DCTR-304-1: High-Performance Storage for the Data Center

Confidential

- Thursday August 7th, 1:25 PM to 02:30 PM

PRO **DCTR-304-1: High-Performance Storage for the Data Center**

Ballroom F, Santa Clara Convention Center, First Floor

Data Center Storage and Memory

Organizer:

Jonathan Hinkle, Senior Director - Azure Memory and Storage Pathfinding, Microsoft

Presenters:

Johann Lombardi, Senior Distinguish Technologist, HPE/Linux Foundation

Paul McLeod, Storage Product Director, Supermicro

Presentation Session Description:

In this session, we delve into innovative approaches and technologies shaping the future of data storage and management. A common thread throughout the presentations is the focus on optimizing performance and efficiency in increasingly demanding datacenter environments. One presentation explores the evolution of software-defined storage, highlighting alternative architectures like DPU-powered JBOF arrays and NVIDIA Grace CPU integrations that promise enhanced power efficiency and performance. Meanwhile, another presentation addresses the pressing issue of datacenter cooling, assessing the impact of immersion cooling on SSD stability and reliability. The study evaluates single-phase and two-phase liquids, offering insights into optimizing storage solutions in these alternative cooling scenarios. Together, these presentations underscore the critical balance between cutting-edge hardware advancements and innovative cooling strategies, emphasizing the importance of reliability and efficiency in next-generation datacenters.

Supermicro

Paul Mcleod - Product Director, Storage

A seasoned industry professional with over 25 years of experience creating storage solutions. An early champion of Software Defined Storage as well as the new and emerging methods being used to create scalable data pipelines for large scale HPC and AI environments.

DCTR-304-1: High-Performance Storage for the Data Center

Paul McLeod

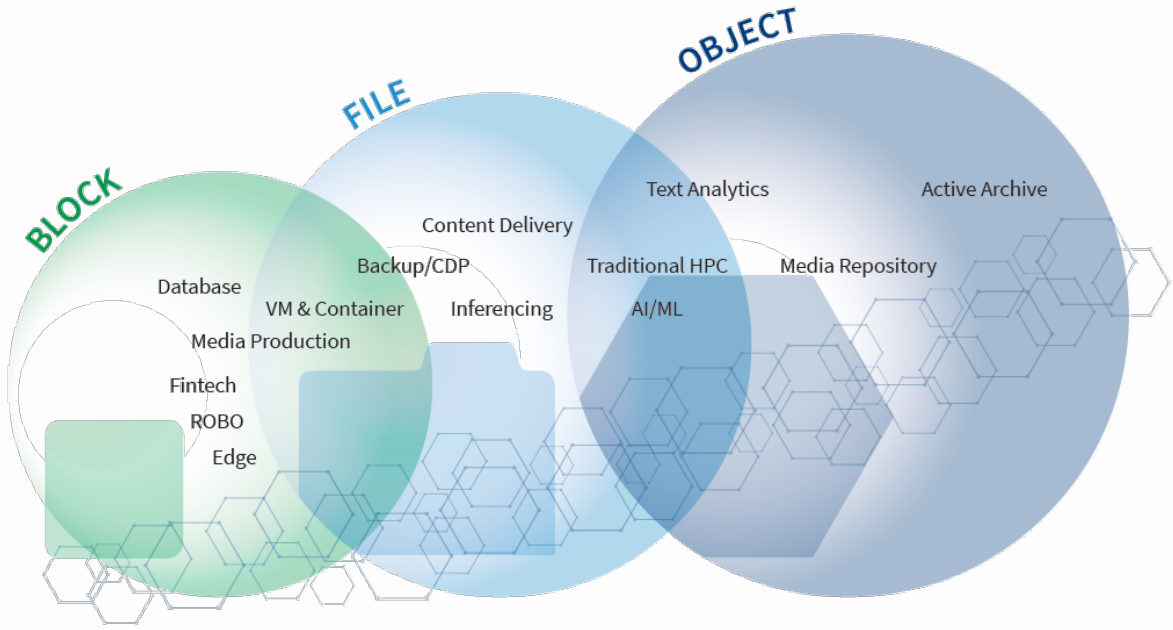
Agenda

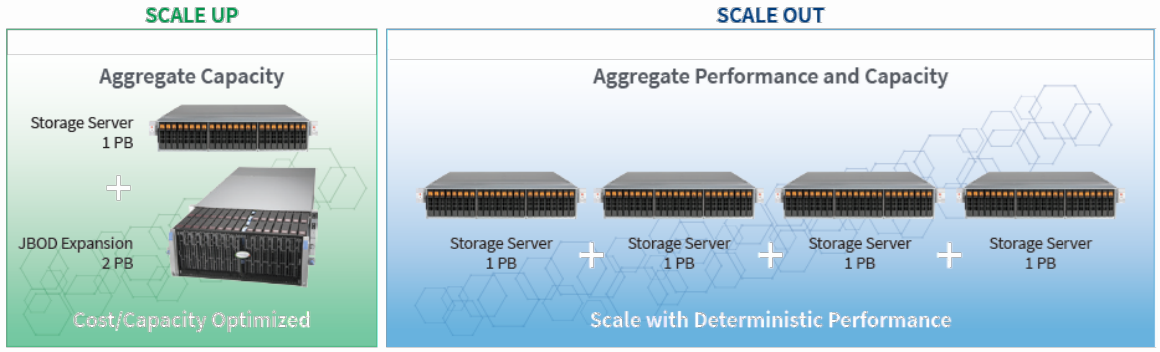
Confidential

- Supermicro – Software Defined Storage
- Architectural considerations for scaling storage
  - Clustered Storage
    - Scale-up & Scale out
- Enabling Technology for large-scale deployments
  - Petascale Family featuring EDSFF bays
  - AMD Epyc, Intel Xeon, NVIDIA Grace Superchip CPUs
  - NVIDIA BlueField-3 self-host mode DPU+JBOF
- Improving Server Efficiency
  - Airflow
  - CPU design
  - Smart Networking Improving Data flow
- Summary

