Modernizing HPC Infrastructure and Administration with VAST Data

Table of Contents

Chapter 1: HPC Infrastructure Basics

Introduction - The Foundation of Modern HPC Infrastructure

High-Performance Computing (HPC) is evolving faster than ever, with AI and data-intensive workloads now converging in the same infrastructure. Initially designed for compute-heavy workloads, conventional parallel file systems like Lustre and GPFS are increasingly burdened by modern AI-driven demands for low-latency, high-throughput data access. Many organizations struggle with complex tuning, data movement challenges, and high administrative overhead—limiting research, innovation, and overall productivity.

The "Transforming High-Performance Computing with VAST Data" white paper introduced the fundamental challenges of HPC storage and how the VAST Data Platform removes conventional bottlenecks. This guide provides an in-depth technical deep dive into how modernizing HPC storage with the VAST Data Platform eliminates these limitations, simplifying HPC administration while delivering HPC and AI performance in one unified architecture.

Who Should Read This Guide?

This guide is designed for:

- **HPC Administrators** – Deep dive into storage, networking, and workload optimization.

- **IT Decision-Makers & Storage Architects** – Evaluate next-gen HPC platforms that unify AI and HPC.

- **Researchers & Engineers** – Enable high-performance, scalable, and efficient HPC storage solutions.

Unlike the primary HPC white paper, which introduces key HPC concepts, this document dives deeper into HPC architecture, job scheduling, networking considerations, and best practices for storage administration.

Suppose you are new to HPC storage and the VAST Data Platform. In that case, we recommend starting with the "Transforming High-Performance Computing with VAST Data" white paper before proceeding with this more technical deep dive.

Why This Guide?

For those seeking a detailed understanding of HPC infrastructure, this guide explores:

- **Key Technologies** – Deep dive into NVMe-over-Fabrics (NVMe-oF), Remote Direct Memory Access (RDMA), and parallel computing frameworks like MPI that power modern HPC.

- **Network Considerations** – Compare Ethernet vs. InfiniBand and understand their impact on performance and scalability.

- **Parallel File Systems** – Analyze the limitations of Lustre and GPFS while exploring how VAST's Disaggregated Shared Everything (DASE) architecture outperforms them.

- **Optimized Administration** – Learn how VAST simplifies storage management and eliminates complex, high-maintenance HPC storage configurations.

The Convergence of HPC and AI—Why VAST Data is the Ultimate Platform

HPC is no longer just about scientific simulations and research computing—it is now intrinsically linked to AI-driven workloads. Conventional HPC storage architectures struggle to balance AI's need for rapid, low-latency data access with the throughput-heavy demands of parallel HPC applications.

The VAST Data Platform is designed to solve this challenge, providing a unified infrastructure that seamlessly supports:

- **High-Performance Simulations** – Ideal for research institutions, genomics, and climate modeling, where data throughput is mission-critical.

- **AI Model Training & Inference** – Run AI-driven workloads without requiring a separate storage infrastructure or excessive data movement.

- **Financial Modeling & Risk Analysis** – Process real-time analytics with ultra-low-latency and high-throughput performance for maximum insight.

Unlike conventional parallel file systems that require manual tuning, complex failover mechanisms, and extensive administration, VAST eliminates operational complexity while delivering:

- **Ultra-Fast Storage** – All-flash NVMe-oF and RDMA architecture ensures extreme speed without bottlenecks.

- **Multiprotocol Access** – Seamless NFS, SMB, S3, and Block support enables cross-platform HPC and AI workloads.

- **Simplified Administration** – Automated, complaint-free storage management dramatically reduces the operational burden.

How to Use This Guide

This guide is structured to provide technical depth on HPC infrastructure while always keeping a VAST-centric focus on how the VAST Data Platform simplifies and accelerates HPC + AI workloads.

As you explore the fundamentals of HPC architecture in this guide, remember how the VAST Data Platform revolutionizes performance, efficiency, and scalability for modern HPC environments. Whether operating a university research lab, a financial institution, a government HPC facility, or an AI-driven enterprise, VAST ensures that your storage infrastructure is no longer a limitation but a competitive advantage.

Overview of HPC Architecture

HPC systems comprise multiple specialized components working in unison to maximize computational power, data handling, and efficiency. Each component has a specific role, allowing the entire system to handle large-scale tasks with minimal bottlenecks. Here's an overview of the key elements:

- **Management Nodes:** These nodes act as the brains of the HPC cluster, overseeing the entire system's operation. Management nodes handle workload scheduling, resource allocation, system monitoring, and job queuing, ensuring that compute resources are used efficiently.

- **Head Nodes:** The head nodes are the primary access points for users submitting tasks to the cluster. Users log into the head nodes to compile code, configure jobs, and monitor progress. The head nodes often have dedicated network access to enable fast data movement and are linked to Network File System (NFS) storage for users' home directories.

- **Compute Nodes (Worker Nodes):** Compute nodes carry out the actual processing of tasks. These nodes, typically equipped with powerful CPUs and GPUs, are optimized for intensive computational work. Each compute node works in parallel with others, processing a portion of the workload and passing results back to the head node. In large clusters, thousands of compute nodes may work together to solve highly complex problems.

- **Scratch Storage:** Scratch storage provides temporary, high-speed storage that compute nodes can use to read and write data quickly during calculations. This storage is optimized for speed rather than durability, as it's primarily used for intermediate data processing, and data is often deleted when tasks are complete. Scratch storage is typically located close to the compute nodes, providing fast access to data during active computation.

Each of these components work together to enable the efficient processing of tasks, focusing on maximizing computational resources, minimizing latency, and improving throughput.

NVMe-oF and RDMA Explained

Two key technologies that enhance data transfer efficiency in HPC are NVMe over Fabrics (NVMe-oF) and Remote Direct Memory Access (RDMA). These protocols can operate over Ethernet and InfiniBand networks, making them versatile solutions for modern HPC architectures.

NVMe over Fabrics (NVMe-oF)

NVMe-oF is designed to extend the NVMe protocol (primarily for fast data storage) over a network, allowing for high-speed access to NVMe storage devices from remote nodes. NVMe-oF is highly beneficial in HPC for accessing large datasets stored on NVMe devices across a network.

- **Ethernet:** NVMe-oF can run over Ethernet networks using protocols like RDMA over Converged Ethernet (RoCE). This allows NVMe's benefits to be extended across Ethernet infrastructures.

- **InfiniBand:** NVMe-oF over InfiniBand combines InfiniBand's low latency with NVMe's high-speed data access, creating an ideal setup for accessing large datasets in HPC.

Remote Direct Memory Access (RDMA)

RDMA enables data to be transferred directly between memory locations on different computers without involving the CPU, bypassing conventional bottlenecks in data transfer. This allows for much faster data movement with lower latency, as CPU cycles are freed up for other tasks. RDMA's benefits vary slightly depending on the type of network it runs over:

- **Ethernet:** RDMA can be used on Ethernet networks through protocols like RDMA over Converged Ethernet (RoCE) and Internet Wide Area RDMA Protocol (iWARP). This allows HPC systems to benefit from RDMA on Ethernet-based infrastructures, making it accessible in a broader range of data center environments.

- **InfiniBand:** RDMA was initially designed for InfiniBand, which offers a high-speed, low-latency communication channel between nodes. In HPC environments where latency is critical, RDMA over InfiniBand delivers optimal performance.

NVMe-oF and RDMA are critical in modern HPC setups. They allow for efficient data movement, reduced latency, and optimized CPU usage, all essential for handling complex, data-intensive tasks.

