OPEN COMPOSABLE COMPATIBILITY LAB

Technical Brief

BeeGFS Tiered Storage with OpenFlex® Data24-4200 and Ultrastar® Data60: Performance and Compatibility in a Disaggregated Architecture

Overview

High-performance computing (HPC) and data-intensive workloads require storage architectures that can efficiently balance performance, scalability, and cost. Tiered storage is a proven approach to address these requirements by combining high-performance flash storage for active data with high-capacity disk storage for less latency-sensitive workloads.

This technical brief presents the results of a WD proof of concept (PoC) validating BeeGFS tiered storage across a disaggregated architecture using OpenFlex Data24 4200 NVMe-oF™ storage platform and Ultrastar Data60 HDD-based hybrid storage platform. The configuration demonstrates compatibility between BeeGFS and WD platforms while evaluating performance characteristics of both storage tiers and the efficiency of data movement between them.

The PoC environment integrates NVMe™-based high-performance nodes with SAS-based high-capacity nodes under a unified BeeGFS file system. Benchmark testing shows that the NVMe tier delivers high aggregate throughput, with read performance exceeding 90 GiB/s and write performance approaching 88 GiB/s. In contrast, the capacity tier provides cost-efficient storage, enabling a balanced and scalable tiered architecture.

In addition to raw performance validation, the PoC highlights the importance of efficient data movement between tiers. Native BeeGFS data migration tools demonstrated significantly higher throughput compared to traditional Linux copy methods, achieving up to an order of magnitude improvement in transfer performance.

Results also indicate that overall system performance is influenced by factors such as network bandwidth, workload parallelism, and storage media characteristics. High-performance tiers were primarily network-bound at scale, while capacity tiers were constrained by disk performance, reinforcing the need for intelligent tiering strategies.

The findings confirm that BeeGFS can effectively leverage WD OpenFlex Data24 4200 and Ultrastar Data60 platforms to enable a scalable, high-performance, and cost-efficient tiered storage solution. This architecture supports a wide range of HPC and AI workloads by providing fast access to active data while maintaining economical storage for large datasets.

The following sections describe the platform components, architecture, and performance results that validate this approach.

OpenFlex Data24 4000 Series Storage Platform

The OpenFlex Data24 4200 is a shared NVMe-oF storage platform built for workloads that need very high bandwidth and low latency. It takes NVMe SSD performance and makes it available to multiple servers over Ethernet. This lets teams treat flash as a pooled resource rather than something trapped inside individual compute nodes.

At a high level, the Data24 4000 series uses NVMe over Fabrics (NVMe-oF) so hosts can access NVMe SSDs across the network with latency designed to mimic direct attached storage. The platform uses WD RapidFlex™ A2000 NVMe-oF controllers, which provides up to 12 ports of 100Gb Ethernet and supports host connectivity using RDMA or TCP. It also supports a small, direct-attached design where up to six dual-pathed hosts can be connected without a switch.

The Data24 4200 is a 24 drive NVMe 2U chassis designed for high-density and parallel access. The platform includes dual I/O modules for parallel paths to storage. It supports multiple fabric connections, which deliver high aggregate bandwidth while keeping latency predictable. It also supports dual-ported, vendor-agnostic NVMe SSDs, which reinforces an open design and avoids lock-in to a single drive vendor.

BeeGFS Overview

BeeGFS is a high-performance, software-defined parallel file system designed for data-intensive workloads in HPC, AI, and analytics environments. It enables concurrent access from multiple clients by distributing data and metadata across multiple services and storage nodes, allowing the system to scale performance and capacity independently.

The architecture separates management, metadata, and storage services, providing flexibility to scale each component based on workload requirements. Data is striped across multiple storage targets to enable parallel I/O and high throughput, while distributed metadata services ensure efficient handling of file operations at scale.

BeeGFS integrates well with disaggregated storage architectures and supports high-speed Ethernet and RDMA networking, allowing it to take full advantage of NVMe-over-Fabrics platforms such as OpenFlex Data24. It presents a single POSIX-compliant namespace across both performance and capacity tiers, enabling transparent access to data regardless of underlying storage media.

Ultrastar Data60 Overview

The Ultrastar Data60 is a high-density Just a Bunch of Drives (JBOD) platform used to add large amounts of HDD capacity in a small rack footprint. In this tiering architecture, it represents the capacity tier that complements the NVMe-based performance tier.

The Data60 is part of the Ultrastar JBOD family, with a 4U form factor that supports up to 60 HDDs. It is positioned for enterprise and big data storage use cases, including disaggregated storage for software-defined storage environments, large scale backup, and archive deployments. The platform supports both high availability connectivity and lower cost options, depending on how the enclosure is deployed and what the storage software expects.

AI and HPC workflows create a mix of data temperatures. Some data needs flash class performance, and much more data needs economical capacity for retention and reuse. Within a tiered design, the Data60 provides the warm and cold capacity layer. It is a good fit for datasets that must stay online and accessible, but do not need NVMe latency. This separation helps keep the NVMe tier focused on active work, while the HDD tier absorbs growth in project data, intermediate results, and long-lived retention datasets.

The Data60 is typically attached to storage servers using external SAS connectivity. Those storage servers then present the HDD capacity to the file system as a managed pool. This keeps the architecture modular. It also makes it easier to scale capacity by adding enclosures without changing the NVMe tier design.

