

Success Story

TACC: Revolutionizing Data Management at Scale

VAST | TACC TEXAS ADVANCED COMPUTING CENTER

Home to some of the world's most powerful supercomputers, the Texas Advanced Computing Center (TACC) drives cutting-edge research with High-Performance Computing (HPC). Based at the University of Texas at Austin, TACC is a National Science Foundation Leadership-Class Computing Facility, providing the computational power behind discoveries that advance science and society. Specializing in HPC, data analysis, scientific visualization, and storage systems, TACC equips researchers across disciplines with the tools to process massive datasets, model complex phenomena, and solve problems once thought impossible.

Researchers at TACC tackle some of the world's most pressing challenges, requiring advanced computational resources to complete their work. Managing massive amounts of research data efficiently is critical to ensuring the integrity and availability of these resources.

Managing Billions of Files Efficiently

TACC operates multiple supercomputing clusters, including Stampede3 and Vista, which serve as foundational platforms for cutting-edge research. However, the sheer volume of data stored in these clusters created significant operational challenges.

According to Nic Lewis, Senior Large Scale Systems Administrator at TACC, one of the biggest hurdles was managing the ephemeral scratch file system, where researchers dump large amounts of data without any storage quotas. "Given that this is a shared system with hundreds to thousands of researchers, you have to have a way to keep things tidy," Lewis explained. To prevent storage overload, TACC regularly purged old, untouched data, but the existing process was inefficient and time-consuming.

> *Given that this is a shared system with hundreds to thousands of researchers, you have to have a way to keep things tidy.*
>
> **- Nic Lewis, Senior Large Scale Systems Administrator at TACC**

Prior to implementing VAST Data, TACC relied on the Lustre file system metadata utility to scan files, which was a slow process. “We would use Lustre’s list files utility, but ultimately it was still crawling the filesystem file-by-file,” said Lewis. “We would do that, as well as a find command to scroll the entirety of the file system, which took a week to potentially even longer.”

> *We would use Lustre’s list files utility, but ultimately it was still crawling the filesystem file-by-file. We would do that, as well as a find command to scroll the entirety of the file system, which took a week to potentially even longer.”*
>
> **- Nic Lewis, Senior Large Scale Systems Administrator at TACC**

The VAST Catalog Streamlines Data Management

To better manage its extensive amount of research data, TACC sought a more efficient solution and identified VAST Data’s Catalog feature as a game-changer. When TACC migrated to the VAST Data Platform, Catalog was an early feature that stood out. “The VAST Catalog was always a feature that we were eyeing early on and was one of the things that we were looking at when we first selected VAST as our data platform,” said Lewis.

With the VAST Catalog, TACC can now efficiently locate and manage files across its system in real-time. “Leveraging the catalog, I’m able to make direct queries to locate files with a set access time, create a dataset of those files, and then move them into VAST’s trashcan, which is significantly faster than traditional deletion methods,” Lewis shared. “Since moving files, I can stage the files and have VAST handle the deletions in parallel, which is sublime.”



A Quantum Leap in Efficiency

With the VAST Catalog, TACC has drastically improved the efficiency of its data management processes. Previously, the purge process on Lustre took days, sometimes weeks, and often had to be terminated early due to metadata server contention. Now, with the VAST Catalog, that same process takes mere hours.

“What would’ve taken me potentially four days with our prior purge process now takes me maybe a couple hours depending on the number of files,” Lewis reported. In a recent purge, TACC efficiently managed nearly 875 million files in just four hours—a process that previously could have taken weeks with Lustre.

Beyond speed, the VAST Catalog also eliminates the need for continuous metadata queries, reducing system contention and improving overall cluster performance. “The last purge that I ran encompassed 875 million files, which would’ve been a lot to crawl on Lustre, especially since we’re in a production environment where users are also performing metadata-heavy tasks,” Lewis noted.

Expanding Use Cases

Lewis envisions broader adoption of the VAST Catalog across TACC as its capabilities continue to evolve. As TACC expands its AI storage infrastructure with the upcoming Horizon NSF leadership-class facility, the VAST Catalog is expected to play an even greater role in optimizing data management and accessibility.

Looking ahead, TACC plans to leverage the VAST Catalog to enhance its noisy neighbor analysis, gain deeper insights into data lifecycles, and refine data curation strategies for researchers. The ability to provide real-time metadata insights can empower researchers to manage their own data more efficiently, further improving TACC’s overall storage ecosystem. “I’ve just only very slightly scratched the surface of the VAST Catalog,” Lewis said. “It’s a powerful tool, and I’m looking forward to exploring it more as our use case matures.”

About VAST Data

VAST Data is the data platform company built for the AI era. As the new standard for enterprise AI infrastructure, organizations trust the VAST Data Platform to serve their most data-intensive computing needs. VAST Data empowers enterprises to unlock the full potential of their data by providing AI infrastructure that is simple, scalable, and architected from the ground up to power deep learning and GPU-accelerated data centers and clouds. Launched in 2019, VAST Data is the fastest growing data infrastructure company in history. For more information, please visit and follow VAST Data on X (formerly Twitter) and LinkedIn.

Scan here to learn more.

> *The last purge that I ran encompassed 875 million files, which would’ve been a lot to crawl on Lustre, especially since we’re in a production environment where users are also performing metadata-heavy tasks.*
>
> **- Nic Lewis, Senior Large Scale Systems Administrator at TACC**

©2025 VAST Data, Inc. All rights reserved. All trademarks belong to their respective owners.
05/2025