SuperStorage

- Purpose-built Storage Hardware
- Embracing Innovation & Scale
- Software Defined
  - ✓Large Selection of ISV Partners
  - ✓Certs & Reference Architectures
  - ✓Flexible/Extensible Platforms to Build on
  - ✓Rack Scale Solutions





Supermicro Software-Defined Storage Partners

Confidential

Leading Storage Building Blocks for the Software-Defined Data Center

Scale Performance	On-Prem / Hybrid	Scale Capacity
Traditional/Monolithic SAN/NAS	Entry Enterprise/Tiered	Backup, Object Storage
AI/HPC/PFS, Scale-out SAN	VM/HCI & Converged	Scale-out NAS, Archive
VM & Containerized @Scale	DAS/JBOD/JBOF	Analytic Applications

MemVerge, ddn, VAST, WEKA, HAMMERSPACE, Qumulo, Quantum, NUTANIX, vmware, MINIO, PROXMOX, OpenZFS, Microsoft, IBM, OSNEXUS, VEEAM, SCALITY, ceph, Red Hat, SUPERMICRO, lightbits, croit, CLOUDIAN, Graid Technology Inc., COMMVAULT, ubuntu

Storage Product Families

Confidential

ENTERPRISE

Capacity Optimized - Cloud Density 3.5" Storage

- Optimized for enterprise data centers and edge deployments
- Front access, double-sided, product families
- 2U/3U/4U form factors
- 16-36 3.5" LFF Bays

TOP LOADING

- Maximum capacity for cloud-scale storage in 4U FF
- Highest storage capacity and $/TB value
- Single node and dual-server systems
- 60, 90 drives in 4U
- Dual Intel CPUs

SIMPLY DOUBLE

- High density LFF storage in 2U FF
- 2U Form Factor
- Up to 24x 3.5" bays and up to 4x Gen5 NVMe SSDs for caching
- Single Intel (X13) or AMD (H13) CPU

Hard Drive Products

HIGH AVAILABILITY DUAL-PORT

TCO Optimized All-Flash Storage

- High Availability Dual Port (Active-Active or Active-Passive)
- Shared Everything Architecture (Scale-up Building Block) in 2U
- All-flash 24x U.2 NVMe drives
- Single Intel Xeon 5th Gen CPU, dual server configuration

PETASCALE ALL-FLASH

- Revolutionary storage performance and capacity
- Highest IOPS and BW
- All-flash EDSFF NVMe in 1U and 2U
- Up to 32x E3.S NVMe SSDs
- Dual Intel and Single AMD CPUs