Hardware Topology

With these platform components in place, the following section outlines how they are combined in a unified tiered storage architecture. The solution combines a disaggregated NVMe performance tier with a high-capacity HDD tier under a single BeeGFS file system. The architecture was designed to demonstrate compatibility between BeeGFS and WD storage platforms while also showing how high-performance and high-capacity storage can be combined in one scalable environment. The solution separates storage, compute, and networking so that each layer can be scaled independently as workload requirements change.

HDD Storage System

- 1 × Ultrastar Data60
- 60 × Ultrastar DC HC550 18TB HDDs (30 drives used for this configuration)

Client Nodes

- 4 × Dell PowerEdge R750 servers
- NICs: 2 × NVIDIA ConnectX-7

Switching Fabric

- Mellanox® SN4600 Ethernet switch
- 64 ports, up to 200GbE

High-Performance Nodes (NVMe Tier)

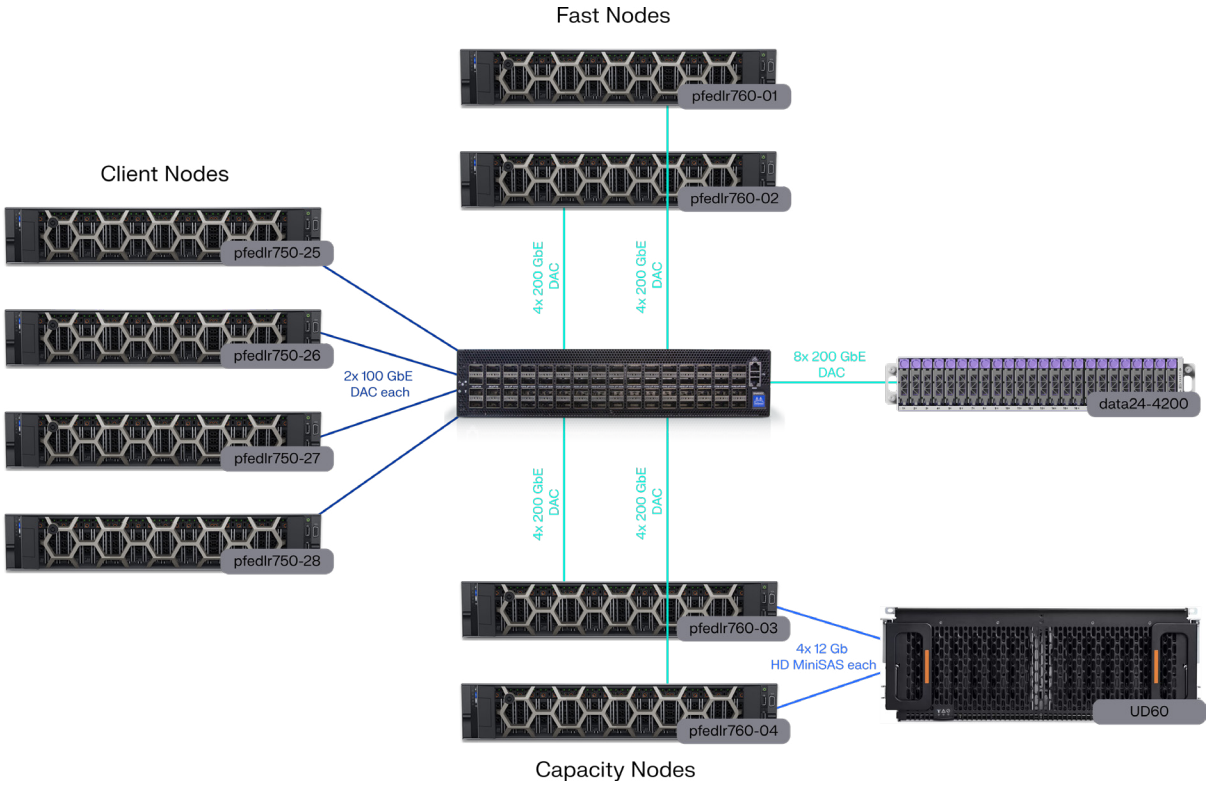
- 2 × Dell® PowerEdge® R760 servers
- NICs: 2 × NVIDIA ConnectX®-7

NVMe Storage System

- 1 × OpenFlex Data24 4200
- 24 × 15.36TB NVMe SSDs (DapuStor R5100D)

Capacity Nodes (HDD Tier)

- 2 × Dell PowerEdge R760 servers
- NICs: 2 × NVIDIA ConnectX-7
- SAS HBA: 2 × Broadcom® 9600-16e



BeeGFS Topology

For BeeGFS storage allocation, the implementation used five NVMe devices per BeeGFS storage service, resulting in ten NVMe devices per server and twenty NVMe devices allocated as BeeGFS storage targets across the two high-performance servers. In addition, two NVMe devices were partitioned for metadata, creating four metadata partitions in total, with one partition assigned to each BeeGFS metadata service. Two NVMe drives remained unallocated during the main test configuration. Those spare drives were used for supplemental testing and would be strong candidates for metadata redundancy in a production design.

