



Bridging the Gap Between Cloud-Like Aspirations & Realities in HPC

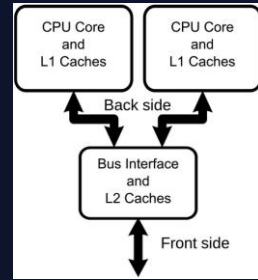
Phil Nass

Presales Director - APJ, VAST Data



Two Decades of Trends and Transformations in HPC

2000



2010



2020

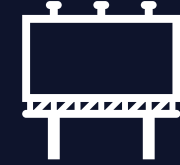


2023



GPU Computing Applications
Application Acceleration Engines (AXEs) ScenIX, CompeX, Optix, PhysX
Foundation Libraries CUBLAS, CUFFT, CULA, NVCUVID/VENC, NVPP, Magma
Development Environment C, C++, Fortran, Python, Java, OpenCL, Direct Compute, ...
CUDA Compute Architecture





USER EXPECTATIONS



Cluster & Resource Availability



Secured Multi-Tenancy



Latest Technologies to Support AI Initiatives



Multi-Protocol, Consistent & Available Datasets with PFS levels of performance

REALITIES IN HPC CENTRES



Budget Constraints & Cost Recovery



HPC Skillset Shortage



Support Forward-Looking Initiatives while Keeping the Lights On

VAST Data Power the Leading AI CSPs & HPC Sites

Single Platform for Rich Catalog of Data Services

- ✓ Secure Multi-Tenancy (QoS)
- ✓ Enterprise-Grade Security
- ✓ PFS Performance
- ✓ Simplified Operations
- ✓ Unbeatable Economics



VAST is the Data Platform for CoreWeave's 14 Data Centers; Partnering on BlueField-Enabled AI Architecture



100s of PBs of VAST Deployed for G42 AI Cloud: "AI Data Lake" Powering Exaflops of Computing



DownUnder GeoSolutions (DUG) is a Western Australian geosciences company that provides HPC-aaS solutions across multiple geographies.



Texas Advanced Computing Center (TACC) partnered with VAST to move beyond Lustre for Exascale

2023 HPC SYSPROS Survey

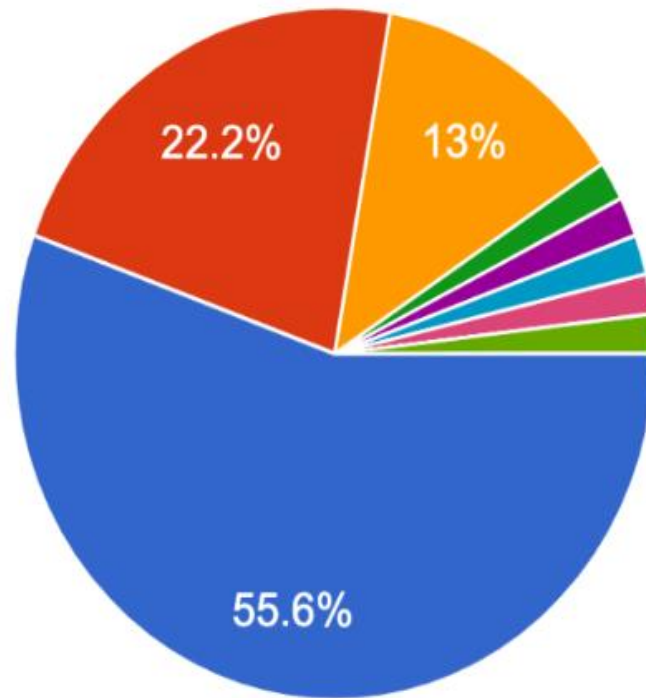
Author: Jenett Tillotson, Senior HPC Systems Engineer, National Center for Atmospheric Research

3M Company
Agriculture Victoria Research
Ansys
Brown University
Chan Zuckerberg Biohub
Cornell University
Digital Research Alliance of Canada
Harvard SEAS
Howard Hughes Medical Institute
Indiana University
LBNL/NERSC
Leidos Inc. - Centers for Disease Control & Prevention
Michigan State University
Middle Tennessee State University
Minnesota Supercomputing Institute
Mississippi State University
NAG/bp
NCAR
NCAR/UCAR
NCSA

NIH
NIH
NIH/NHLBI
NIHM
P&G (AFDS)
Pawsey Supercomputing Centre
Penn State Institute for Computational and Data Sciences
TACC
Tennessee Tech University
Texas A&M
The MITRE Corporation
The University of Arizona
UC Davis
UIUC/NCSA
University of British Columbia
University of California at Berkeley
University of Chicago
University of Pennsylvania
University of Pennsylvania
University of Utah

What type of institution do you work for?

54 responses



- Academic
- Government
- Industry
- Non-profit research nondegree granting
- Non-profit and FFRDC's
- Government funded for academic use - not sure which to tick
- Non-profit academic/research
- Government Owned, Contractor Operated.

