



SUPERMICRO PETASCALE ALL-FLASH GEN 5 JBOF STORAGE SOLUTION WITH NVIDIA® BLUEFIELD® -3 DPU

High Performance All-Flash Solution Provides High-efficiency Platform for Scale-Up and Scale-Out Storage



Supernode E3.S JBOF SSG-229J-5BE36JBF



Supernode U.2 JBOF SSG-229J-5BU24JBF

Executive Summary

The Supernode Gen 5 JBOF provides an all-flash storage building block for either scale-out or scale-up storage. The Supernode Gen 5 JBOF is a 2U system available with either U.2 storage bays or E3.S storage bays. It uses the NVIDIA BlueField-3 DPU (data processing unit), which combines the functions of a networking NIC with hardware acceleration and a storage accelerator that runs the storage application. This system supports two canisters that contain the BlueField-3 DPU. This dual port design provides high availability failover between the canisters and enables up to two BlueField-3 DPUs per canister, with up to four in total.

TABLE OF CONTENTS

- Executive Summary 1
- What is a JBOF? 1
- What are the Benefits of a JBOF Storage Solution?..... 2
- Supernode JBOF Specs: 5
- Example Deployment Architectures 6
- Summary & Further Information 6

What is a JBOF?

JBOF, or “Just a Bunch of Flash,” is an intelligent all-flash storage array that uses a DPU rather than the separate CPU, memory, and NIC found in traditional storage servers. To unpack this further, a DPU is an intelligent dedicated special purpose processor which is optimized for networking and storage workloads. The DPU combines the functions of a networking card, a networking



data accelerator, and a storage processing accelerator into a single ASIC (Application Specific Integrated Circuit). We will discuss the NVIDIA BlueField-3 DPU as the self-hosted DPU used in this solution.

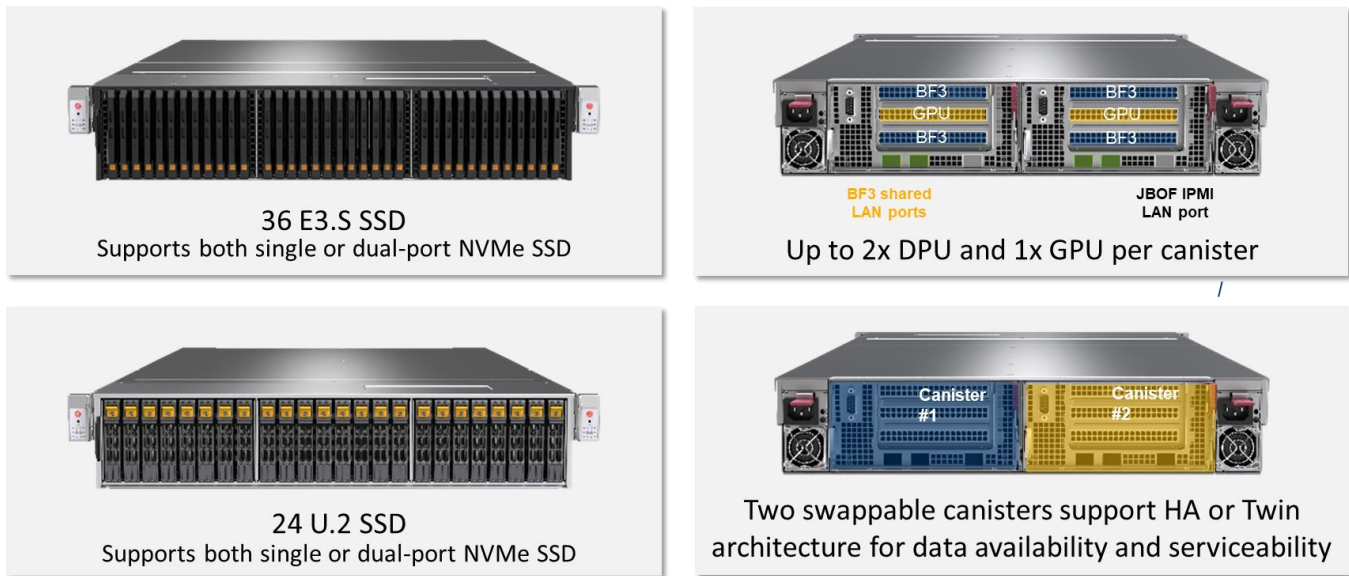


Figure 1 - Supermicro All-Flash JBOF with Either E3.S or U.2 SSD Storage

What are the Benefits of a JBOF Storage Solution?