Flash Storage Servers

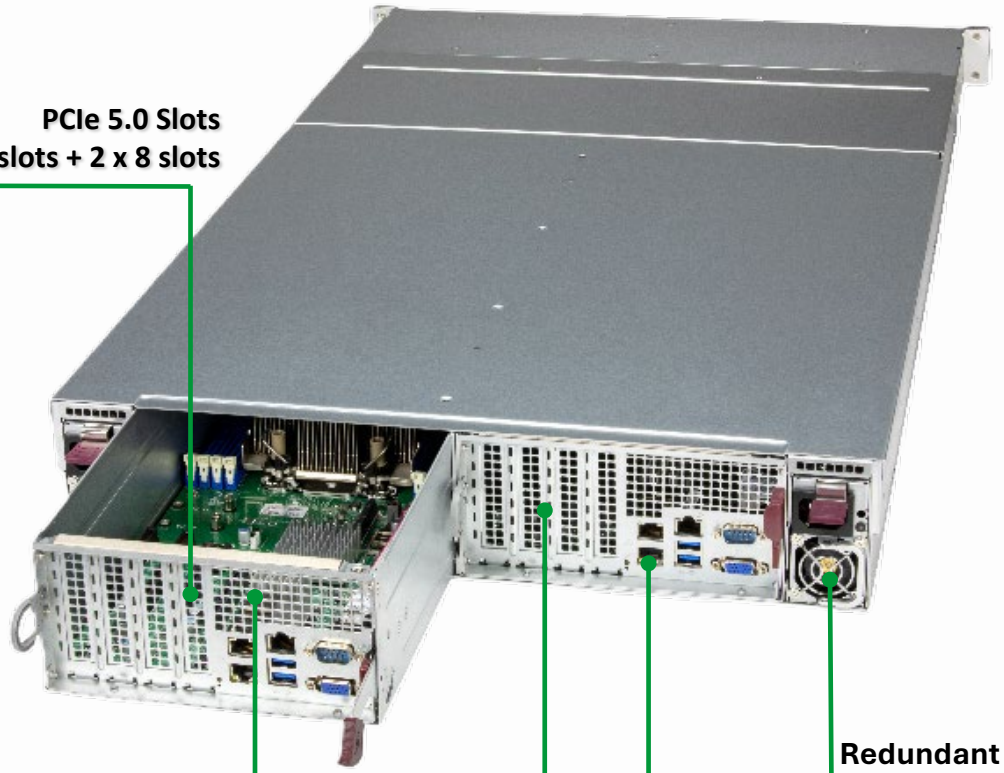
Architecture

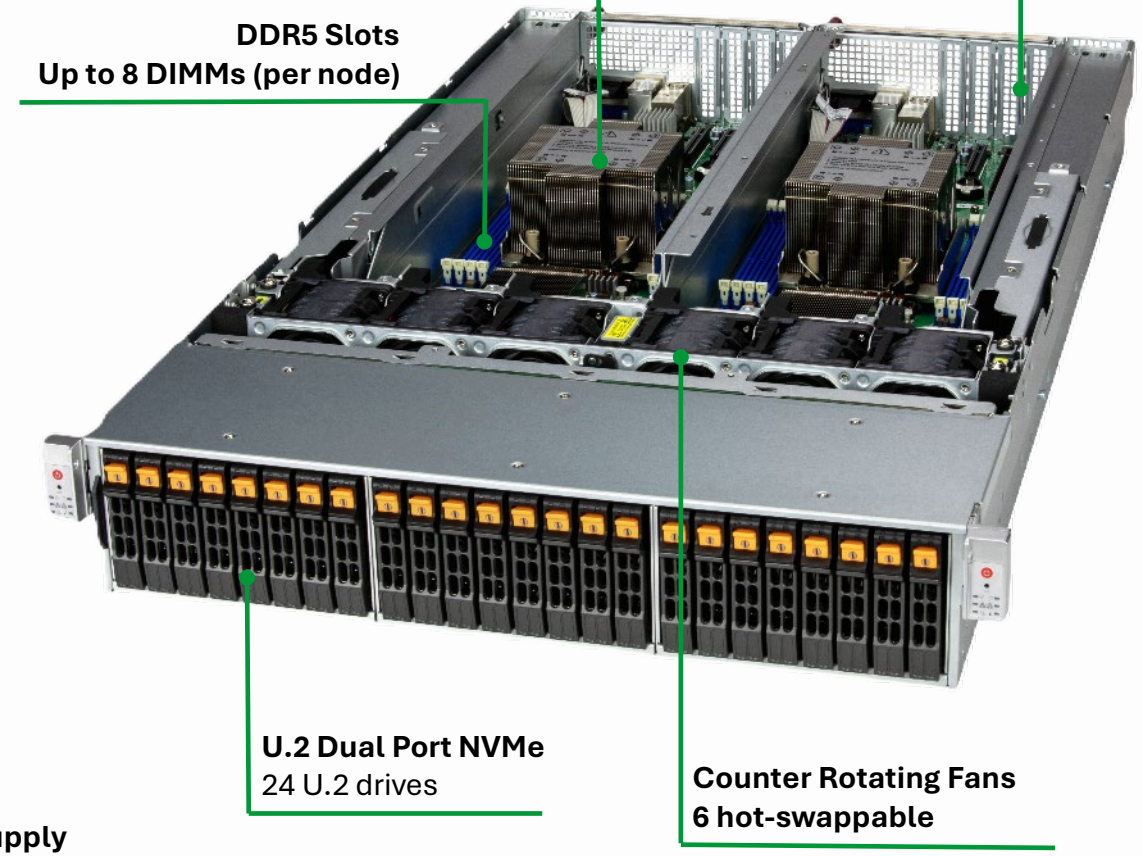
Scale-up | Scale-out

Dual-port High Availability Storage

Confidential

Model Name: SSG-221E-DN2R24R

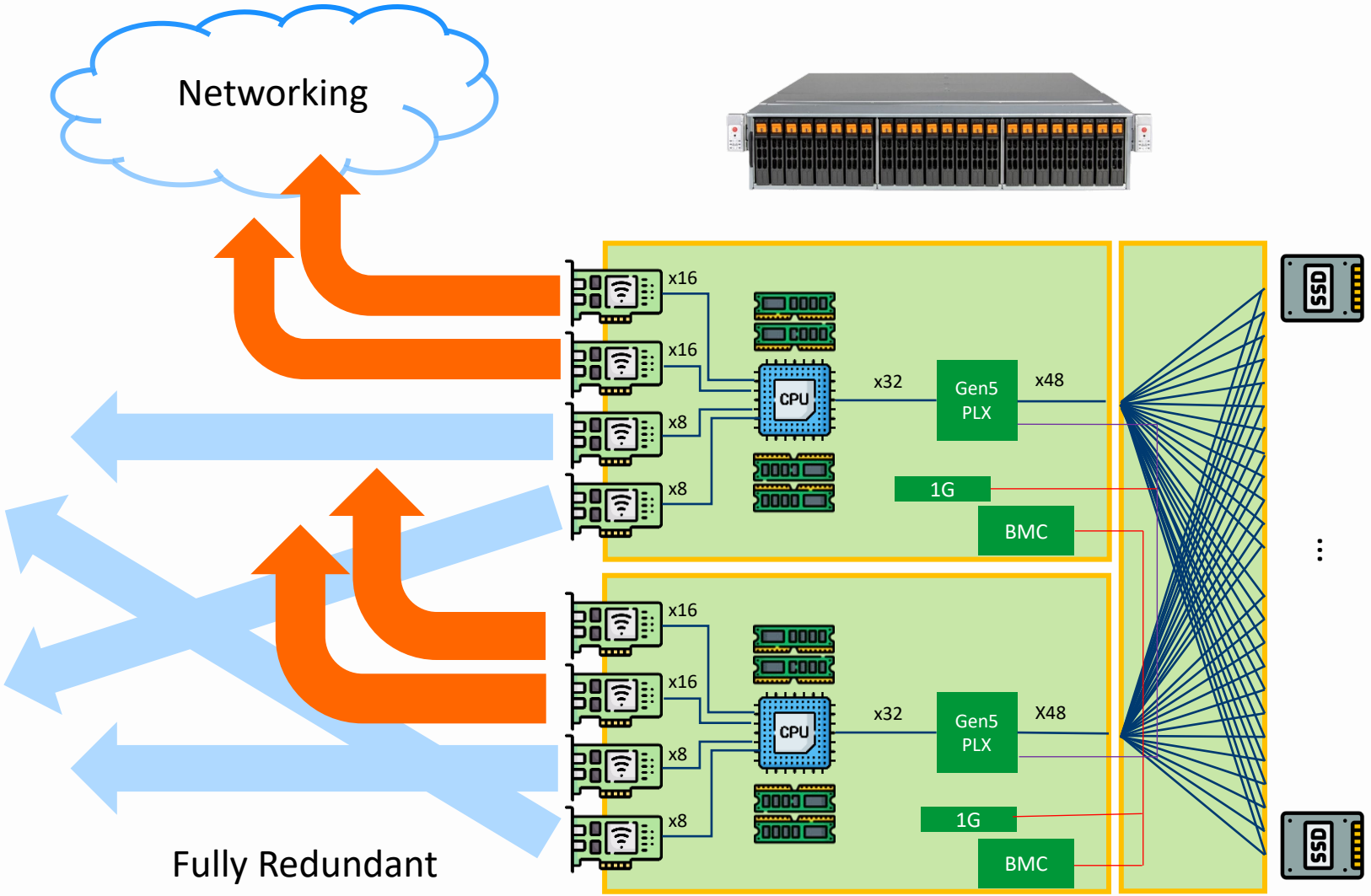




Dual-port Scale-up Architecture