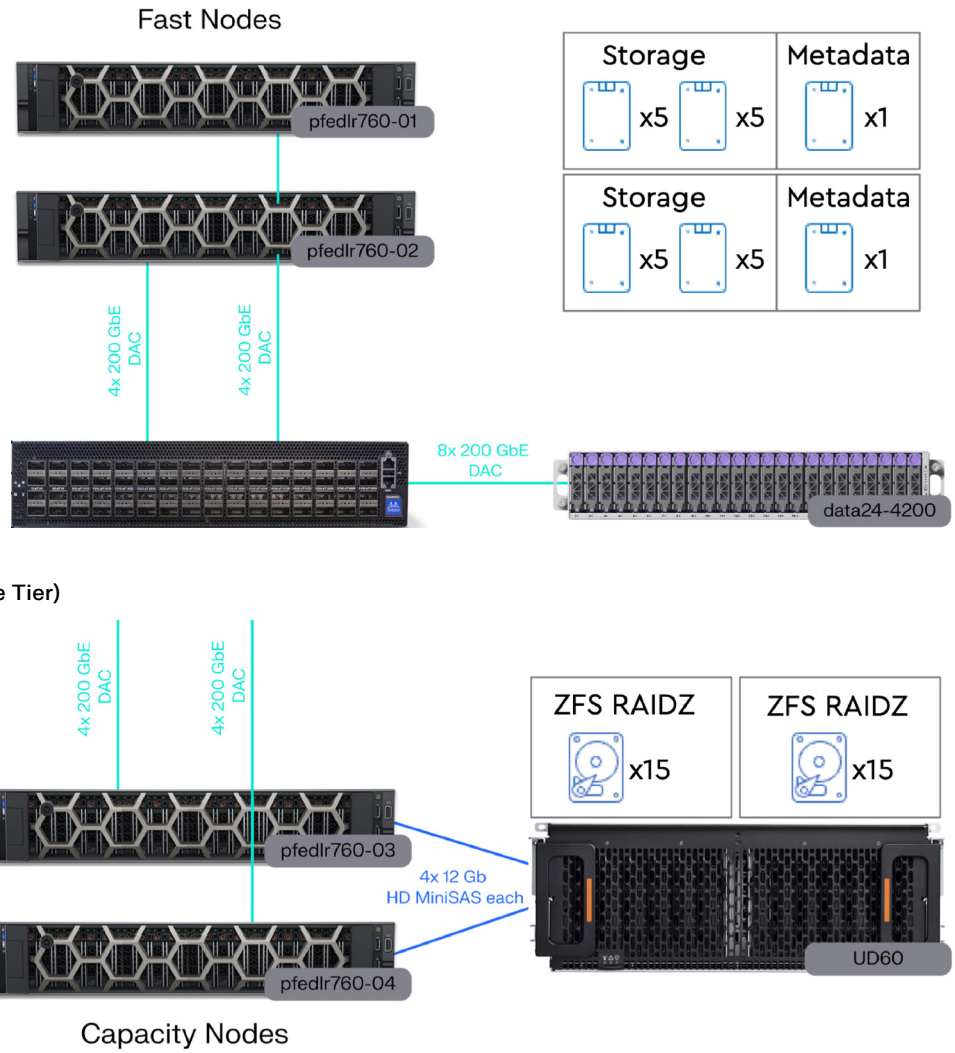
The capacity tier was built on two additional Dell PowerEdge R760 servers to a Ultrastar Data60 enclosure. Each capacity server ran two BeeGFS storage services, for a total of four capacity-tier storage services. The Ultrastar Data60 provided dense HDD-based expansion for the BeeGFS file system, and in this configuration the storage nodes used ZFS RAID-Z pools built from 15 drives each.

Management Service

- 1 × BeeGFS management (mgmtd) service
- Provides cluster configuration and service coordination

Metadata Services

- 4 × BeeGFS metadata services
- 2 NVMe drives partitioned for metadata
- 4 total metadata partitions (1 per metadata service)
- Distributed across high-performance nodes
- Handle namespace operations and file metadata



High-Performance Storage Services (NVMe Tier)

- 4 × BeeGFS storage services
- Deployed on 2 high-performance nodes
- 5 NVMe drives per storage service (20 total targets)
- Provide high-throughput access to NVMe storage
- 2 NVMe drives reserved for testing or redundancy

Capacity Storage Services (HDD Tier)

- 4 × BeeGFS storage services
- Deployed on 2 capacity nodes
- Backed by ZFS RAID-Z pools of HDDs
- Provide high-capacity storage tier

Client Access

- 4 × BeeGFS clients
- Parallel access to all metadata and storage services
- Enable distributed I/O across both tiers

Data Distribution

- Data striped across multiple storage targets
- Metadata distributed across multiple services
- Enables scalable throughput and avoids single-node bottlenecks

Performance

With the system architecture defined, performance testing was conducted to validate scalability, throughput, and data movement behavior across both storage tiers. Performance testing was conducted to evaluate the scalability and throughput of the BeeGFS file system across both NVMe and HDD tiers. The following sections highlight how the system behaves under increasing levels of parallelism and different workload characteristics.

BeeGFS Bench Results

The BeeGFS benchmark tests measure raw file system throughput by generating I/O directly from the storage nodes, providing a clear view of how the system scales under increasing levels of parallelism. These results highlight the efficiency of BeeGFS in distributing workloads across multiple storage targets and services, particularly within the NVMe performance tier.

