Western Digital.

Solution Brief

Cost Effective All-Flash BeeGFS HA Solution with Xinnor, Powered by OpenFlex™ Data24 4000 Series NVMe-oF™ Storage Platform

XINNOR | BeeGFS®

Highlights

- Gives your business a boost by accelerating application performance with OpenFlex Data24 4000 Series.
- Combines multiple storage servers to provide a highly scalable shared network file system with striped file contents providing high throughput satisfying demands of large numbers of clients with BeeGFS.
- Extends high availability and robust data protection with Xinnor xiRAID Classic on a shared storage clustered environment.
- Provides up to 1474.56TB[1] capacity in a 2U unit OpenFlex Data24 4000 Series.
- Enables server and storage consolidation, reducing capital and licensing costs.

Benefits

- Enables multiple servers to share NVMe flash storage as if it were local.
- Leverages low latency fabric to fully utilize IOPS and capacity.
- Provides more efficient use of large capacity SSDs at low latency.
- Provides open composability through mature NVMe-oF standard.
- Balances access to eliminate over-subscription and maintain NVMe performance.
- Cost-effective and scalable storage solutions.
- High availability (HA) and robust data protection.
- Accelerated application deployment and performance enhancements.
- Easy data management.
- Optimized for highly concurrent access.
- High Performance

Executive Summary

In high-performance computing (HPC), designing a well-balanced storage system to achieve optimal performance presents significant challenges. A typical storage system consists of a variety of considerations, including file system choices, file system tuning, disk drives, storage controllers, IO cards, network cards, and switch strategies. Configuring these components for best performance, manageability, and future scalability requires a great deal of planning and organization.

The Western Digital validated design for HPC with BeeGFS, Xinnor HA solution using OpenFlex Data24 showcases ease of use, high-throughput, scale-out, parallel file system storage with well described performance characteristic.[2]

Problem Statement

In recent years, the requirement for I/O performance has increased dramatically with increasing demand for both traditional HPC I/O bandwidth but also for increased IOPS and metadata performance. Modern IO intensive workloads like Big Data, ML, AI, and Internet of Things (IoT) workloads require data infrastructure designed to scale storage and compute independently to ensure both are provisioned efficiently and effectively.

Western Digital's OpenFlex Composable Infrastructure leverages NVMe-oF to enable HPC environments to thrive by extending the high performance of NVMe™ flash to shared storage. OpenFlex composable Infrastructure provides the flexibility to meet varying requirements depending on data workload and performance requirements and is built to deliver streaming performance in HPC environments. With low latency and consistently high bandwidth, data is accelerated to the speed of flash and can be shared across hundreds of hosts.

Xinnor xiRAID is a high-performance software RAID, with xiRAID Classic it supports high availability, the possibility to build a cluster of server nodes capable of surviving multiple drive failures, as well as a complete server node failure.

The BeeGFS filesystem has steadily gained popularity as it solves bottleneck issues at the file system layer and is optimized for highly concurrent access to shared files and was specifically designed for data-intensive HPC workloads.

[1] One terabyte (TB) is equal to one trillion bytes. Actual user capacity may be less due to operating environment.
[2] This document is not an endorsement of xiRAID or BeeGFS by Western Digital, and no warranty of either product is either expressed or implied.

Cost Effective All-Flash BeeGFS HA Solution with Xinnor, Powered by OpenFlex Data24 4000 Series NVMe-oF Storage Platform

OpenFlex Data24 Overview

Flash technology has revolutionized the performance of storage systems, and Western Digital leads the way with NVMe-oF technology that propels flash storage to its full potential. Western Digital's OpenFlex Data24 4000 series NVMe-oF storage platform extends the high performance of NVMe flash to shared storage. Like the original OpenFlex Data24 and the OpenFlex Data24 3200 series, it provides low latency sharing of NVMe SSDs over a high-performance Ethernet fabric to deliver similar performance to locally attached NVMe SSDs. Industry-leading connectivity, using Western Digital RapidFlex™ network adapters, allows up to 12 hosts to be attached without a switch.