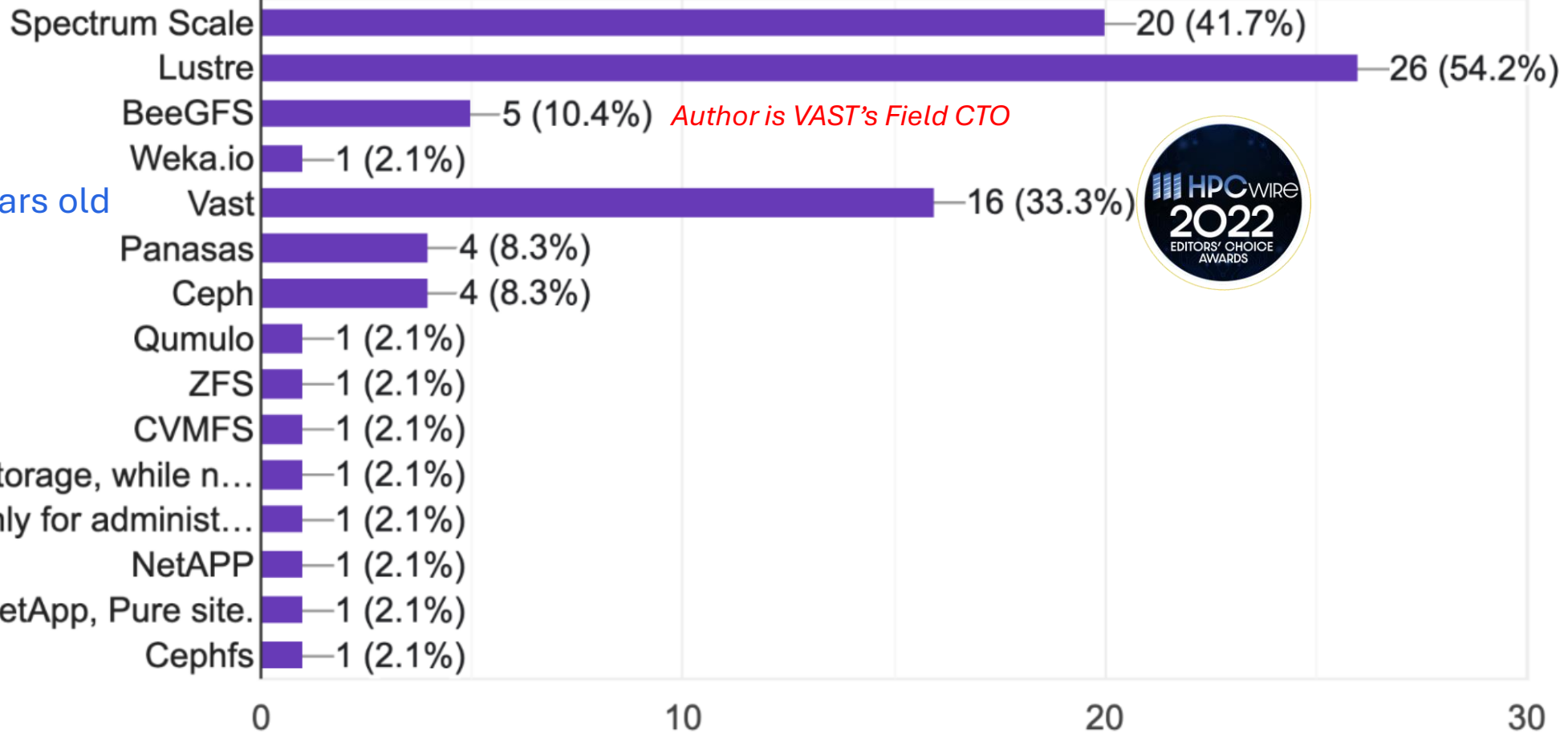
What clustered file systems do you have at your institution (select all that apply)?

48 responses

25 years old!

20 years old!

Less than 4 years old



Source: 2023 HPC SYSPROS Survey – Jenett Tillotson,
Senior HPC Systems Engineer, National Center for Atmospheric Research (UCAR)



“VAST hasn’t just changed the way we store data; it is changing our relationship with data.

Success Story

- Eliminated any user trouble tickets vs. 1 per day with Lustre
- HPC codes run up to 20x faster
- Quality of service controls for cloud tenants
- A simple, VAST-managed solution serving ~15,000 Compute Nodes





“We partner with VAST Data to deliver a multi-tenant and zero-trust environment purpose-built for accelerated computing”

Success Story

- Single clusters consisting of 100s to 10,000s of GPUs
- VAST provided the ideal mix of AI-Scale, Cloud-Scale and Enterprise Capability
- Delivering ML, VFX and Streaming solutions that are up to 35x faster and 80% less expensive than legacy cloud platforms



Aligning To Build Better AI and Data Pipelines



Generative AI & LLMs

Data Analytics / Science

Digital Twin

Classic AI (CV, NLP)

CPU/GPU Server Layer

Hardware Infrastructure | System Software | Acceleration Libraries | Application Frameworks



Data Ingest & Serving

Scalable throughput via standard NFS, SMB, S3 (object) access for ingest and homedirs

Data Prep

Integrated database for ultra efficient ELT operations and real-time queries

Model Training

High-bandwidth, parallel data access for accelerated compute platforms (ie. DGX SuperPOD)

Inferencing

Low-latency access for AI applications making real-time predictions

VAST Data Platform

Single File / Object / Table Namespace | Embarrassingly-Parallel | 100% Online Operations | Industry-best Data Reduction | Multi-Tenant Ready



VAST Data

The Data Platform for the AI Era

Faster Time to Results

Measured at 15x faster for nuclear physics, 6x faster for AlphaFold.

Secured Multi-Tenancy

Powering the leading AI Clouds and multiple tenants

Larger Workloads

Scales to 10x larger than existing parallel filesystems.

More Flexible Open Data Architecture

Unlock Cloud and S3 access to local file (POSIX) datasets.

Far More Sustainable

Less eWaste, longer lifetime, lower power consumption.

Far Less Downtime

Already delivering 99.999% uptime (under 6 mins per year).

A person wearing a white protective suit and a gold helmet is shown in profile. The word "VAST" is overlaid in white, stylized text. Above the letters "A" and "S" are three white arrows pointing downwards. The background is a blurred outdoor scene with trees and a light sky.

VAST