READ

```
beegfs bench start -D -n 24 -b 512kiB -s 10GiB --nodes 15,25,35,45 -r
| READ Test Summary |
| METRIC    | THROUGHPUT | TARGET ID | TARGET ALIAS       | ON NODE          |
| Minimum   | 4.41GiB/s  | s:205     | target_storage_205 | node_storage_25  |
| Maximum   | 4.68GiB/s  | s:104     | target_storage_104 | node_storage_15  |
| Average   | 4.56GiB/s  | -         | -                  |                  |
| Aggregate | 91.22GiB/s | -         | -                  |                  |
```

WRITE

```
beegfs bench start -D -n 24 -b 512kiB -s 10GiB --nodes 15,25,35,45
| WRITE Test Summary |
| METRIC    | THROUGHPUT | TARGET ID | TARGET ALIAS       | ON NODE          |
| Minimum   | 3.68GiB/s  | s:303     | target_storage_303 | node_storage_35  |
| Maximum   | 4.87GiB/s  | s:201     | target_storage_201 | node_storage_25  |
| Average   | 4.40GiB/s  | -         | -                  |                  |
| Aggregate | 87.93GiB/s | -         | -                  |                  |
```

The BeeGFS bench results for the high-performance NVMe tier deliver strong aggregate throughput and consistent per-target performance across all storage nodes. Read performance achieved an aggregate throughput of 91.22 GiB/s, with individual targets ranging from 4.41 GiB/s to 4.68 GiB/s and an average of 4.56 GiB/s, indicating well-balanced distribution of I/O across the cluster.

Write performance follows a similar pattern, reaching 87.93 GiB/s aggregate throughput, with slightly wider variation per target (3.68 GiB/s to 4.87 GiB/s) and an average of 4.40 GiB/s, which aligns with expected overhead for write operations on flash media. The narrow performance spread between minimum and maximum values across both read and write results demonstrates effective striping and load balancing by BeeGFS, ensuring that no single storage target becomes a bottleneck while maintaining high parallel efficiency across the NVMe-backed services. This enables applications to scale efficiently without being constrained by individual storage nodes.

READ

```
beegfs bench start -D -n 24 -b 512kiB -s 10GiB --nodes 16,17,18,19 -r
| READ Test Summary |
| METRIC    | THROUGHPUT | TARGET ID | TARGET ALIAS       | ON NODE          |
| Minimum   | 1.35GiB/s  | s:501     | target_storage_501 | node_storage_19  |
| Maximum   | 1.49GiB/s  | s:701     | target_storage_701 | node_storage_17  |
| Average   | 1.40GiB/s  | -         | -                  |                  |
| Aggregate | 5.62GiB/s  | -         | -                  |                  |
```

WRITE

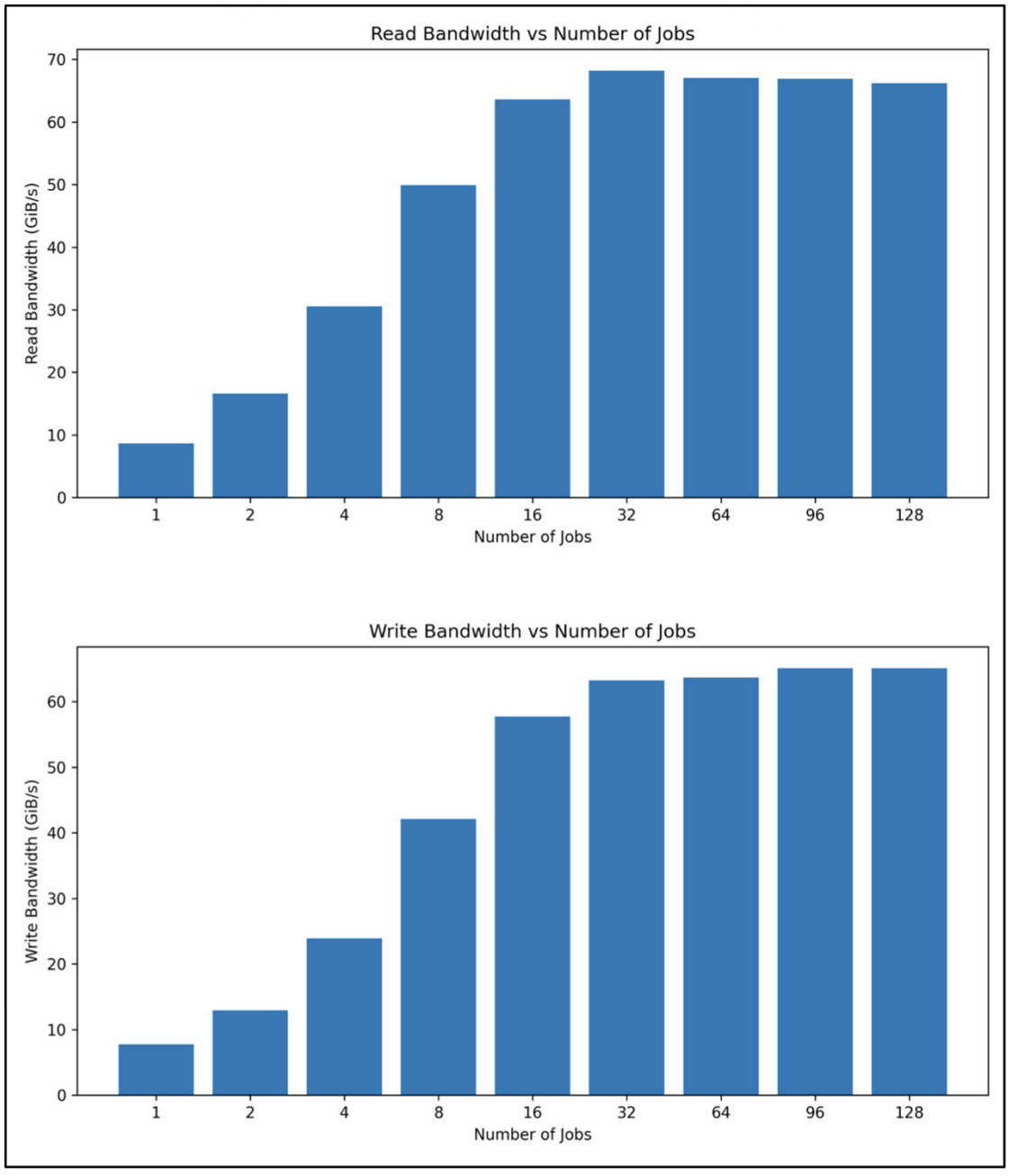
```
beegfs bench start -D -n 24 -b 512kiB -s 10GiB --nodes 16,17,18,19
| WRITE Test Summary |
| METRIC    | THROUGHPUT  | TARGET ID | TARGET ALIAS       | ON NODE          |
| Minimum   | 964.79MiB/s | s:601     | target_storage_601 | node_storage_16  |
| Maximum   | 1.02GiB/s   | s:501     | target_storage_501 | node_storage_19  |
| Average   | 995.74MiB/s | -         | -                  |                  |
| Aggregate | 3.89GiB/s   | -         | -                  |                  |
```