There are several benefits to implementing a JBOF Storage Solution as compared with traditional server-based storage:

Lower Power:

The power consumed by the JBOF is a combination of the control plane power (either the DPU or the CPU/memory/NIC) and the power used by the SSD media. Because the DPU is explicitly designed for networking and storage acceleration, it consumes around 50% less power than an equivalent storage server.. The power consumed by the is the same for the DPU-based system and the CPU-based system. For a 24 SSD system, the total system power for a JBOF system is 10-15% less than a storage server. Aggregated over 100 JBOF systems used, this amounts to up to 20KW saved directly and an additional 10KW saved in data center cooling, assuming a PUE (Power Usage Effectiveness) of 1.5 for a typical data center.

Higher Performance:

The DPU is a single ASIC that integrates multiple CPU cores, hardware offload and acceleration engines, and memory and network and PCIe interfaces. The integrated Ethernet or Infiniband 400Gbps network interfaces eliminate multiple data transfers from the NIC to the CPU and then to SSDs found in server-based design. In addition to a more direct data path, the DPU has hardware-based acceleration for networking (such as encryption, cryptography, and streaming) and storage (such as GPU Direct Storage, NVMe/TCP acceleration, erasure coding for RAID, data deduplication and decompression). The DPU also supports NVMe-over-Fabric target offload acceleration, removing a data copy between the network interface and the CPU. In testing with the NVIDIA BlueField-3 DPU, we have found the following results:

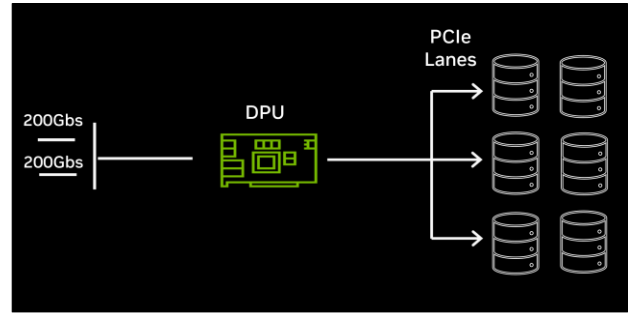
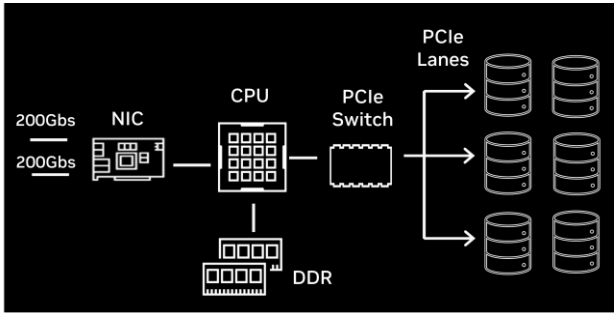


Figure 2 - x86 CPU base JBOF (left) and DPU based JBOF (right)

	Without NVMe Offload	With NVMe Offload	% Difference
512B* Block Size*	6M IOPs	8M IOPs	33.3% Better
4K* Block Size*	2M IOPs	5M IOPs	150% Better
Latency* (excl. SSD)	15 usec	5 usec	66.6% Lower
4KB Block Latency**	100 usec	86 usec	14% Lower

*Data from NVIDIA testing of BlueField-3

**Data from Supermicro testing with BlueField-3 and Gen 5 JBOF.

This data shows that the DPU has a higher IO rate (IOPs) and lower latency than a CPU-based implementation without NVMe-of offload.

Integration:

The integrated design of the DPU with the JBOF goes beyond the special purpose design of the DPU ASIC. The storage application, either object storage or file system, runs in user-mode on the DPU rather than requiring any kernel level access, preventing inadvertent data access to system level functions. The JBOF system manages the discrete control of DPUs, nodes and chassis providing high availability operation with nearly zero service downtime.

NVIDIA BlueField-3:

The NVIDIA BlueField-3 DPU is optimized to offload and accelerate networking, storage, security and management features. Because the BlueField DPU unifies networking, CPU, memory controllers, PCIe switch bi-furcation, and traffic accelerators on one chip, data can move directly between SSDs and the high-speed network ports without needing to be processed by an external CPU or cross a separate PCIe bus multiple times. This allows low latency storage access and makes the system more efficient in terms of price-performance and power efficiency.