Ethernet vs. InfiniBand Network Considerations

In HPC environments, network selection is critical in determining performance, scalability, and cost-effectiveness. Ethernet and InfiniBand are the two most commonly used network technologies, each with specific advantages depending on workload requirements and infrastructure goals.

Selecting the right interconnect can significantly impact data movement efficiency, application performance, and total infrastructure cost. Conventional parallel file systems like Lustre and GPFS are susceptible to network selection, often requiring extensive manual tuning to ensure optimal performance. This leads to administrative overhead and unpredictable network bottlenecks, making network choice even more critical in modern HPC and AI environments.

Why a Customer Would Choose Ethernet Over InfiniBand

Ethernet remains the most widely used networking technology in data centers and HPC environments due to its cost-effectiveness, ubiquity, and ease of management. It provides a good balance of performance and affordability for many HPC workloads.

- **Cost Efficiency:** Ethernet is generally more affordable than InfiniBand in hardware and maintenance. Ethernet offers a cost-effective interconnectivity solution for HPC environments where budget constraints are a priority.

- **Compatibility and Ubiquity:** As a widely adopted networking standard, Ethernet ensures broad compatibility across devices, servers, and storage systems. Its presence in data centers worldwide makes it a practical choice for organizations preferring standardized, universal networking.

- **Simplified Management:** Ethernet networks are generally easier to deploy, manage, and maintain. A broader talent pool of IT professionals is already familiar with Ethernet configurations, reducing the need for specialized expertise.

- **Sufficient Performance for Many HPC Workloads:** While low-latency communication is essential for tightly coupled HPC applications, many workloads—such as data processing, analytics, and cloud-based HPC—perform adequately over Ethernet without requiring InfiniBand's advanced capabilities.

Why a Customer Would Choose InfiniBand Over Ethernet

InfiniBand delivers unmatched low latency and high bandwidth for HPC workloads requiring extreme performance. It is preferred for tightly coupled, high-performance applications such as AI model training, scientific simulations, and large-scale parallel processing.

- **Low Latency and High Bandwidth:** InfiniBand provides significantly lower latency and higher bandwidth than Ethernet, essential for compute-intensive HPC applications such as AI model training, climate simulations, and molecular dynamics, where fast inter-node data exchange is critical.

- **Enhanced Scalability for Large HPC Clusters:** InfiniBand is designed to support large-scale HPC clusters and high-performance workloads with better scalability than Ethernet. Its architecture minimizes latency as clusters expand, making it the preferred choice for large supercomputing environments.

- **RDMA (Remote Direct Memory Access):** InfiniBand natively supports RDMA, allowing direct memory-to-memory transfers between nodes without CPU involvement. This feature reduces overhead and decreases latency, providing a significant advantage in HPC workloads requiring fast, frequent data exchanges.

- **Optimized for Parallel Processing**: InfiniBand is purpose-built for highly parallel HPC workloads, enabling seamless, high-speed communication across nodes to ensure efficient data flow in synchronized, multi-node operations.

How Conventional Parallel Storage Impacts Network Choice

Conventional HPC parallel file systems like Lustre and GPFS are highly dependent on network infrastructure, often requiring extensive manual tuning to maintain performance:

- **Ethernet Bottlenecks**: Lustre over Ethernet requires significant tuning to minimize latency and optimize performance, adding complexity to network management.

- **Complex RDMA Tuning:** GPFS over InfiniBand demands complex RDMA adjustments, increasing administrative overhead and slowing deployment.

VAST Data Advantage: Supporting Both Ethernet and InfiniBand

Unlike conventional HPC parallel file systems requiring tuning for different network infrastructures, the VAST Data Platform is designed for seamless performance across Ethernet and InfiniBand.

- **Seamless RDMA Support Across Both Ethernet & InfiniBand:** No manual tuning required.

- **NVMe-over-Fabrics (NVMe-oF) with RDMA:** Ensuring high-throughput storage performance regardless of interconnect choice.

- **Dynamic Workload Adaptability:** VAST automatically optimizes data movement, avoiding conventional bottlenecks caused by network selection.

- **Multiprotocol Access:** Supports NFS, SMB, S3, and Block natively, without requiring performance-draining gateway solutions.

- **No Need for Additional Gateways:** Unlike conventional HPC storage, which often requires protocol translation gateways for multi-network environments, VAST natively supports both Ethernet and InfiniBand without data translation layers, reducing latency, complexity, and failure points.

Whether an HPC center runs on high-bandwidth InfiniBand or cost-effective Ethernet, VAST eliminates tuning complexities and ensures consistent, optimized performance for any workload.

NVMe-oF, RDMA, and RoCE with Genomics Example

Imagine a research team analyzing genomic data to develop new treatments. Every day, they process terabytes of data from thousands of genetic sequences. Speed is critical—they need results fast to move the project forward. Here's where VAST Data's platform, with NVMe-oF, RDMA, and RoCE, becomes the backbone of their operation.

When the scientists submit a job, the compute nodes spring to action, contacting the VAST Data storage system. Instead of waiting for the conventional slow storage read times, NVMe-oF enables these nodes to access massive data volumes at speeds as if the data were on the nodes themselves.

Meanwhile, RDMA allows the compute nodes to pull data directly from the storage, bypassing CPU bottlenecks. This direct memory access ensures the data flows smoothly without interruption, significantly reducing processing delays. It's like having a dedicated express lane for data between storage and computing. This happens over RoCE, which carries RDMA over the lab's Ethernet network.

With the VAST Data Platform running NVMe-oF, RDMA, and RoCE, the researchers experience ultra-fast data processing, allowing them to complete what would take days in a matter of hours. As data flows effortlessly from storage to compute, insights come to life faster, accelerating breakthroughs and keeping the research moving forward

at lightning speed.

Message Passing Interface (MPI) and Parallel Writing

Message Passing Interface (MPI) is a standardized and portable message-passing system designed to function on various parallel computing architectures. It is fundamental in HPC, particularly for distributed computing environments where tasks are distributed across multiple nodes, allowing them to work simultaneously on parts of a problem. MPI enables data exchanges between processes running on different nodes, thus facilitating parallel operations essential for handling large-scale scientific, engineering, and AI applications.

MPI and Parallel Writing in HPC

In conventional parallel writing, MPI enables multiple processes to write to storage concurrently, increasing data throughput and improving efficiency for I/O-intensive applications. MPI's approach leverages the system's network, passing messages between nodes and coordinating write operations to avoid conflicts and ensure data consistency. However, this coordination, while effective, can introduce latency and communication overhead, mainly as datasets grow and the demand for high-throughput, low-latency storage rises.

How VAST Data Enhances MPI Workflows

VAST Data addresses these challenges with its Disaggregated Shared Everything (DASE) architecture. Unlike conventional shared-nothing architectures, VAST Data's DASE model allows any compute node to access data directly across the cluster without relying on a centralized controller. VAST Data's NVMe-over-Fabrics integration with RDMA (Remote Direct Memory Access) enables high-speed, low-latency communication, facilitating parallel reads and writes that MPI can leverage. The VAST Data Platform improves upon typical MPI-based I/O workflows:

- **Optimized Data Access:** VAST Data's globally accessible data structures eliminate east-west traffic between nodes for metadata synchronization, enabling parallel access and reducing bottlenecks.

- **Enhanced Data Resilience:** Unlike competitors who rely on centralized metadata managers, VAST Data's architecture supports decentralized access, enabling resilience and reducing potential I/O failures in MPI workflows.

- **Scalability and Isolation:** VAST Data's system can handle hundreds of thousands of parallel MPI processes without experiencing significant performance degradation, thanks to its ability to isolate 'noisy neighbors' in parallel operations, a limitation in many shared-nothing designs.