OpenFlex Data24 4000 series uses Western Digital's RapidFlex A2000 Fabric Bridge devices to provide 12-ports of 100GbE which can connect to RDMA and/or TCP configured host ports. While RoCE (RDMA over Converged Ethernet) connections have historically been preferred in data centers, TCP offers greater ease-of-use and is sometimes preferred. OpenFlex Data24 4000 series offers the flexibility of connecting to either RoCE or TCP host ports for optimum usage. OpenFlex Data24 4000 series enables PCIe® Gen4 performance throughout the chassis, bringing the full performance capability of each SSD to the Ethernet fabric. PCIe Gen4 SSDs from multiple brands are supported. Storage infrastructures built on OpenFlex benefit from accelerated performance, improved responsiveness, and increased the agility of your business.

XINNOR xiRAID Classic Overview

Software RAID offers high flexibility, zero associated hardware costs, and vendor-agnosticism in terms of compatibility. It's worth noticing that software RAID is currently the only option to support the new class of NVMe-oF JBOF (EBOF) devices for disaggregated storage in the CDI world. Xinnor xiRAID allows for the creation of high-performance RAID from NVMe and SAS/SATA SSD for the most demanding enterprise-grade tasks, ensuring fast and effective access to data. It is easy to maintain and more suitable for operating in large server infrastructures.

xiRAID Classic is purpose-built for the new types of flash drives. With I/O handling parallelization and lockless data path, xiRAID arrays have very small RAID penalties and perform very close to the raw hardware capabilities. xiRAID Classic is a lightweight software module with low CPU and RAM usage. Among other features, xiRAID Classic supports high availability, the possibility to build a cluster of server nodes capable of surviving multiple drive failures, as well as a complete server node failure. This feature is critical for building highly resilient HPC storage clusters.

BeeGFS Overview

BeeGFS is a hardware-independent POSIX parallel file system (a.k.a., software-defined parallel storage) developed with a strong focus on performance and designed for ease of use, simple installation and easy management. It is designed for all performance-oriented environments, including HPC, AI, deep learning, life sciences, oil, gas, media, and entertainment.

One of the most fundamental concepts of BeeGFS is the strict avoidance of architectural bottlenecks or locking situations in the cluster through the user space architecture. BeeGFS is striping the file content on multiple storage nodes, plus it is distributing file system metadata across multiple metadata servers, which benefits small, large systems, and metadata-intensive applications.

BeeGFS is built on highly efficient and scalable multithreaded core components with native RDMA support. File system nodes can serve RDMA (InfiniBand, Omni-Path, RoCE, and TCP/IP) network connections at the same time and automatically switch to a redundant connection path in case any of them fail.

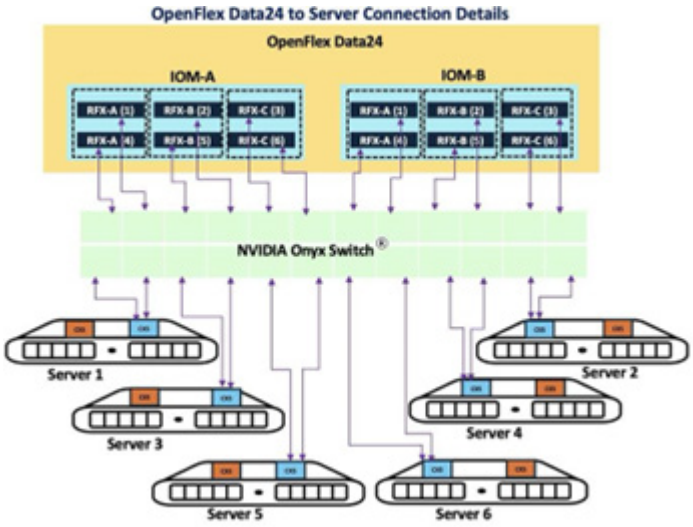
BeeGFS and Xinnor xiRAID HA Deployment Topology with OpenFlex Data24