SSG-221E-DN2R24R
+
947HE2C-R2K05JBOD

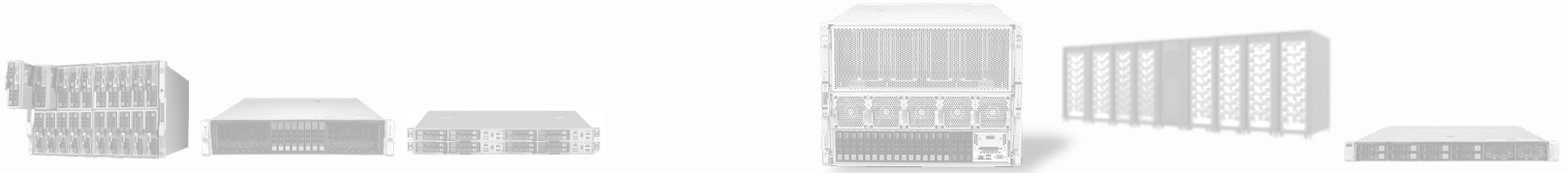
90 Bay SAS JBOD



Fully Redundant
JBOD Expansion

Scale-Up + Scale-out using PFS

Confidential



HPC

AI Pipelines

Parallel File System - Namespace

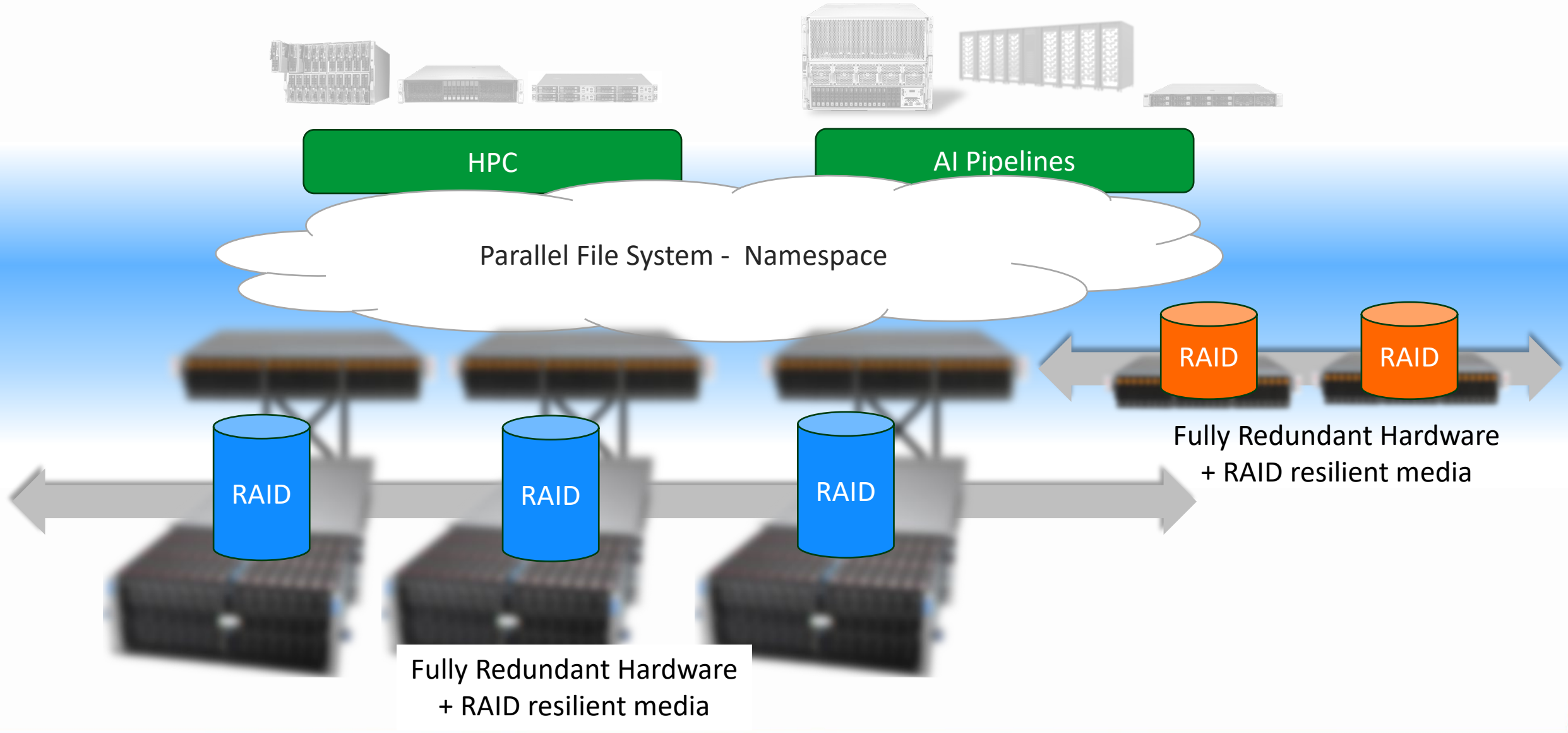
Scale Your Access-speed (Metadata)