- **Cost Efficiency:** VAST Data's all-flash infrastructure provides flash performance with the economics of hard drives, reducing TCO while maintaining the performance necessary for MPI-driven workloads.

HPC Software Stack

Software in HPC environments is specialized to manage the complexity of multi-node operations and resource allocation. Some critical components of the HPC software stack include:

- **Job Scheduling (e.g., Slurm):** Job schedulers are essential in HPC, managing the submission, prioritization, and execution of jobs across the compute nodes. Slurm (Simple Linux Utility for Resource Management) is one of HPC's most widely used schedulers. It allows administrators to define the allocation of resources, queue jobs, and manage task dependencies, ensuring efficient resource use across the cluster.

- **Nameservers and Management Software:** Nameservers handle the organization and addressing of resources across nodes, enabling efficient data retrieval and communication. Management software provides administrators with tools to monitor cluster health, manage resources, and troubleshoot issues, which are vital for ensuring HPC performance.

- **Linux Compatibility and Security:** Most HPC environments are Linux-based, as Linux offers the flexibility, compatibility, and performance needed to manage HPC resources. Security tools are also integrated within the Linux ecosystem to protect sensitive data, especially in healthcare and financial services.

Together, these tools form the backbone of an HPC software stack, allowing users to submit, track, and manage tasks across an extensive array of resources with maximum efficiency.

Checkpoints in HPC vs. AI

Checkpointing is a critical component of both HPC and AI workloads, but these two domains have fundamentally different storage access patterns. In HPC environments, checkpointing ensures fault tolerance, allowing large-scale simulations to resume from saved states rather than restarting from scratch after a failure. These HPC checkpoints are typically extensive sequential data writes that occur periodically to preserve computational progress.

In contrast, AI workloads rely on frequent, incremental checkpointing, where model weights and training states are saved at high frequency (often every few minutes or even seconds). Unlike HPC, where checkpoints serve as simulation recovery points, AI checkpoints enable iterative improvements in machine learning models—ensuring that training can resume without losing progress. These AI checkpoints comprise thousands to millions of small, random writes, significantly increasing metadata overhead and straining conventional storage systems.

The problem arises because conventional parallel file systems like Lustre and GPFS were designed for HPC checkpointing—not AI. These legacy systems treat all checkpointing the same, assuming workloads operate with large, infrequent writes rather than small, continuous updates. This results in:

- **Metadata Bottlenecks:** Conventional file systems rely on centralized metadata servers, which become overwhelmed when AI workloads generate millions of small checkpoint files every few minutes.

- **Write Amplification Issues:** AI checkpointing creates excessive small writes, leading to fragmentation and storage inefficiencies that slow down retrieval performance.

- **Complex File System Tuning:** HPC admins must manually adjust caching, striping, and data tiering settings to balance HPC and AI checkpointing, adding unnecessary administrative overhead.

- **Performance Degradation Over Time:** The mixed workload nature of HPC+AI leads to checkpointing bottlenecks, reducing both simulation restart efficiency and AI model training speed.

How VAST Data Eliminates This Problem

- **Optimized for HPC & AI Checkpointing Together:** VAST's disaggregated storage architecture intelligently adapts to different checkpointing workloads without requiring manual tuning.

- **No Impact from Small, Frequent Writes:** Unlike Lustre and GPFS, VAST efficiently handles AI checkpointing without metadata congestion.

- **Automated, Real-Time I/O Optimization:** VAST dynamically adjusts data placement based on workload type, ensuring fast recovery for HPC and rapid resumption for AI models.

- **Scales AI Checkpointing Without Bottlenecks:** Unlike conventional parallel storage, which struggles under the exponential growth of AI model checkpoints, VAST effortlessly scales AI training workloads without metadata congestion or manual tuning.

The VAST Data Platform is built for modern HPC environments, delivering seamless performance whether checkpointing a multi-petabyte simulation or saving an AI model every few minutes.

Workflow – Preparing and Submitting an HPC Job

HPC jobs are structured to allow complex tasks to be broken down and processed in parallel across compute nodes. Here's a high-level workflow for preparing and submitting an HPC job:

1. **Login to Head Node:** Users log in to the head node to begin the job setup process. This node is the user interface for configuring tasks, compiling code, and defining resource requirements.

2. **Job Script Preparation:** Users prepare a job script, which includes instructions on the required resources (e.g., number of nodes, memory, CPU/GPU usage), and the executable command. These job scripts often include shell commands and specific scheduler directives that define how the job will run in the cluster.

3. **Job Submission:** The job is submitted to the job scheduler (e.g., Slurm), which places it in a queue, allocates resources, and manages its execution. The scheduler ensures that each job can access the necessary resources while balancing the load across the cluster.

4. **Job Execution and Monitoring**: The job executes resources across the designated compute nodes once resources are allocated. Users can monitor the job's status from the head node or a management console, tracking metrics like completion time, resource usage, and errors.

5. **Result Collection and Cleanup:** After completing the job, results are stored in the designated output location (often on scratch storage for immediate access). Users can then transfer data as needed and clean up temporary files, freeing resources for the next set of tasks.

This structured workflow allows HPC environments to maximize their computational capabilities, handling multiple jobs concurrently and ensuring the efficient management of each job's data and resources.

Summary of HPC Infrastructure Essentials

The infrastructure supporting High-Performance Computing (HPC) is both powerful and intricate, and it is designed to manage the scale and speed demanded by complex computational workloads. This chapter has explored the essential components of HPC environments, covering hardware, networking, software stack, and workflow elements that collectively enable high-performance results.

1. **Key Hardware Elements:** An HPC cluster typically includes management nodes for resource allocation and system oversight, head nodes for user interaction and job configuration, and compute nodes for intensive processing tasks. Scratch storage, optimized for speed rather than durability, allows for rapid data access during computations and temporary data handling.

2. **Networking Protocols and Performance:** HPC systems require robust networking solutions to transfer data across nodes efficiently. Ethernet and InfiniBand provide high-bandwidth and low-latency connections, which are crucial for minimizing delays and preventing bottlenecks in data-heavy workflows. The integration of NVMe-oF and RDMA, as seen in VAST Data's platform, further enhances data flow, especially for genomics and AI-driven applications, by reducing CPU dependency and accelerating data access.

3. **HPC Software Stack:** Job schedulers like Slurm allocate tasks across the cluster, nameservers organize resource addressing, and Linux-based tools offer the flexibility and security essential for data-intensive sectors like healthcare and finance. The structured approach to job submission and execution allows for parallel processing across nodes, maximizing computational capabilities and ensuring the efficient management of data and resources.

4. **Checkpointing and Workflow Management:** HPC workflows demand periodic checkpointing for data integrity and recovery in long-running tasks. VAST Data's architecture supports both HPC's large-scale checkpoints and AI's frequent, iterative checkpoints on a single platform, providing a versatile solution that simplifies management and ensures seamless data accessibility.

By understanding these foundational elements, HPC administrators and users are equipped to leverage the infrastructure effectively, supporting advanced computational needs and driving breakthroughs in research and industry. This chapter sets the stage for exploring how VAST Data's unique storage capabilities address the challenges of HPC environments, simplifying data management, and enhancing workflow efficiency.

Chapter 2: Redefining HPC Storage with VAST Data

HPC Data Storage Needs

In HPC environments, storage systems play a multifaceted role, supporting not only high-speed scratch storage but also providing resources for user-specific workspaces, long-term data retention, and efficient data access across varied workflows. Conventional HPC storage often relies on multiple storage tiers, requiring data movement between scratch, home, project, and archival tiers. However, VAST Data's Disaggregated Shared Everything (DASE) architecture simplifies this landscape by eliminating the need for data pre-movement across storage types, allowing data to remain accessible in a unified, high-performance layer.

