads. For instance, module loads dropped from nine seconds to just 2.5 seconds, leading to a more responsive system overall. “Our user support staff were very excited. It felt like a new system because of how responsive it was,” Maloney shared.

2. **Data Efficiency**
   With VAST’s advanced compression and deduplication, NCSA achieved significant data reduction, allowing them to provide larger quotas with the same physical storage capacity.

3. **Streamlined Operations**
   VAST’s support for multiple fabric types (Ethernet, InfiniBand, Slingshot) and multi-tenancy capabilities allowed NCSA to consolidate its infrastructure. This reduced the number of file systems NCSA had to manage, resulting in fewer maintenance windows and more uptime for its clusters.

4. **Scalability for Future Growth**
   With the addition of GPU-optimized clusters like Delta AI, NCSA can now handle larger, more complex AI workloads. VAST’s read-optimized Data Platform ensures GPU resources are fully utilized by minimizing data loading times. “The ability to launch large containers across multiple nodes quickly is critical for our GPU workloads. VAST’s performance allows us to keep those expensive GPUs fully utilized, which is a huge win for us,” Maloney explained.

Expanding AI Capabilities

NCSA has adopted and integrated advanced AI infrastructure to accelerate scientific research across multiple fields, helping scientists tackle complex problems faster and with greater accuracy. NCSA’s focus on AI began with the deployment of Delta AI, one of its flagship systems designed to provide unparalleled computation power for large-scale AI workloads. Delta AI includes 700 Grace Hopper GPUs, making it one of the most powerful AI-focused systems in the research community. This system is aimed at supporting a broad range of AI applications, from machine learning and deep learning to computational biology, materials science, and climate modeling. It also advances NCSA’s ability to process complex datasets, train AI models, and conduct high-performance simulations across various scientific domains.

Key to NCSA’s AI infrastructure is the VAST Data Platform, which provides high-speed access to critical datasets and models without bottlenecks. AI workloads, especially in large containerized environments, need a data platform with extremely high read performance and low latency. VAST accelerates data access, especially for container-heavy AI workloads, maximizing GPU utilization and streamlining AI research pipelines at NCSA

Future Plans

As NCSA continues to expand its HPC offerings, it is exploring further use cases for VAST Data. Maloney is excited about VAST’s roadmap, particularly with upcoming features like metadata compression and enhanced write performance. “Metadata compression will allow us to increase user quotas even further. It’s a huge quality-of-life improvement for our users, and we’re excited to implement it as soon as possible,” said Maloney.

©2024 VAST Data, Inc. All rights reserved. All trademarks belong to their respective owners.
11/2024