The BeeGFS bench results for the high-capacity HDD tier reflect significantly lower throughput compared to the NVMe tier, as expected, with performance primarily constrained by disk characteristics rather than parallelism or network bandwidth. Read performance achieved an aggregate throughput of 5.62 GiB/s, with individual targets ranging narrowly from 1.35 GiB/s to 1.49 GiB/s and an average of 1.40 GiB/s, indicating consistent behavior across all storage nodes.

Write performance was lower, reaching 3.89 GiB/s aggregate throughput, with per-target performance centered around ~1 GiB/s and minimal variance between nodes. The tight clustering of results across both read and write tests demonstrates balanced workload distribution by BeeGFS, while the overall lower throughput reflects the inherent limitations of HDD-based storage, reinforcing the role of the capacity tier as a cost-efficient solution for less performance-sensitive data.

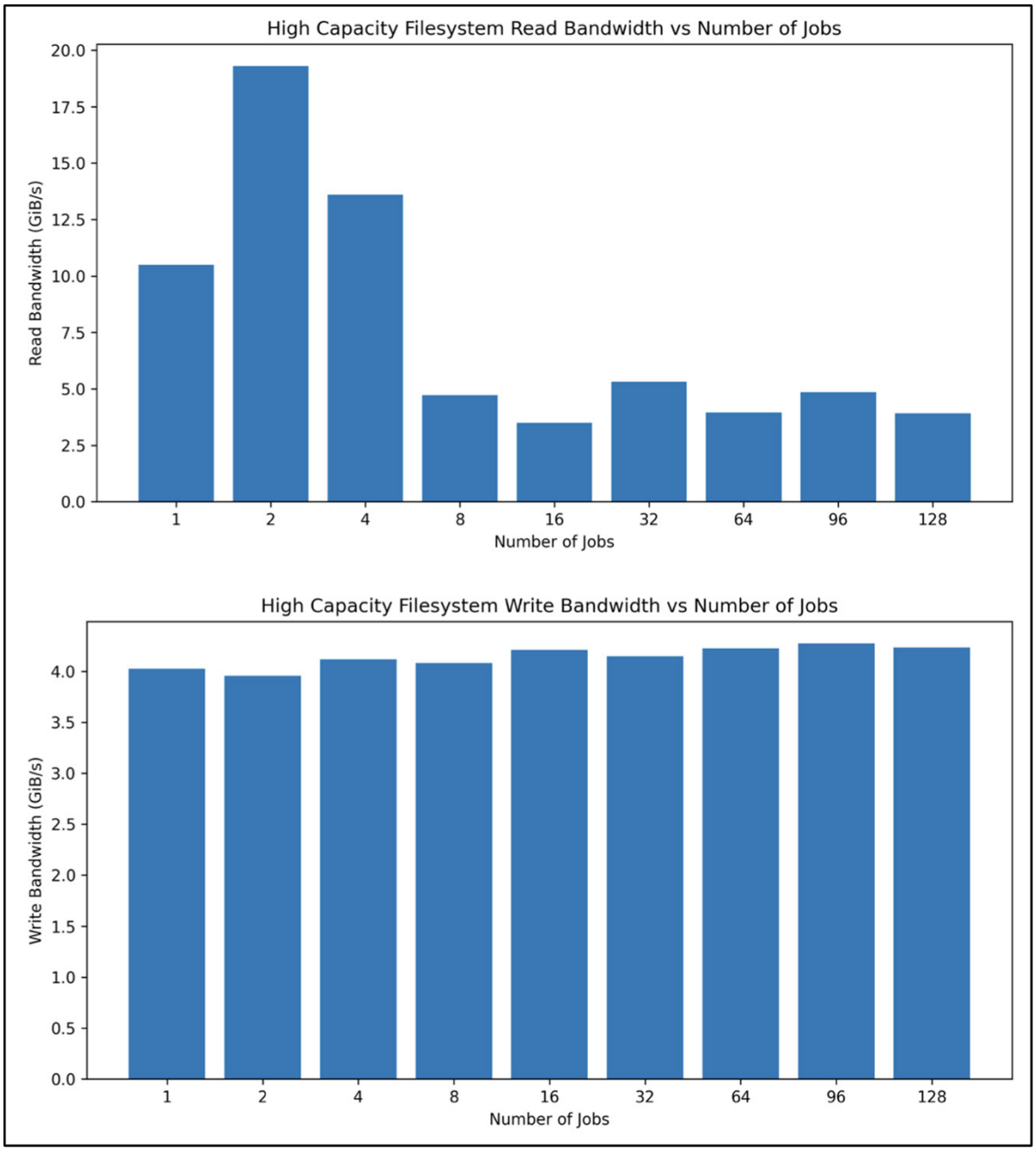
FIO Results

The Flexible I/O Tester (FIO) benchmarking tests were conducted to evaluate filesystem performance under client-driven workloads, complementing the BeeGFS bench results by simulating real application I/O patterns across multiple clients. These tests provide insight into how the system scales with increasing concurrency, highlighting the interaction between storage, network, and client parallelism in sequential workload scenarios



The FIO results for the high-performance NVMe tier show strong scaling characteristics as concurrency increases, with both read and write bandwidth improving steadily with additional jobs until approaching a saturation point. Read throughput scales efficiently from 8.86 GiB/s at low job counts to 68.2 GiB/s at peak. Performance gains begin to level off around 32 jobs, indicating that system limits such as network bandwidth or aggregate storage throughput are being reached.

Write performance follows a similar trend, scaling consistently to approximately 63.2 GiB/s, with slightly lower efficiency compared to reads due to inherent write overhead. Beyond 32 jobs, both read and write performance plateau with minimal gains, demonstrating that the system achieves optimal utilization at moderate concurrency levels while maintaining stable and predictable performance under high load. This behavior makes the NVMe tier well suited for highly concurrent HPC and AI workloads that rely on sustained parallel I/O.



The FIO results for the high-capacity HDD tier highlight the performance limitations of disk-based storage, particularly under increasing concurrency. Read performance shows inconsistent scaling as job count increases, with an initial peak of 19.3 GiB/s at low concurrency followed by a noticeable drop and fluctuation as additional jobs are introduced. This behavior indicates contention at the disk level, where increased parallel access leads to seek overhead and reduced efficiency rather than improved throughput.