This section outlines the various types of storage in HPC environments. It explains how VAST Data's platform unifies these needs, delivering fast, scalable access without the bottlenecks and complexity of conventional multi-tiered storage.

Types of Storage in HPC

- **Scratch Storage:** Scratch storage is high-performance, temporary storage optimized for active computational workloads. HPC jobs rely on scratch storage to read and write massive amounts of data during execution, making high throughput and low latency essential for maximizing efficiency. Traditionally, scratch storage serves as a transient data space, requiring data movement as tasks are completed.

- **NFS Home Directories:** Home directories, typically managed through the Network File System (NFS), store user-specific files, configurations, and smaller datasets. These directories provide a personal workspace separate from high-speed scratch storage and are vital for managing individual workloads and user data.

- **Project and Bulk Storage:** HPC workflows often require project-specific or bulk storage for large datasets accessed intermittently. Examples include research archives, reference datasets, and result storage. This tier generally serves data that doesn't require the I/O intensity of scratch storage but must be readily available for extended periods, such as when revisiting past research.

- **Archive and Backup Storage:** Long-term archival storage is essential for compliance, historical data preservation, and data redundancy. This storage tier is accessed infrequently but provides critical support for data integrity and risk management, ensuring that information is preserved over the long term.

Eliminating Data Pre-Movement Requirements

In conventional HPC environments, data must frequently move between different storage tiers, often starting in scratch storage for processing and then moving to home, project or archival tiers for long-term retention. This process adds time and complexity, as data transfer and pre-processing can create bottlenecks in workflows.

VAST Data's DASE architecture simplifies this by providing a single, high-performance storage layer that supports both scratch and long-term storage needs without requiring data pre-movement. Data can remain in place from ingestion through processing and archiving, reducing the administrative burden of managing transfers between tiers and accelerating data availability across workflows.

By maintaining data accessibility across its lifecycle, VAST Data's approach enables HPC environments to operate more efficiently, allowing researchers and administrators to focus on computation rather than data logistics.

DASE and the Power of Embarrassingly Parallelism

Its capability for embarrassingly parallelism is at the heart of VAST Data's Disaggregated Shared Everything (DASE) architecture. This term highlights the architecture's ability to handle large, independent workloads simultaneously across multiple nodes without interdependency. This parallelism is a unique strength in HPC because it allows the system to scale effortlessly while maintaining top performance, even as data and compute demands increase.

Key Benefits of Embarrassingly Parallelism with DASE

1. **Linear Scalability for Growing HPC Needs:** DASE's design allows each node to work independently on data, enabling performance to scale linearly as additional nodes are added. In conventional storage systems, increasing nodes can introduce coordination overhead, reducing the returns on added resources. With DASE, there is no bottleneck or interdependency, meaning that each new resource contributes directly to throughput and computational power. For HPC, where scalability is essential, this architecture is ideal for supporting massive data growth and complex computations without sacrificing speed.

2. **High Throughput and Minimal Latency:** DASE's parallelism enables continuous, high-speed data access across all nodes, making it possible to support multiple, high-demand workloads simultaneously. Each node has direct access to all data, bypassing the need for data copies or serialized processes, which reduces latency. This architecture is particularly advantageous for HPC tasks requiring fast, synchronous access to data, such as large-scale simulations and real-time analytics.

3. **Optimal Performance for HPC and AI Workloads:** In HPC and AI, embarrassingly parallel tasks are common, as they allow complex computations to be broken down into smaller, manageable segments that can be processed in parallel. DASE's architecture supports these tasks by allowing data-intensive workloads, such as AI model training or scientific simulations, to be distributed across the storage cluster without interference. This parallelism means faster computation times, quicker access to results, and increased productivity in HPC environments.

4. **Fault Isolation and Enhanced Reliability:** DASE's shared-everything design ensures that each node operates independently, which limits the impact of any single node failure. This setup makes DASE inherently fault-tolerant, as the system can isolate and bypass failing nodes without affecting overall performance. For HPC environments, which rely on continuous uptime and high reliability, DASE provides a dependable infrastructure that reduces downtime and ensures data availability.

5. **Simplified Management with Reduced Complexity:** Conventional storage architectures often require complex node coordination, especially in high-parallelism scenarios. DASE eliminates the need for such coordination, streamlining management and reducing operational overhead. The VAST Data Platform architecture enables a simplified, efficient environment that supports high levels of parallel processing without requiring intensive monitoring or configuration adjustments.

VAST Data Advantage: Unlocking New Levels of HPC Performance

With its inherent support for embarrassingly parallelism, VAST Data's DASE architecture empowers HPC environments to achieve new performance, efficiency, and reliability levels. By eliminating data bottlenecks, ensuring fault isolation, and reducing management complexity, DASE enables HPC users to tackle larger, more complex workloads easily. This capability accelerates time-to-results and allows for flexible scaling, making DASE a future-ready solution for evolving HPC and AI demands.

Conventional Parallel File Systems - Lustre, and GPFS

In HPC environments, managing large datasets and providing efficient, simultaneous access across compute nodes is crucial for running complex simulations, AI model training, and intensive research workflows. Traditionally, parallel file systems like Lustre and GPFS have been the primary solutions for distributing data storage and enabling multiple nodes to access data concurrently. These systems, each with its unique architecture, have helped HPC environments achieve the scale and performance required to handle massive data demands.

Key Roles of Conventional Parallel File Systems in HPC

Parallel file systems are designed to address the specific challenges of HPC by offering several core capabilities:

- **Data Distribution Across Nodes:** By breaking down files and distributing data blocks across multiple storage nodes, parallel file systems enable multiple clients to read and write simultaneously, reducing bottlenecks and improving access speeds.

- **High Throughput for Data-Intensive Applications:** In fields like genomics, climate modeling, and physics simulations, high-throughput data access is paramount. Parallel file systems optimize performance by allowing many nodes to process data in parallel.

- **Support for Large-Scale Data Environments:** With petabytes of data storage and thousands of nodes, HPC environments demand storage solutions that can expand quickly. Parallel file systems support horizontal scalability, making it possible to add more storage nodes as data and workload demands grow.

Lustre - Architecture and Limitations

Architecture Overview

Lustre is an open-source parallel file system widely used in HPC due to its capability for high throughput. Lustre operates on a shared-nothing architecture, with independently functioning servers designated by function. Its architecture includes:

- **Metadata Server (MDS) and Metadata Target (MDT):** These servers manage metadata, including file names and permissions, essential for efficient file retrieval.

- **Object Storage Server (OSS) and Object Storage Target (OST):** These components handle actual file storage, allowing data to be striped across multiple OSTs to increase parallel access.

- **Management Server (MGS):** The MGS manages the configuration information for all servers, helping coordinate activities across the system.

While Lustre's architecture enables large-scale, parallel data access, it requires local file systems like ZFS or ldiskfs to manage data redundancy and availability, which can add complexity.

Limitations

Lustre, while widely adopted in HPC environments, comes with operational challenges that require specialized expertise and ongoing manual intervention. These limitations include:

- **Metadata Bottlenecks:** Lustre relies on centralized metadata servers that can create a performance bottleneck as file counts grow, slowing down access times for large datasets.

- **Manual Striping & Tuning:** Optimizing Lustre performance requires careful striping decisions at the file level—an additional burden for administrators.