Scale-out across Scale-up nodes

Scale Your Data (Throughput & Capacity)

Scale-Up + Scale-out using PFS

Confidential



Scale-out share nothing with EC

Confidential

HPC

AI Pipelines

File & Object - Namespace

Erasure Code Data Protection
Backend Network

Scale Data (Capacity & Performance)

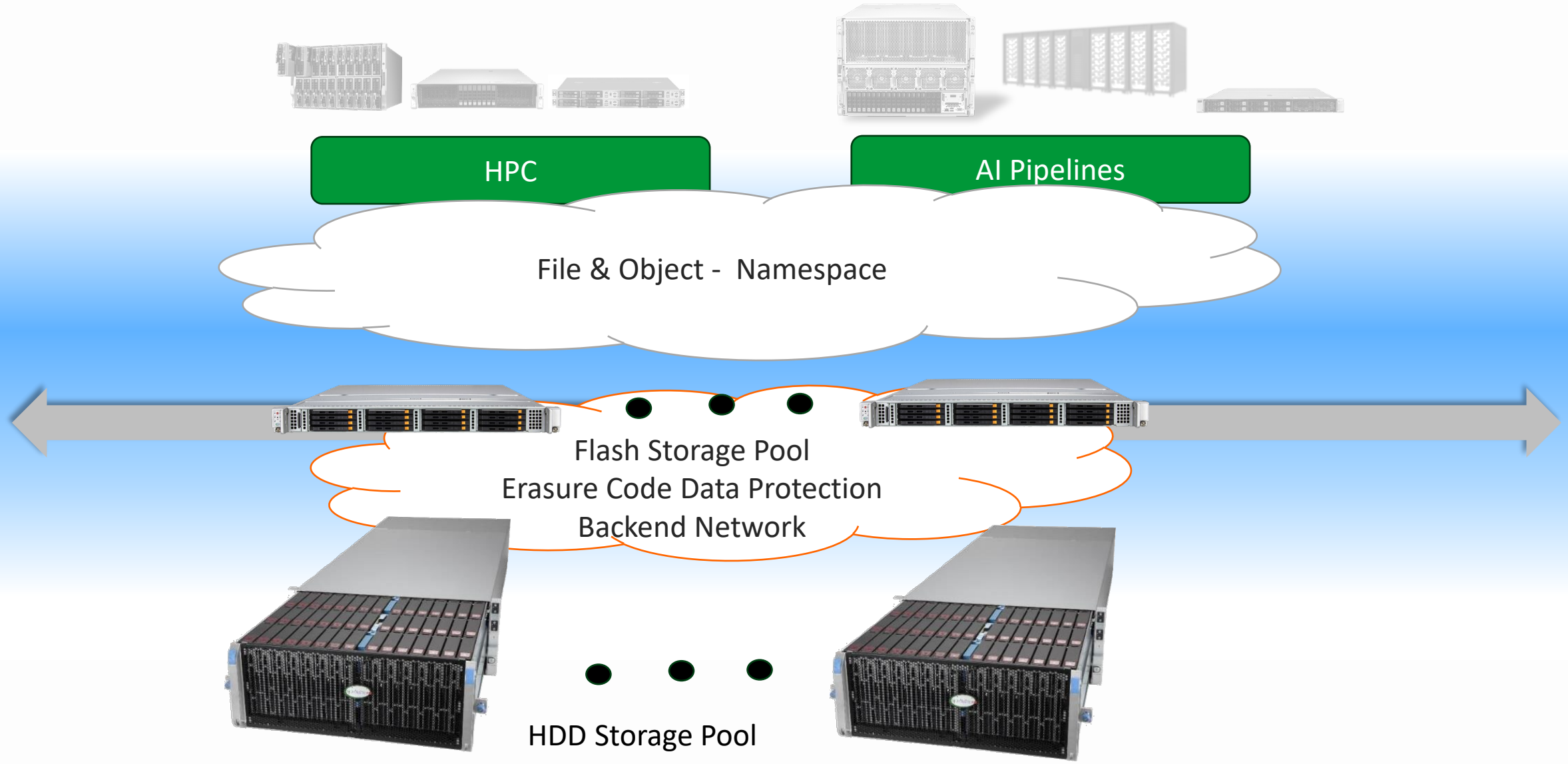
- ✓ Simplified management (homogeneous)
- ✓ latency performance
- ✓ Data placement/scatter, distributes workload
- ✓ Network Port aggregation= bandwidth