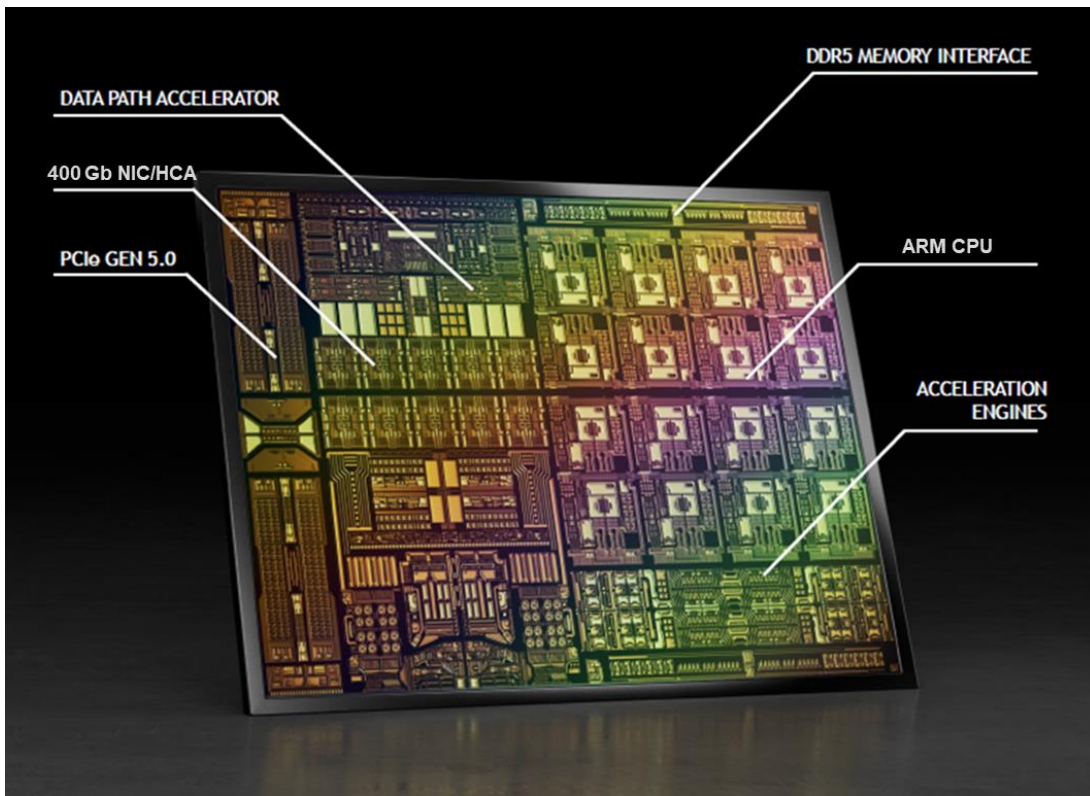
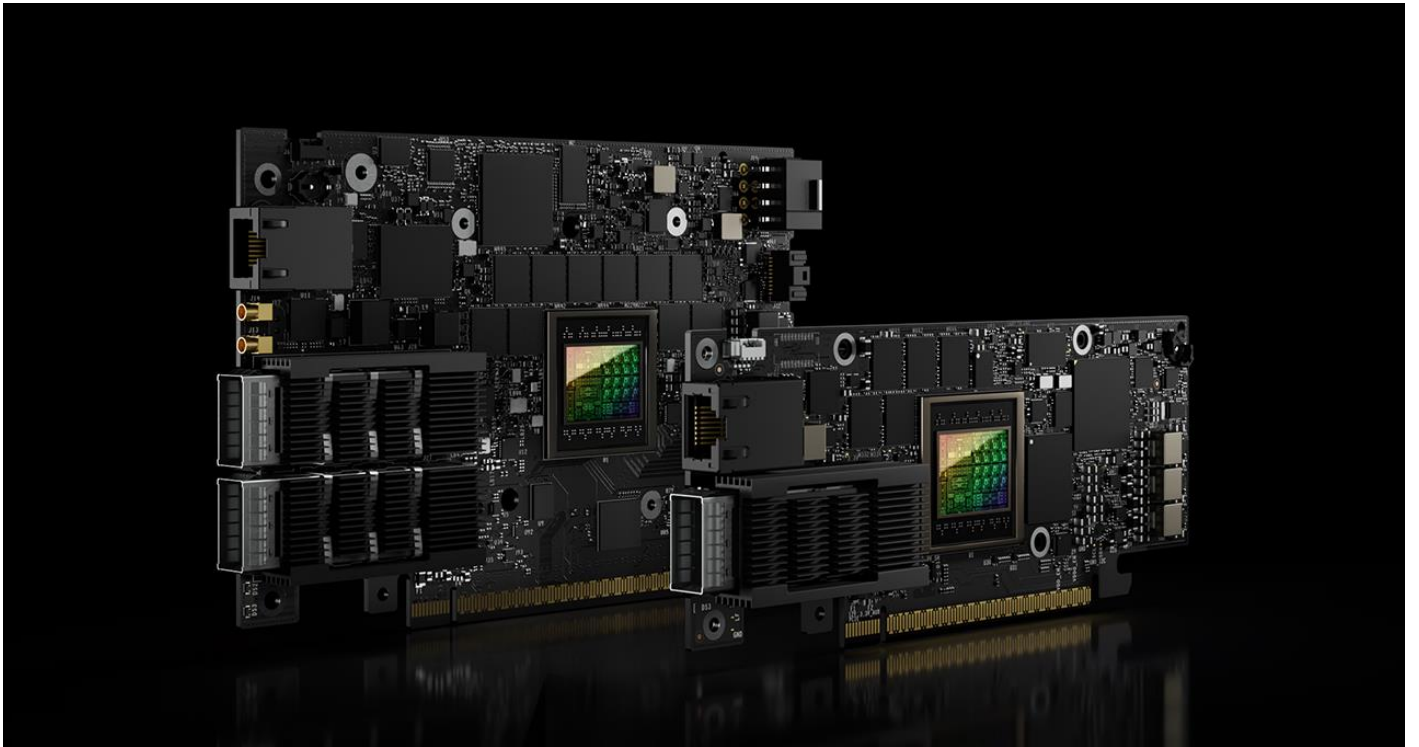