- **Upgrade & Compatibility Issues:** Lustre's frequent updates require recompiling kernel modules and adjusting storage backends, which can introduce disruptions.

VAST Data Advantage

VAST Data's DASE architecture eliminates Lustre's complexities by offering a single, high-performance storage layer with built-in high availability, enterprise features, and low-latency data access without the need for extensive tuning or

configuration. VAST does this by:

- **No Centralized Metadata Bottlenecks:** VAST's disaggregated architecture distributes metadata dynamically.

- **No Manual Striping Required:** VAST automates data placement with intelligent, high-performance data distribution.

- **Non-Disruptive Software Updates:** VAST's platform allows seamless updates without requiring compute downtime.

GPFS – Architecture and Limitations

Architecture Overview

GPFS is an enterprise-grade parallel file system that uses a Network Shared Disk (NSD) model, allowing multiple nodes to access shared storage resources. Key components include:

- **Cluster Manager:** Manages configuration and coordinates the namespace across nodes.

- **Administrative Nodes:** Oversee configuration, monitoring, and resource allocation.

- **Clients:** Access data in the GPFS cluster and manage file operations, which can introduce performance dependencies.

GPFS also allows for customized block sizes and pools, optimizing performance for different workload requirements, and includes features like snapshots, encryption, and replication for enterprise use.

Limitations

GPFS, while offering robust enterprise-level parallel storage, comes with administrative overhead and tuning requirements that can impact efficiency:

- **Client-Driven Metadata Processing:** Unlike disaggregated architectures, GPFS relies on client nodes to process metadata, leading to performance degradation if metadata-heavy workloads increase.

- **Manual Caching Policies:** GPFS requires tuning cache sizes and block allocation to optimize performance—a process that varies across workloads and can result in inefficiencies.

- **Replication Overhead:** GPFS achieves resiliency through multi-site replication, increasing storage capacity demands and complexity.

VAST Data Advantage

With the VAST Data Platform unified architecture, administrators can avoid GPFS's operational complexities. VAST Data's built-in high availability and multiprotocol support enable seamless cross-site access, simplifying multi-cluster management and ensuring reliable performance across all workloads.

- **No Manual Caching Policies:** VAST automates performance tuning.

- **Lower storage overhead:** VAST reduces reliance on complex replication models through global data resilience.

- **Optimized for HPC + AI:** GPFS was designed for HPC first, whereas VAST natively supports mixed HPC and AI workloads.

Conventional Parallel File Systems - Complexity and Constraints

While Lustre and GPFS have served HPC well by enabling these capabilities, they introduce significant architectural complexity and operational challenges. Each system relies on unique components and architectures that add to the administrative burden and limit flexibility in evolving workloads.

1. **Metadata Management Overhead:** In parallel file systems, metadata servers handle data about file locations, permissions, and access patterns. Managing metadata efficiently is critical but challenging, as it can become a bottleneck when multiple users access files concurrently. In Lustre, for instance, metadata management requires specialized tuning, while in GPFS, client nodes play a role in managing metadata, complicating failover and recovery.

2. **Dependency on Specialized Hardware and Configurations:** Many conventional parallel file systems rely on dedicated servers for different functions, like object storage and metadata, leading to higher infrastructure costs. High availability in these systems often requires additional third-party tools and configurations, increasing both setup complexity and total cost of ownership.

3. **Complex Failover Mechanisms:** Ensuring data availability in case of hardware or network failure is essential in HPC. Parallel file systems often have technically demanding failover mechanisms, and in some cases, they rely on replication across nodes or clusters, which adds latency and reduces overall system efficiency.

4. **Management and Tuning Requirements:** Achieving optimal performance with Lustre or GPFS requires constant tuning of parameters like data striping, cache size, and read-ahead settings. These systems typically demand significant HPC expertise for day-to-day operations, making them costly and labor-intensive to maintain. As data volumes grow, these tuning requirements can intensify, making it challenging to keep up with evolving workloads.

5. **Challenges with Mixed Workloads and AI:** Conventional parallel file systems were initially designed to support HPC workloads with large, sequential read/write patterns. However, modern HPC environments increasingly incorporate AI and machine learning workloads involving smaller, random I/O operations. These file systems struggle to handle mixed workloads efficiently, resulting in performance degradation and underutilization of resources.

Transition to Simplified, Unified Storage with DASE

The VAST Data Platform introduces a revolutionary approach to HPC data storage that is fundamentally different from conventional parallel file systems like Lustre and GPFS. While historically integral to HPC environments, these legacy

systems are built on architectures that often require complex configuration, maintenance, and tuning to achieve optimal performance. VAST Data's Disaggregated Shared Everything (DASE) architecture overcomes these limitations with a streamlined, high-performance solution that provides unified, low-latency data access for mixed workloads, including HPC and AI.

Unique Advantages of VAST Data's DASE Architecture

1. **Disaggregated Shared Everything (DASE)**
   Unlike conventional parallel file systems that rely on tightly coupled compute and storage resources, DASE allows storage and compute to scale independently. This architecture eliminates the need for data movement across different storage tiers, which reduces administrative overhead and enables more efficient workflows. Each storage node in DASE can access data across the cluster directly, ensuring high throughput and minimizing bottlenecks.

2. **Unified Storage Layer with Multiprotocol Access**
   VAST Data's platform natively supports NFS, SMB, S3, and Block without needing gateway-based access. This native multiprotocol support means organizations can use a single storage layer across HPC, AI, and conventional workloads. By contrast, conventional parallel file systems often require data copies or gateways to enable multiprotocol access, which adds latency and reduces scalability.

3. **High-Performance NVMe-over-Fabrics (NVMe-oF) with RDMA**
   VAST's use of NVMe-oF with RDMA enables low-latency, high-speed data access directly over the network, bypassing CPU involvement in data transfers. This feature provides significantly faster data access than conventional file systems, often constrained by legacy networking protocols. This setup is especially advantageous for AI-driven workloads, where small, random I/O is common and requires high IOPS performance.

4. **Built-In Data Protection and Simplified Management**
   VAST's platform includes built-in features like snapshots, deduplication, and replication, providing robust data protection without needing third-party tools. Conventional systems often lack these capabilities or require complex integrations, adding to the administrative burden. With VAST, administrators can manage data protection seamlessly, reducing time spent on routine tasks and enabling them to focus on strategic improvements.

5. **Optimized for Mixed Workloads (HPC + AI)**
   Many parallel file systems are optimized primarily for sequential read and write patterns, common in conventional HPC tasks but less suitable for AI workloads, often involving random I/O operations. VAST Data's architecture is optimized for both, providing the flexibility to support simultaneous HPC and AI workloads without compromising performance.

VAST Data vs. Conventional Parallel File Systems Comparison Table

Feature	VAST Data (DASE)	Lustre	GPFS
Architecture	Disaggregated Shared Everything	Shared-Nothing	Network Shared Disk (NSD)
Scalability	Independent Storage/ Compute	Complex Tuning Required	Limited by Client Dependencies
Data Access Protocols	NFS, SMB, S3, Block (Native)	Primarily POSIX	POSIX, NFS via Gateways
Data Movement	No Pre-Movement Required	Requires Tiering	Manual Across Clusters
Network Protocol	NVMe-oF with RDMA	TCP/IP with Tuning	Mixed, Requires Tuning
Performance Optimization	Native Support for HPC and AI	Sequential I/O Optimized	HPC Optimized, Limited AI
Data Protection	Built-In Snapshots, Deduplication	Basic, Lacks Snapshots	Snapshots, Encryption
Administrative Complexity	Simplified, Centralized Control	High; Tuning, Striping	Moderate; Multi-Cluster Config

Summary of HPC Storage Redefined with VAST Data

Conventional parallel file systems like Lustre and GPFS have enabled HPC data access at scale but have inherent complexities impacting management, tuning, and scalability. VAST Data's DASE architecture and NVMe-oF with RDMA technology offer a simplified yet powerful alternative, supporting high-throughput access, real-time data processing, and multiprotocol compatibility without the conventional operational overhead. This streamlined approach meets the demands of modern HPC, allowing organizations to increase productivity and focus on innovation rather than storage management.