Scale Metadata (Performance)

- ✓ Optimal access/hot-data placement
- ✓ Distributed Metadata
- ✓ Close-to-client or even on-client

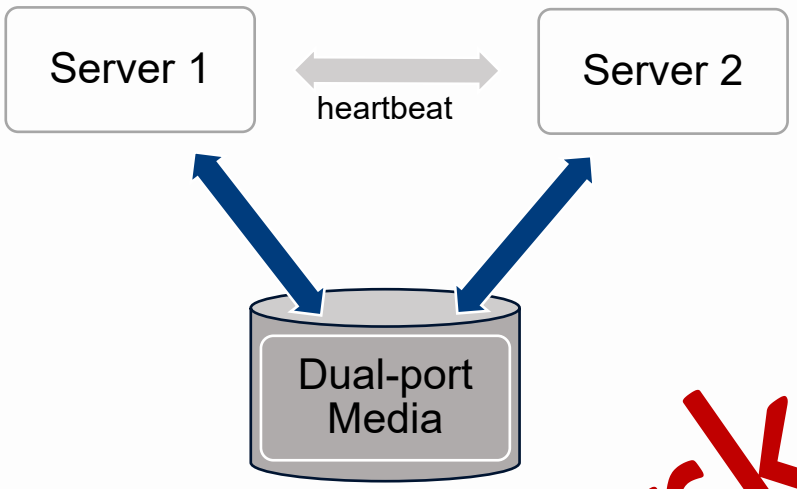
Scale-out share nothing with EC

Confidential

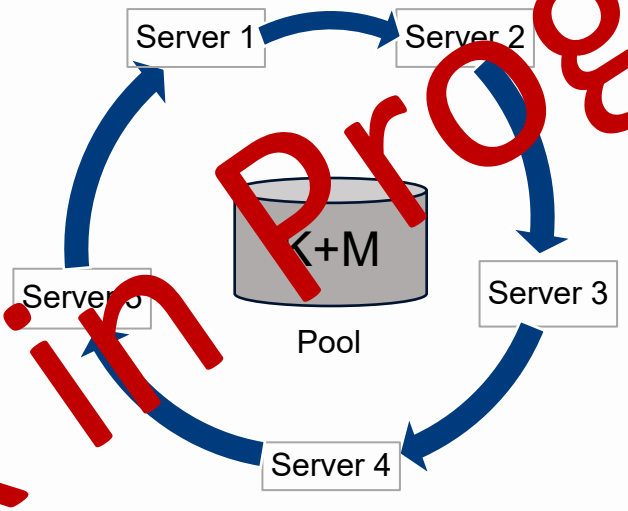


Architecture Efficiency Discussion

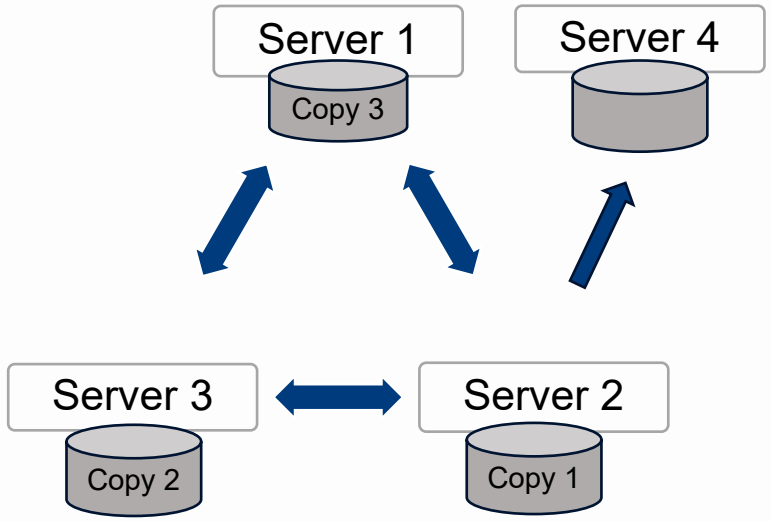
Confidential



Failover



Erasure code



Replication

Work in Progress

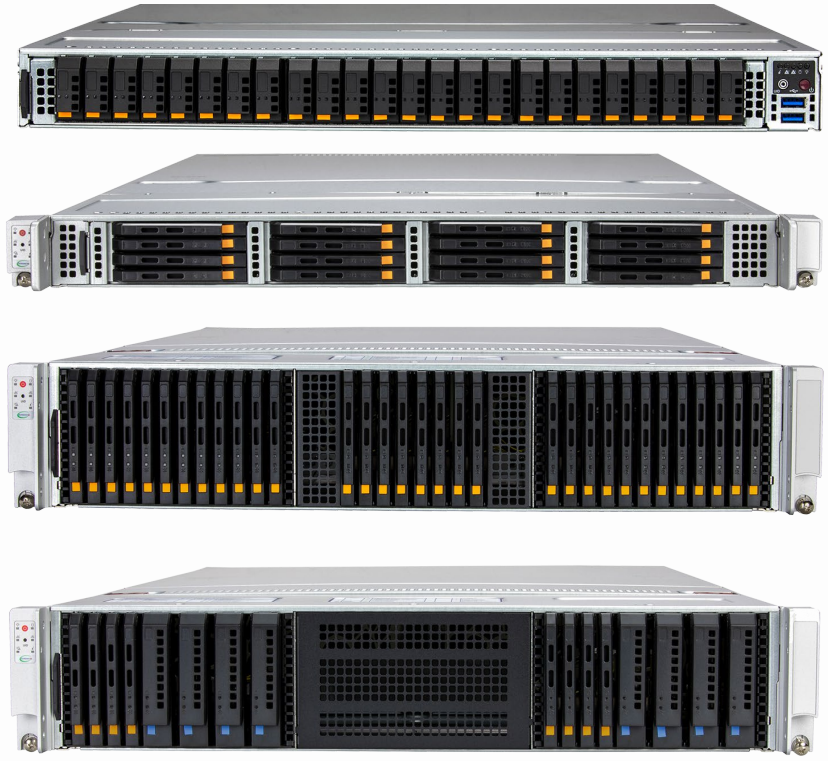
Scale-Out

Hardware Design

EDSFF Petascale All-Flash Highlights

Confidential

Flagship All-Flash Storage Servers Designed for Large Scale Workloads



NVIDIA | intel Xeon | AMD EPYC | 1 CPU | 2 CPU

TCO Optimized Design for Large-Scale Flash Applications

Optimized thermal design with EDSFF spec, balanced PCIe lanes for front SSD and rear I/O provides balanced non-blocking bandwidth for network-based clients

High Efficiency EDSFF Architecture Takes Performance to the Next Level

Allowing for more performance per watt than U.2 NVMe, as well as provides a new media bays standard for advanced technologies like FPGA and CXL devices

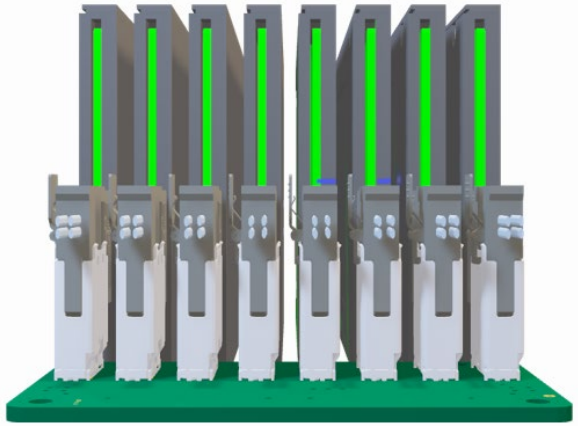
CXL Type-3 Expansion of Memory Bandwidth and Capacity