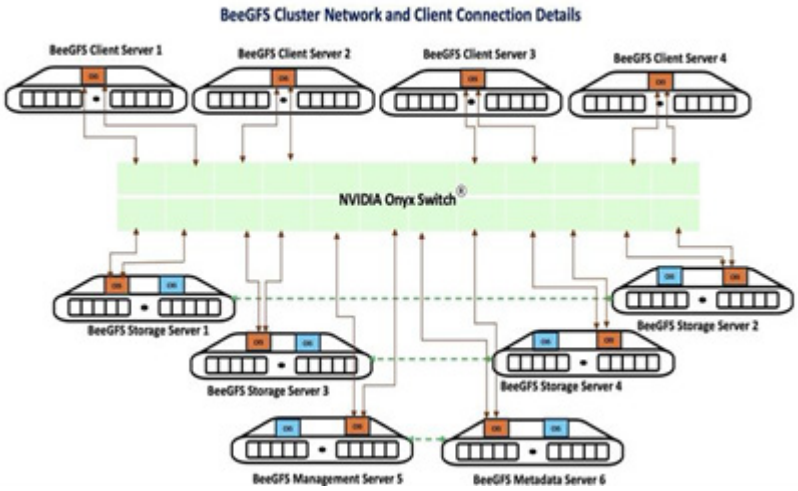
Western Digital is proud to partner with BeeGFS and Xinnor. The solution enables an innovative, highly available, high-performance parallel file system designed to provide the performance, scalability, and flexibility required to run the most demanding HPC applications. The solution is tested to ensure it has no single point of failure and can survive multiple drive failures as well as the failure of any server within the cluster. The Xinnor xiRAID Classic high availability feature ensures HA in case of server failure by automatically failing over the RAIDs to the paired server with Pacemaker Cluster Shell cluster. BeeGFS solves bottleneck issues at the file system layer and is optimized for highly concurrent access to shared files and was specifically designed for data-intensive HPC workloads. Western Digital OpenFlex Data24 with BeeGFS and Xinnor xiRAID Classic provides a highly available and scalable shared network file system with striped file contents for HPC environments.

The PoC has been done to validate the HA features and performance. Please refer to the "High Capacity BeeGFS HA Solution with Xinnor Powered by OpenFlex Data24" reference architecture for more details.

This allows users to overcome the performance limitations of single servers, single network interconnects, or a limited number of storage devices. BeeGFS clients directly access shared storage and can communicate with multiple servers simultaneously, giving your applications truly parallel access to the shared data. In such a system, high throughput demands of large numbers of clients can easily be satisfied, but even a single client can benefit from the aggregated performance of all the servers in the system.

Cost Effective All-Flash BeeGFS HA Solution with Xinnor, Powered by OpenFlex Data24 4000 Series NVMe-oF Storage Platform





Performance Results

The measured performance numbers are specific for the current configuration of BeeGFS HA topology in conjunction with OpenFlex Data24 using xiRAID RAID5 volumes. For complete details refer to the High Capacity BeeGFS HA Solution with Xinnor Powered by OpenFlex Data24 4000 Series reference architecture.

The performance achieved with IOR on Xinnor RAID5 volumes with BeeGFS configuration:

BeeGFS Performance on Xinnor RAID5 Volumes

IOR Test Results	Throughput (GBps)
Read	61.5 GBps
Write	24.4 GBps

Note: The performance results are captured after running multiple iterations with different data size, transfer size. All of the performance data are subject to change depending on the Drive model, capacity, workloads, servers, CPU, tunings, and HA topology used.

Conclusion

BeeGFS on Western Digital's OpenFlex Data24 Storage with Xinnor xiRAID Classic addresses the need of IT/HPC with a well-designed solution that is easy to manage and fully supported. The deployment of a highly available BeeGFS solution using xiRAID Classic showcases the capability of combining high-performance RAID storage with advanced cluster management. Using this solution based on the latest NVMe-oF and CDI technologies, organizations can quickly deploy a proven, self-service, composable infrastructure solution, helping customers move to a more flexible, variable cost model.

The deployment of BeeGFS on Western Digital's OpenFlex Data24, combined with Xinnor xiRAID, provides a reliable, high-performance, and highly available storage solution for IT and HPC needs. With advanced NVMe-oF and composable infrastructure technologies, organizations can achieve:

- Cost-effective and scalable storage solutions.
- High availability (HA) and robust data protection.
- Accelerated application deployment and performance enhancements.

Western Digital.

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

May 2025 / Revision 01