VAST Data redefines HPC storage by eliminating the need for data pre-movement, consolidating scratch and long-term storage needs into a single layer, and providing consistent, low-latency access across workflows. This unified platform reduces time and complexity ensuring that researchers, engineers, and data scientists have the seamless data access required for advanced research, AI model training, and complex simulations. In doing so, VAST Data delivers a future-ready solution that enhances the effectiveness of HPC environments and supports the next generation of scientific discovery and industrial advancement.

In the next chapter, we'll explore broader challenges in the HPC space, including data movement and gravity issues, along with how VAST Data provides a modern solution to meet these demands.

Chapter 3: Administering the VAST Data Platform for HPC

Simplifying HPC Platform Administration

The efficient administration of the VAST Data Platform is essential for ensuring optimal performance, accessibility, and scalability in HPC environments. As HPC workloads demand high throughput, real-time data access, and robust failover capabilities, administrators benefit from the VAST platform's simplified management and intuitive control features, which reduce the complexity and overhead typically associated with conventional parallel file systems.

This chapter will cover essential tasks and concepts for administering VAST in an HPC environment, including system configuration, permissions management, network setup, and basic troubleshooting. We'll also discuss some differences between administering the VAST Data Platform and conventional HPC storage solutions like Lustre and GPFS, highlighting VAST's advantages in simplicity and scalability.

Basic System Administration

The VAST Data Platform is designed to simplify storage administration, making it accessible to administrators even without extensive HPC-specific experience. Key components in the VAST architecture, such as storage clusters, data management settings, and protocol configurations, are unified within a single interface, allowing for centralized control over the system.

Core Administrative Tasks

- **Configuring Storage Clusters:** VAST storage clusters can accommodate specific workload requirements, such as data retention policies, replication, and access protocols (NFS, SMB, S3, and Block). Administrators can create storage policies and assign roles to different datasets according to performance and availability needs.

- **Managing Global Namespace:** One of the VAST platform's key features is its Global Namespace, which allows data to be accessible across various locations without needing to move data physically. Administrators can set access permissions and define data residency requirements across this namespace, ensuring efficient access without duplicating datasets.

- **Snapshot and Backup Management:** VAST supports space-efficient snapshots that allow administrators to capture and restore data states without impacting system performance. Configuring regular snapshots helps safeguard against data loss and enables fast recovery in case of an error or data corruption.

System Monitoring and Maintenance

The VAST Data Platform provides real-time system monitoring and performance metrics, helping administrators track key metrics such as data throughput, latency, and storage utilization. The platform's intuitive dashboard offers a consolidated system view, highlighting areas that may require attention, such as nodes nearing storage capacity or applications experiencing high latency.

- **Real-Time Alerts:** Administrators can set up alerts for specific thresholds, such as storage capacity limits or latency issues, allowing for proactive intervention.

- **Performance Tuning:** While VAST minimizes the need for extensive tuning, administrators can adjust caching policies and replication settings to fine-tune performance based on workload requirements.

Permissions and Access Management

Permissions management is critical for ensuring data security and accessibility in an HPC environment, where multiple users and teams may need to access specific datasets simultaneously. The VAST Data Platform includes built-in permission controls to help administrators manage user and group access efficiently.

User and Group Permissions

- **Role-Based Access Control (RBAC):** VAST supports RBAC, allowing administrators to assign permissions based on user roles and groups. This ensures users only access the datasets and system functions relevant to their work.

- **Multi-Tenancy Support:** VAST's platform enables multi-tenancy, allowing different user groups (such as research teams or departments) to securely access shared resources without compromising data privacy. This feature is particularly useful in HPC environments where multiple research teams may need isolated yet secure access to shared data.

Data Security and Compliance

- **Encryption:** The VAST Data Platform supports at-rest and in-transit encryption, helping secure sensitive data against unauthorized access. This is particularly important for Healthcare and Financial Research data, where compliance requirements demand strict data security measures.

- **Audit and Logging:** Administrators can enable detailed audit logging to track data access patterns, user actions, and administrative changes. These logs provide a comprehensive view of data interactions and are helpful for compliance and security monitoring.

Network Configuration and Optimization

Networking is essential in HPC environments, where data must move between compute nodes and storage with minimal latency. VAST Data's Disaggregated Shared Everything (DASE) architecture enables efficient network use, whether over Ethernet or InfiniBand while optimizing data access speeds through advanced protocols like RDMA (Remote Direct Memory Access) and NVMe over Fabrics (NVMe-oF).

Network Protocols and Configuration

- **Configuring RDMA and NVMe-oF:** VAST allows administrators to enable RDMA and NVMe-oF over both Ethernet and InfiniBand networks. RDMA enables direct memory access across nodes without CPU intervention, reducing latency, while NVMe-oF extends NVMe performance across networked storage, accelerating data access. Administrators can configure these protocols to optimize data flow and minimize bottlenecks based on workload requirements.

- **Setting Up Multi-Path Networking:** VAST supports multi-path configurations for environments with high redundancy requirements, ensuring continuous data access even if one path fails. This is crucial in HPC, where interruptions in data flow can disrupt compute tasks and delay workflows.

Network Security

- **Network Segmentation and Access Controls:** Administrators can segment network traffic based on user roles, workload requirements, or research departments, reducing the risk of unauthorized access and minimizing the attack surface. This level of control helps secure data flows and maintain system integrity.

- **Firewall Configuration:** To further enhance security, VAST integrates with standard firewall protocols, allowing administrators to control inbound and outbound network traffic and set IP whitelisting or blacklisting as needed.

Troubleshooting and Maintenance

The VAST Data Platform is built for reliability, but administrators may occasionally need to troubleshoot issues related to network latency, data access, or storage capacity. The platform's monitoring and alerting capabilities and automated maintenance features simplify the troubleshooting process.

Proactive Monitoring and Alerts

VAST provides real-time alerts for key performance indicators, enabling administrators to proactively address potential issues before they affect workloads. Alerts are highly configurable, allowing notifications for storage thresholds, latency, or unexpected changes in data access patterns.

Automated Failover and High Availability

VAST's built-in High-Availability (HA) capabilities ensure minimal disruption in the event of hardware or network failures. Unlike conventional HPC storage solutions, which may require third-party solutions for failover, VAST's architecture provides seamless failover within the platform, ensuring continuous data availability.

Basic Troubleshooting Steps

- **Latency Issues:** If latency issues arise, administrators can review performance metrics on the VAST dashboard, checking for any network bottlenecks or high I/O demand from specific applications. Protocol settings (e.g., RDMA, NVMe-oF) can also be adjusted to optimize data access speeds.

- **Storage Capacity Warnings:** For storage nearing capacity, VAST's snapshot management and deduplication capabilities allow administrators to clear space without impacting data integrity. Thanks to the platform's scalability, additional capacity can be added seamlessly.

- **Node Failures:** In case of a node failure, VAST automatically reroutes data access to available nodes. Administrators can review logs to assess the cause of the failure and perform maintenance as necessary, knowing that data availability remains uninterrupted.