In contrast, write performance remains relatively stable across all job counts, maintaining throughput in approximately the 4 GiB/s range with minimal variation. This stability suggests that sequential or buffered write behavior is less sensitive to contention than reads on the HDD tier. Overall, the results reinforce that the capacity tier is optimized for consistent, moderate throughput rather than high parallel scalability, making it well suited for less performance-sensitive data while leaving highly parallel workloads to the NVMe performance tier. This supports the role of the capacity tier as a cost-efficient layer for less performance-sensitive workloads.

MDTest Results

MDTest is a metadata performance benchmarking tool commonly used in HPC environments to evaluate the efficiency of file system operations such as file and directory creation, deletion, stat, and traversal. Unlike throughput-focused tools like FIO or BeeGFS bench, MDTest specifically measures how well the file system handles large volumes of metadata operations under parallel workloads. This makes it particularly valuable for comparing configurations across different systems, as many real-world workloads are heavily influenced by metadata performance rather than raw bandwidth. Metrics such as file creation rates and directory operations can be used as reference indicators, allowing direct comparison against similar architectures, scaling configurations, or competing file system deployments.

```
mpirun --allow-run-as-root -np 128 -host
192.168.2.55:32,192.168.2.56:32,192.168.2.57:32,192.168.2.58:32 /root/ior/src/mdtest
-d /mnt/beegfs/mdtest/ -I 2000 -i 2 -F -D -P -N 1 -t -u -b 8

SUMMARY rate: (of 2 iterations)
Operation Max Min Mean Std Dev
--------- --- --- ---- -------
Directory creation 33988.953 33085.686 33537.320 638.706
Directory stat 372393.985 360754.401 366574.193 8230.429
Directory rename 152458.729 143538.701 147998.715 6307.412
Directory removal 28145.833 27470.514 27808.173 477.523
File creation 155402.036 137462.073 146432.055 12685.469
File stat 635722.974 602269.556 618996.265 23655.138
File read 207909.675 196228.729 202069.202 8259.676
File removal 136992.756 126814.548 131903.652 7197.080
Tree creation 14.710 14.534 14.622 0.125
Tree removal 3.360 2.816 3.088 0.384

SUMMARY time: (of 2 iterations)
Operation Max Min Mean Std Dev
--------- --- --- ---- -------
Directory creation 7.737 7.532 7.635 0.145
Directory stat 0.710 0.687 0.699 0.016
Directory rename 1.783 1.679 1.731 0.074
Directory removal 9.319 9.095 9.207 0.158
File creation 1.862 1.647 1.755 0.152
File stat 0.425 0.403 0.414 0.016
File read 1.305 1.231 1.268 0.052
File removal 2.019 1.869 1.944 0.106
Tree creation 0.069 0.068 0.068 0.001
Tree removal 0.355 0.298 0.326 0.041
```

The MDTest results demonstrate strong metadata performance across the BeeGFS cluster, particularly in file and directory operations that are critical for HPC and AI workloads. File creation rates exceed ~146K operations per second on average, with file stat and read operations reaching even higher levels, indicating that the distributed metadata services are handling parallel workloads efficiently. Directory operations follow similar trends, with consistent performance across creation, stat, rename, and removal, and relatively low standard deviation across all metrics. These results confirm that the metadata layer scales effectively with concurrency and maintains stable performance, which is essential for workloads involving large numbers of small files or frequent metadata access.