Up to 32 DIMMs plus up to 8 CXL expansion bays remove bottlenecks for memory constrained applications, like AI inferencing and in memory database

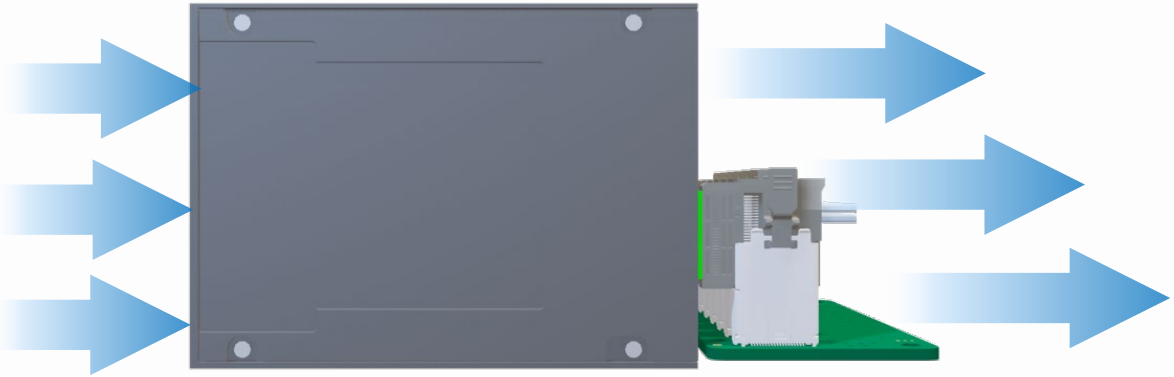
Product Specs

- Three motherboard platforms to choose from:
  - Dual Intel® Xeon® 6700 series
  - NVIDIA Grace Superchip (1U/16 E3.S)
  - Single AMD EPYC™ 9004/9005 series
- Optimized thermal design with EDFF media support
  - 1U E1.S and E3.S and 2U E3.S 1T and 2T (CXL)
  - Up to 8 CXL modules in 2U
- Balanced PCIe lanes for optimal I/O performance
  - Up to 1.9 PB in 2U using E3.S media
  - Up to 30M 4KB random read IOPS
  - Up to 230 GB/s sequential read bandwidth

EDSFF Storage Bay Design (E3.S)



EDSFF bay design optimizes airflow



Orthogonal Connector

Improving Efficiency of Air-cooled designs

Supermicro's Petascale Family

Now with NVIDIA Grace CPU Superchip

Confidential

Software Defined Storage

- Available Integrated with leading GDS certified AI Storage platforms – Weka

Seamless Integration with the NVIDIA AI Stack

- Design for tight integration to industry leading NVIDIA GPU and Networking technology (DOCA, CUDA and GDS)