Comparison with Conventional HPC Storage Administration

Administering the VAST Data Platform offers significant advantages over conventional parallel file systems, such as Lustre and GPFS, which require extensive tuning and third-party tools for high availability and failover. VAST's integrated approach simplifies many of the challenges commonly associated with HPC storage.

Simplified Management

- **Minimal Tuning Requirements:** Unlike Lustre and GPFS, which demand extensive tuning and configuration adjustments to optimize performance, VAST automates most aspects of data placement and performance tuning. This reduces the learning curve for new administrators and ensures predictable performance across workloads.

- **Integrated Data Management Features:** VAST includes built-in support for snapshots, deduplication, and high availability, eliminating the need for separate tools. This consolidation simplifies the administration process and reduces the total cost of ownership.

Reduced Complexity in Scaling

VAST's architecture allows administrators to scale storage independently of compute resources, adding storage as needed without reconfiguring or balancing file striping. This flexibility contrasts with Lustre and GPFS, which require detailed configuration and resource management as environments scale.

Improved Data Accessibility

VAST's Global Namespace enables seamless access to data across locations, allowing administrators to manage permissions and access without duplicating datasets or creating data silos. Conventional systems often struggle with cross-site access, creating administrative challenges and adding complexity to data-sharing workflows.

Eliminating Bottlenecks for Seamless HPC Operations

For HPC administrators, a top priority is creating an environment that meets the rigorous demands of users without interruptions or performance issues. The VAST Data Platform is engineered to minimize the complexity of storage management in HPC, allowing admins to focus on strategic improvements rather than troubleshooting. VAST Data helps HPC admins deliver a smooth, reliable user experience even in the most demanding environments by streamlining data access, enhancing system reliability, and offering easy scalability.

Key Benefits of the VAST Data Platform for HPC Admins

1. **Unified Storage Management:** Conventional HPC storage often requires admins to manage multiple tiers or separate data silos, complicating access and slowing performance. VAST's Disaggregated Shared Everything (DASE) architecture provides a single, high-performance layer, eliminating the need for complex tiering and manual data movement. This unified storage approach allows admins to oversee a simplified storage system that automatically supports various workflows without manual intervention.

2. **Multiprotocol Support with No Gateways:** VAST's multiprotocol support (NFS, SMB, S3, and Block) allows seamless access across protocols without gateways, eliminating the latency and scalability issues common in competitive systems. Admins benefit from a streamlined setup that minimizes user complaints about delays or limited access, as users can access the same data layer in real-time regardless of their protocol needs.

3. **Built-in Resilience and Fault Tolerance:** VAST's architecture offers built-in data resilience, reducing the risk of downtime or data loss. With high fault tolerance, VAST ensures data availability and system reliability even as workloads scale, minimizing the chances of performance-related complaints from users. This resilience makes the platform highly suitable for HPC environments with unpredictable or intensive workloads.

4. **Rapid, Scalable Performance:** VAST's all-flash architecture delivers consistently high throughput and low latency, making it easier for admins to meet performance expectations even as data demands increase. With NVMe-over-Fabrics and RDMA, VAST provides direct, fast access to data, ensuring that even the most data-intensive applications run smoothly. For HPC admins, this means fewer user complaints and less need for performance tuning as demands grow.

5. **Reduced Complexity with DASE Architecture:** VAST's DASE architecture simplifies storage management by allowing compute and storage resources to scale independently. This flexibility provides admins with an easier way to scale infrastructure in response to user needs without the need for complex configuration or restructuring.

VAST Data Advantage: Keeping Users Productive and Complaint-Free

With the VAST Data Platform, HPC administrators can confidently maintain high performance and reliability, supporting their users' most complex workflows without disruptions. The simplified storage environment, combined with high fault tolerance and real-time data access, reduces the risk of user complaints and gives admins peace of mind that their system is equipped to handle demanding workloads.

Summary of VAST Data Platform Administration for HPC

Administering the VAST Data Platform offers a streamlined experience designed to reduce the operational complexities traditionally associated with HPC storage. Key advantages include centralized control, robust data protection, and scalable performance, all of which empower administrators to manage HPC storage efficiently and effectively.

1. **Centralized Management and Simplified Configuration:** VAST's unified interface consolidates setup, monitoring, and configuration, allowing administrators to manage storage clusters, control permissions, and monitor performance from a single platform. This consolidation significantly reduces the need for complex tuning and allows new administrators to manage the platform with ease.

2. **Optimized Network Performance and Failover:** With support for RDMA, NVMe-oF, and multi-path configurations, VAST ensures that data access remains high-performance and uninterrupted. This setup minimizes downtime risks and allows administrators to provide continuous access to data across HPC environments.

3. **Built-In High Availability and Data Protection:** VAST's integrated snapshots, deduplication, and automated failover eliminate the need for third-party solutions, simplifying administration and ensuring reliable data protection. These features also reduce the need for constant monitoring, allowing administrators to allocate resources to strategic initiatives.

4. **Reduced Complexity Compared to Conventional HPC Storage:** By removing tuning requirements and simplifying scaling processes, VAST's DASE architecture enables independent scaling of storage resources without reconfiguring compute infrastructure. This design minimizes administrative intervention, enhances efficiency, and supports environments with high storage demands.

By streamlining management tasks and reducing the need for manual intervention, the VAST Data Platform enables HPC teams to focus on optimizing workflows and supporting user needs, ensuring high productivity and robust system performance in the most demanding HPC applications.

Chapter 4: Summary and Key Takeaways

HPC Storage Landscape and the VAST Data Advantage

Rising performance demands, increased AI integration, and the need for simplified administration have driven the evolution of HPC storage. As discussed in Chapter 1, conventional parallel file systems like Lustre and GPFS struggle with modern AI-driven workloads, metadata bottlenecks, and manual tuning requirements. Chapter 2 examined how VAST Data's Disaggregated Shared Everything (DASE) architecture eliminates these complexities by scaling storage independently of compute, automating metadata management, and providing NVMe-oF powered high-speed access. Finally, in Chapter 3, we explored how VAST dramatically simplifies HPC administration, eliminating the need for manual striping, caching policies, and gateway-based access management.

The VAST Data Platform transforms HPC infrastructure by eliminating legacy constraints, unifying HPC + AI workloads, and delivering high-throughput performance across Ethernet and InfiniBand networks—all without requiring complex tuning or reconfiguration.

HPC Success Stories – Real-World Impact with VAST Data

The following Success Stories provide an in-depth look at how VAST Data has enabled each organization to meet unique HPC challenges with unmatched performance and flexibility.

University of Utah Center for HPC

Overview: The University of Utah's Center for High-Performance Computing supports diverse research applications, from bioinformatics and climate modeling to engineering simulations. The university needed a storage platform that could handle the demands of various scientific disciplines without burdening HPC administrators with complex storage management.

- **Challenge:** The University of Utah Center for HPC faced increasing demands for high-performance data access across numerous research fields. Their conventional storage solutions created data silos and required extensive management, impacting researchers' ability to share data seamlessly and limiting administrative flexibility in scaling resources for new projects.

- **Solution:** VAST's all-flash storage solution provided the university with a unified platform that supports high-performance, multiprotocol access across all research disciplines. With VAST's DASE architecture, the university simplified data management, enabling multiple users and departments to access a single data pool in real-time, regardless of workload requirements.

- **Outcome:** The adoption of VAST Data enabled the University of Utah's HPC center to accelerate research across departments by improving data sharing and reducing latency. Researchers gained consistent, high-speed access to data, fostering interdisciplinary collaboration. VAST's architecture also freed HPC administrators from the challenges of managing complex storage tiers, allowing them to focus on expanding and enhancing the HPC environment for future research needs.