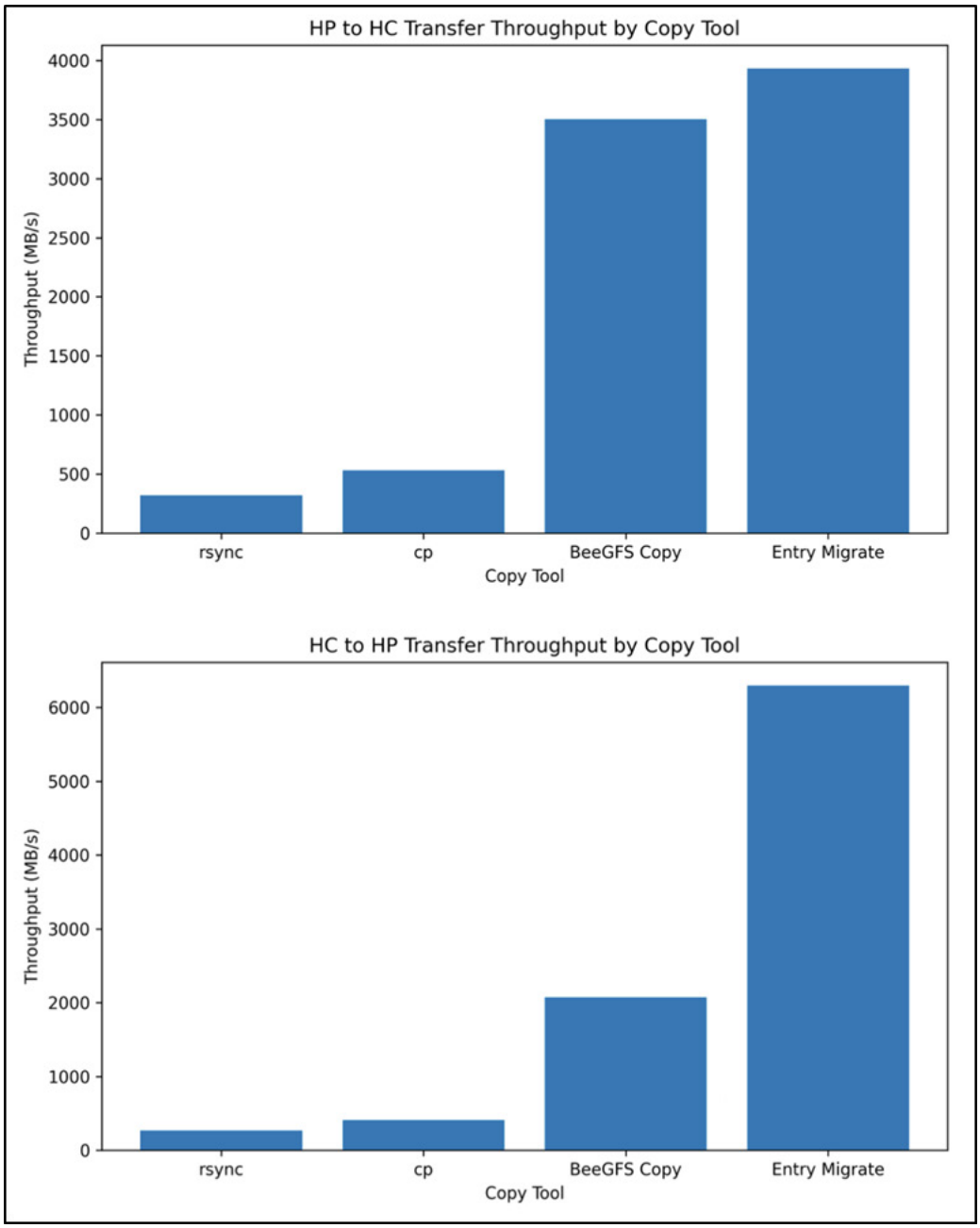
Importantly, these results closely align with the high-capacity node behavior observed earlier because both the NVMe and HDD tiers share the same metadata services in this configuration. As a result, metadata performance remains consistent regardless of the underlying storage tier, reinforcing that metadata operations are decoupled from data path performance. This separation ensures that even when data resides on the HDD tier, metadata-intensive workloads can still benefit from the high-performance characteristics of the NVMe-backed metadata layer. This is particularly important for AI and analytics workflows that involve large numbers of small files and frequent metadata operations.

Data Transfer

To complement the synthetic benchmarks, data transfer tests were conducted to evaluate how the file system performs under common real-world operational workflows. These tests focus on real-world file movement using widely adopted tools such as rsync and cp, alongside native BeeGFS utilities including BeeGFS Entry Migrate and the BeeGFS Copy functionality added in version 8.0. Together, these tools represent a range of usage patterns from simple copy operations to more advanced parallelized data movement within the file system.

These tests provide insight into how efficiently data can be moved within and across tiers in a shared storage environment. By comparing standard Linux utilities with BeeGFS-native capabilities, the results highlight the impact of parallelism, metadata handling, and data placement awareness on transfer performance. This is particularly important in tiered architectures, where data is frequently moved between NVMe and HDD storage to balance performance and capacity while maintaining a unified namespace across the system.