Image 1 - NVIDIA BlueField-3 Networking Platform (Image Courtesy of NVIDIA)

Software Support:

As previously mentioned, the data management software, either file or object, runs natively on the DPU. Two solutions are currently supported, and several others are being tested.

Cloudian: Cloudian HyperStore is an object storage that supports the S3 object protocol as well as NFS and SMB file protocols. It supports very large unstructured data sets with robust data protection and ransomware defense.

Hammerspace: The Hammerspace Global Data Platform is a high-performance, standards-based parallel file system that also supports a single global namespace and data orchestration services. Hammerspace's Hyperscale NAS architecture combines standards based pNFS with the simplicity of NAS.

Supermicro JBOF Specs:

Supermicro has two JBOF products which share the following features:

Form Factor	2U
PCIe Generation	Gen 5
SSDs	Either 32x E3.S or 24x U.2 Single or Dual Port
DPU Support	2x Gen 5 BlueField-3 DPUs per canister (up to 2 canisters per system)
Additional PCIe Gen 5 Slot	1x additional PCIe Gen 5 x16 GPU or SAS HBA
Power & Cooling	Hot-swap redundant Titanium power supplies and high performance counter rotating hot swap fans
Embedded PCIe Switch	1 Gen 5 Embedded Switch per Canister
Dual Port SSD	Dual Port SSD support for Active-Active or Active-Passive Failover
System Management	BMC support for firmware management of PCIe Switch, BMC, and CPLD Support for IPMI 2.0 and Redfish 1.8 Integrated management of BlueField-3 thermals and firmware

SSD Qualified Media Support:

SSD Form-Factor	PCIe Generation	Manufacturer	Single-Port	Dual-Port
E3.S	Gen 5	KIOXIA	KIOXIA CM7 KIOXIA CD8P	KIOXIA CM7
		Micron	Micron 9550	N/A
U.2	Gen 5	KIOXIA	KIOXIA CM7 KIOXIA CD8P	KIOXIA CM7
		Micron	Micron 9550	N/A

Example Deployment Architectures

Single vs. Dual Ported and Scale-Up vs. Scale-Out Architecture:



Figure 4 - Scale-up and Scale-Out Storage Options

The JBOF can be deployed as either a single-ported system with a single canister or a dual-ported system with two canisters connected to shared media. The dual-ported system provides high-availability support for system failover, enabling the failure of one canister to be transferred to the other canister. This is often used in scale-up storage architectures for file servers since multiple users are attached to the storage host controller.

For scale-out storage, either as file or object, a single-port design is commonly used since data protection using erasure coding happens at the system (node) level, and a clustered architecture is used to enable the system to continue to run if an individual node fails.

Either approach can be implemented using one or two NVIDIA BlueField-3 DPUs per canister.

Summary

Supermicro's PCIe Gen 5 JBOF is an All-Flash array supporting either E3.S or U.2 PCIe Gen 5 SSDs and using up to four NVIDIA BlueField-3 DPUs. The single or dual canister designs enable either scale-out storage implementations for object or file storage or scale-up file storage. The advantages of the JBOF storage implementation over the traditional server-based storage system include lower power consumption, by up to 15% per system, faster performance due to lower latency, and a more integrated implementation with the storage software running on the BlueField-3 DPU.

Further Information

<https://www.supermicro.com/en/products/jbof>

SUPERMICRO

As a global leader in high performance, high efficiency server technology and innovation, we develop and provide end-to-end green computing solutions to the data center, cloud computing, enterprise IT, big data, HPC, and embedded markets. Our Building Block Solutions® approach allows us to provide a broad range of SKUs, and enables us to build and deliver application-optimized solutions based upon your requirements.

© 2024 Supermicro, Inc. All rights reserved.

October 2024