Texas Advanced Computing Center (TACC)

Overview: The Texas Advanced Computing Center (TACC) at the University of Texas is one of the leading supercomputing facilities in the world, supporting large-scale scientific applications that serve thousands of researchers. For their new Stampede3 supercomputer, TACC required a high-performance storage solution to power data-intensive applications and enable advanced scientific breakthroughs.

- **Challenge:** TACC's Stampede3 project aimed to deliver next-generation computing for over 10,000 scientists and engineers. Conventional storage solutions risked becoming a bottleneck, as they were not optimized for the simultaneous data access demands of thousands of users running diverse applications. TACC needed a storage platform with extreme scalability and performance to keep pace with Stampede3's computational requirements.

- **Solution:** TACC selected VAST Data's storage platform to support Stampede3's data access needs. With VAST's scalable, all-flash storage and high-throughput DASE architecture, TACC achieved the performance necessary to handle intense, parallel HPC workflows. VAST's platform allowed Stampede3 to deliver multiprotocol data access without compromising speed, reliability, or scalability.

- **Outcome:** Integrating VAST Data into TACC's Stampede3 supercomputer enabled TACC to deliver on-demand, high-speed data access to thousands of users across various disciplines, from astrophysics to biomedical research. The result is a flexible, reliable storage solution that supports diverse scientific applications at scale, empowering researchers with fast access to critical data and facilitating cutting-edge discoveries.

DUG Technology

Overview: DUG Technology, a leader in geoscience and high-performance computing, provides HPC-as-a-service, focusing on seismic data analysis for energy companies. With data demands continuously rising, DUG needed a storage platform to handle petabytes of seismic data while maintaining high performance and reliability.

- **Challenge:** DUG's clients rely on rapid, precise seismic data analysis for critical energy exploration. Conventional storage solutions struggled with the high I/O demands and data movement, slowing down service delivery. DUG needed an HPC storage solution to support ultra-high throughput and scale with their data growth.

- **Solution:** DUG implemented VAST Data's all-flash storage platform, which offers massive scalability, high throughput, and low latency. VAST's DASE architecture allowed DUG to eliminate data silos and enable direct data

access across their infrastructure, supporting their extensive HPC workflows without requiring data migration or pre-processing.

- **Outcome:** By transitioning to VAST's platform, DUG achieved exascale computing capabilities, significantly reducing processing times for seismic analysis. This upgrade allowed DUG to provide on-demand, scalable HPC services to clients, helping them improve operational efficiencies in data-intensive projects. With VAST, DUG can handle increasingly complex datasets and maintain a competitive edge in HPC services for geosciences.

The success stories from the University of Utah, TACC, and DUG Technology illustrate how VAST Data empowers modern HPC environments with unparalleled storage performance, simplicity, and scalability. Whether accelerating genomic research, astrophysics simulations, financial modeling, or seismic data analysis, organizations worldwide are choosing VAST to break free from operational constraints present in conventional parallel file systems like Lustre and GPFS. As HPC continues to evolve, VAST ensures that storage is no longer a limitation but a competitive advantage that drives research, innovation, and discovery forward.

FAQs: Common Questions About HPC with VAST Data

This FAQ section addresses HPC professionals' most common questions when evaluating VAST Data as an alternative to conventional parallel file systems.

General Questions

How does VAST Data differ from conventional HPC storage like Lustre or GPFS?

- VAST eliminates tuning complexity, manual striping, and metadata bottlenecks seen in Lustre and GPFS.

- It provides all-flash, NVMe-oF performance without requiring POSIX-specific constraints.

- HPC and AI workloads run on a single platform without requiring separate storage tiers.

How does VAST Data handle AI-driven workloads differently than conventional HPC storage?

- Conventional parallel file systems were built for large, sequential HPC workloads—they struggle with AI's small, frequent checkpointing writes.

- VAST is built for HPC and AI together, eliminating metadata congestion and performance degradation.

Can VAST Data be used in hybrid cloud or multi-site HPC environments?

- Yes. VAST's Global Namespace enables cross-site data accessibility without pre-movement.

- Supports both on-premises and cloud-based HPC environments with no workflow disruptions.

Performance & Scalability

Does VAST Data require manual tuning like Lustre or GPFS?

- No. VAST automates all performance tuning, metadata management, and data distribution.

- Conventional parallel file systems require striping, caching, and tiering adjustments—VAST eliminates this complexity.

How does VAST scale compared to conventional parallel file systems?

- VAST scales independently, adding storage without reconfiguring.

- Conventional storage requires manual rebalancing and tuning as clusters grow—VAST does not.

Networking & Integration

Does VAST Data require InfiniBand, or can it run on Ethernet?

- VAST supports both InfiniBand and Ethernet natively, with no tuning required for RDMA-based performance.

- Conventional Lustre and GPFS deployments require manual RDMA adjustments—VAST works out of the-box.

How does VAST Data handle multiprotocol access?

- NFS, SMB, S3, and Block are natively supported—no gateways or additional software layers needed.

- Cross-discipline teams (AI, research, financial modeling) can access the same dataset in real-time without performance degradation.

Administration & Security

How does VAST simplify HPC storage administration?

- Built-in snapshots, automated failover, and no-tiered storage mean less administrative overhead.

- Conventional storage requires dedicated metadata servers, manual striping, and extensive monitoring.

Does VAST support multi-tenancy and permissions management?

Yes. VAST provides Role-Based Access Control (RBAC) and supports secure multi-user HPC environments.

Cost & Efficiency

How does VAST Data lower TCO compared to conventional parallel file systems?

- Eliminates extra hardware for metadata management.

- Reduces administrative overhead and tuning complexity.

- Consolidates multiple storage tiers into a single, scalable layer.

What industries benefit most from VAST Data for HPC?

- Life Sciences & Healthcare (Genomics, Medical Imaging, Precision Medicine).

- Higher Education & Research (AI-driven simulations, cross-discipline collaboration).

- Financial Services (Risk modeling, AI-enhanced data analytics).

- Energy & Engineering (Seismic data processing, climate modeling).

Conclusion

High-Performance Computing (HPC) has long been a cornerstone of scientific discovery, engineering innovation, and AI-driven research. However, as workloads become more data-intensive and interdisciplinary, conventional storage architectures like Lustre and GPFS are increasingly inefficient, complex, and challenging to scale. This guide has explored the key infrastructure, networking, and administration challenges that burden HPC storage.

Chapter 1 examined how networking choices like Ethernet and InfiniBand impact HPC performance and why conventional parallel file systems require constant tuning to maintain efficiency. Chapter 2 explored how VAST Data's Disaggregated Shared Everything (DASE) architecture eliminates these limitations, delivering all-flash NVMe-oF storage optimized for both HPC and AI workloads. Chapter 3 highlighted how VAST simplifies administration by reducing the overhead of managing metadata servers, data movement, and performance tuning.

As HPC and AI workloads continue to converge, organizations must move beyond legacy architectures that hinder innovation and increase operational complexity. The VAST Data Platform delivers the scalability, simplicity, and high-performance storage that modern research and AI-driven enterprises demand. By eliminating the bottlenecks of Lustre and GPFS, VAST empowers HPC administrators, researchers, and engineers to focus on accelerating discoveries, optimizing performance, and solving the world's most complex computational challenges.

Are you ready to accelerate your HPC transformation? Contact VAST Data today at hello@vastdata.com to learn how our innovative platform simplifies data management, maximizes performance, and future-proofs your most demanding workloads.