Copy tests were performed with an unstructured dataset of 46 files ranging between 10GB to 100GB with a total size of about 536GB.



For transfers from the high-performance tier to the high-capacity tier, the results show a clear separation between conventional copy methods and BeeGFS-native tools. rsync delivered 322 MB/s and required 27m45s to move the 536GB dataset, while cp improved that to 531 MB/s in 16m49s. In contrast, BeeGFS Copy reached 3501 MB/s and reduced transfer time to 2m26s, while BeeGFS Entry Migrate delivered the best result at 3932 MB/s in 2m16s. These results demonstrate that native BeeGFS methods are dramatically better suited for moving data down to the capacity tier, where preserving file system awareness and using BeeGFS-managed movement provides a major throughput advantage over generic Linux tools.

For transfers from the high-capacity tier back to the high-performance tier, the same pattern holds, but the performance gap becomes even larger. rsync achieved 266 MB/s with a transfer time of 33m29s, and cp improved this to 411 MB/s in 21m45s. BeeGFS Copy increased throughput substantially to 2071 MB/s and completed the transfer in 4m6s, while BeeGFS Entry Migrate again delivered the strongest result at 6297 MB/s in just 1m25s. The particularly strong result for BeeGFS Entry Migrate in the capacity-to-performance direction suggests that BeeGFS-native movement is especially effective when relocating data back onto the faster NVMe tier, where the destination storage is less likely to become the limiting factor.

Taken together, the transfer results show that BeeGFS-native data movement tools provide a substantial operational advantage in a tiered storage environment. Standard file copy tools remain functional for basic transfers, but they operate at only a fraction of the throughput delivered by BeeGFS Copy and BeeGFS Entry Migrate. Across both directions, the BeeGFS-native methods consistently complete the same 536GB transfer in minutes rather than tens of minutes, making them much better aligned with the goals of automated tiering, rapid data relocation, and efficient use of a single multi-tier BeeGFS namespace. This significantly reduces the time required to move data between tiers, enabling faster workflows and more efficient use of storage resources.

High Availability

While high availability (HA) was not implemented as part of this proof of concept due to time and resource constraints, the overall solution architecture is designed with HA in mind. BeeGFS inherently supports distributed services and scale-out deployments, allowing management, metadata, and storage services to run across multiple nodes without introducing single points of failure. This aligns well with disaggregated NVMe and capacity tier architectures, where redundancy and service distribution can be extended across both high-performance and high-capacity tiers.

Enabling HA in this environment would be straightforward through the integration of clustering frameworks such as Corosync and Pacemaker. These tools can manage service failover, node health monitoring, and resource orchestration across both storage tiers, ensuring continuous availability in the event of node or component failure. When combined with BeeGFS service distribution, this approach allows the solution to evolve into a robust, production-ready tiered storage platform capable of maintaining data availability and operational continuity without requiring significant architectural changes.

Conclusions

This PoC demonstrates how BeeGFS with a disaggregated NVMe performance tier and HDD-based capacity tier delivers a scalable, high-performance storage solution for HPC and AI workloads. Across all benchmark categories, the system showed strong parallel scaling behavior. The NVMe tier delivered high aggregate throughput, while the BeeGFS architecture efficiently distributed both data and metadata workloads. The results highlight how a tiered design can balance performance and cost, ensuring that active data benefits from flash-level speed while larger datasets are maintained on economical, high-capacity storage.

The performance results further reinforce the advantages of BeeGFS in a disaggregated environment. Synthetic benchmarks demonstrated predictable scaling as concurrency increased, while metadata testing confirmed the ability to handle large volumes of file operations efficiently. Real-world transfer tests provided additional validation, showing that BeeGFS-native tools such as BeeGFS Copy and BeeGFS Entry Migrate deliver significantly higher throughput compared to traditional file copy methods. These results reinforce the importance of using file system-aware mechanisms for data movement, particularly in environments where data frequently transitions between performance and capacity tiers.

In addition to performance, the solution architecture provides a strong foundation for operational scalability and future expansion. The disaggregated design allows compute, storage, and networking resources to scale independently, while the unified namespace simplifies data access across tiers. Although high availability was not implemented during this PoC, the architecture is inherently designed to support it through standard clustering frameworks, making it well suited for production deployments that require resilience and continuous availability.

Overall, this PoC validates that BeeGFS on a tiered OpenFlex Data24 and Ultrastar Data60 platform delivers a balanced, flexible, and high-performing solution for modern data-intensive workloads. By combining strong parallel I/O performance, efficient metadata handling, and optimized data movement capabilities, the solution enables organizations to meet performance demands while maintaining cost control and operational simplicity in a single, scalable storage environment. This allows organizations to align performance, cost, and operational efficiency within a single, unified storage platform.

5601 Great Oaks Parkway
San Jose, CA 95119, USA
www.westerndigital.com

© 2026 All rights reserved. WD, the WD Design, Western Digital, OpenFlex, RapidFlex, and Ultrastar are registered trademarks or trademarks of Western Digital Corporation or its affiliates in the U.S. and/or other countries. Dell and PowerEdge are trademarks of Dell Inc. or its subsidiaries. The NVMe and NVMe-oF word marks are trademarks of NVM Express, Inc. All other marks are the property of their respective owners. Product specifications subject to change without notice. Pictures shown may vary from actual products. Not all products are available in all regions of the world.