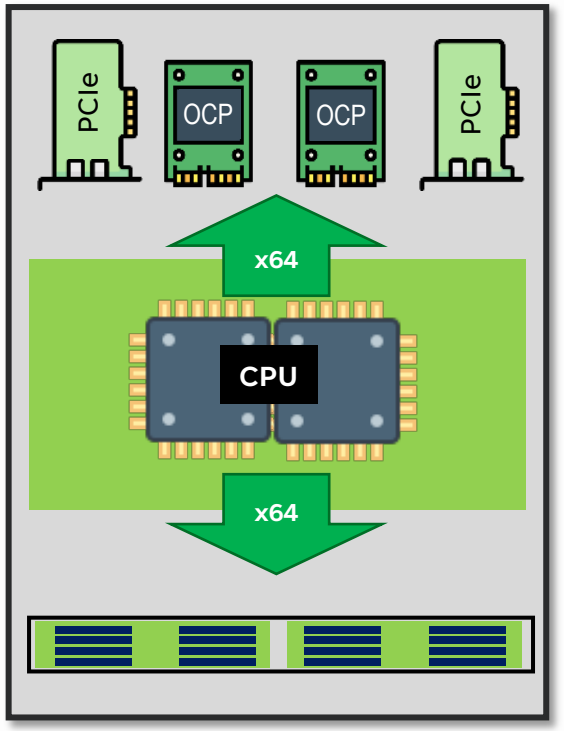
Performance Scaling

- Complete Gen5 design with industry leading memory bandwidth **(up to 1TB/s)** and 16 high-performance E3.S SSD **(220+ Gb/s and 40M IOPS).**

Power Efficiency

- Storage solution with industry leading performance and power efficiency

Optimized IO Architecture



1U 16x E3.S SSD

NVIDIA Grace Superchip Spec

Grace Design Features

- Memory capacity
- High bandwidth between CPUs and to memory
- Efficiency in power consumption and cooling

Grace CPU Superchip Overview	
Core Count	144 Arm Neoverse V2 Cores with 4 x128b SVE2
L1 Cache	64 KB i-cache + 64 KB d-cache
L2 Cache	1 MB per core
L3 Cache	234 MB
LPDDR5x Size	240 GB, 480 GB, and 960GB on-module memory options
Memory Bandwidth	Up to 1 TB/s
NVLink C2C Bandwidth	900 GB/s
PCIe Links	Up to 8 xPCIe Gen5 x16 option to bifurcate
Module Thermal Design Power (TDP)	500 W TDP with memory
Form Factor	Superchip module
Thermal Solution	Air cooled or liquid cooled

Industry's First Storage Server with NVIDIA Grace CPU Superchip

Confidential

NVIDIA Grace CPU Superchip

NVIDIA Connect®X-7 / X-8

Redundant Titanium
1600W Power Supplies

16 Front Hot-Swap E3.S
NVMe drive bays

Petascale JBOF with NVIDIA BlueField Data Processing Unit

Confidential

The Supermicro JBOF with NVIDIA BlueField-3 solution replaces the traditional storage CPU and memory subsystem with the BlueField-3 DPU

Petascale Flash using Bluefield DPU

Confidential

Self-host DPU

- Run Linux based SDS solutions from the DPU

Power Efficient Flash Storage

- Save up 15-20% of the power consumption vs x86 based subsystems of similar capacity

NVIDIA BlueField-3 B3220SH

- 2P QSFP112 200Gb/s > 75w
- 16x Arm A78 cores @2.0GHz
- Gen5 x16 + x16
- 48GB DDR5
- Self-Host support

SSG-229J-5BU24JBF

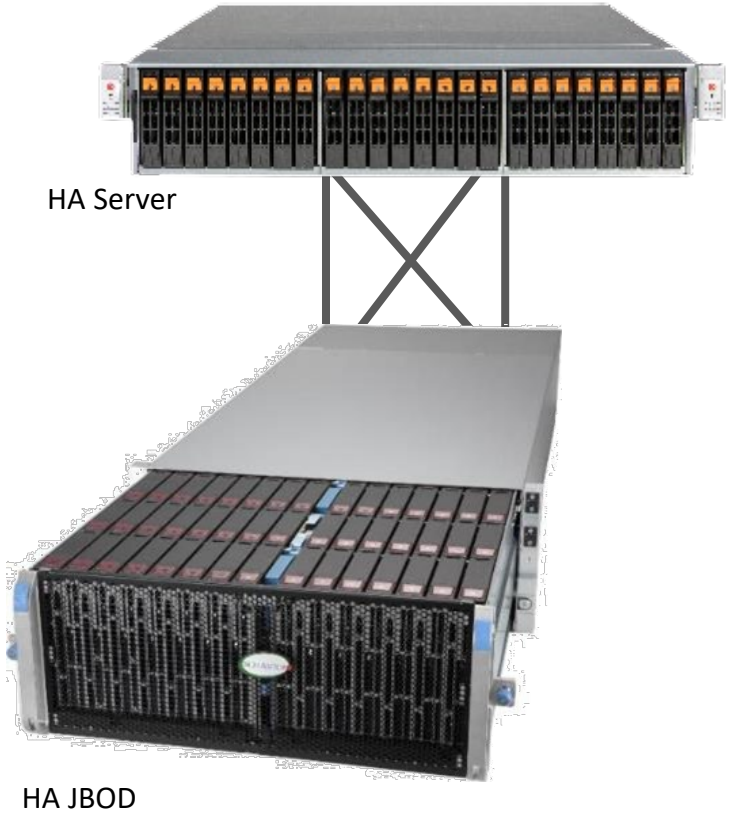
SSG-229J-5BE36JBF

Hot-swap Field Serviceable Architecture

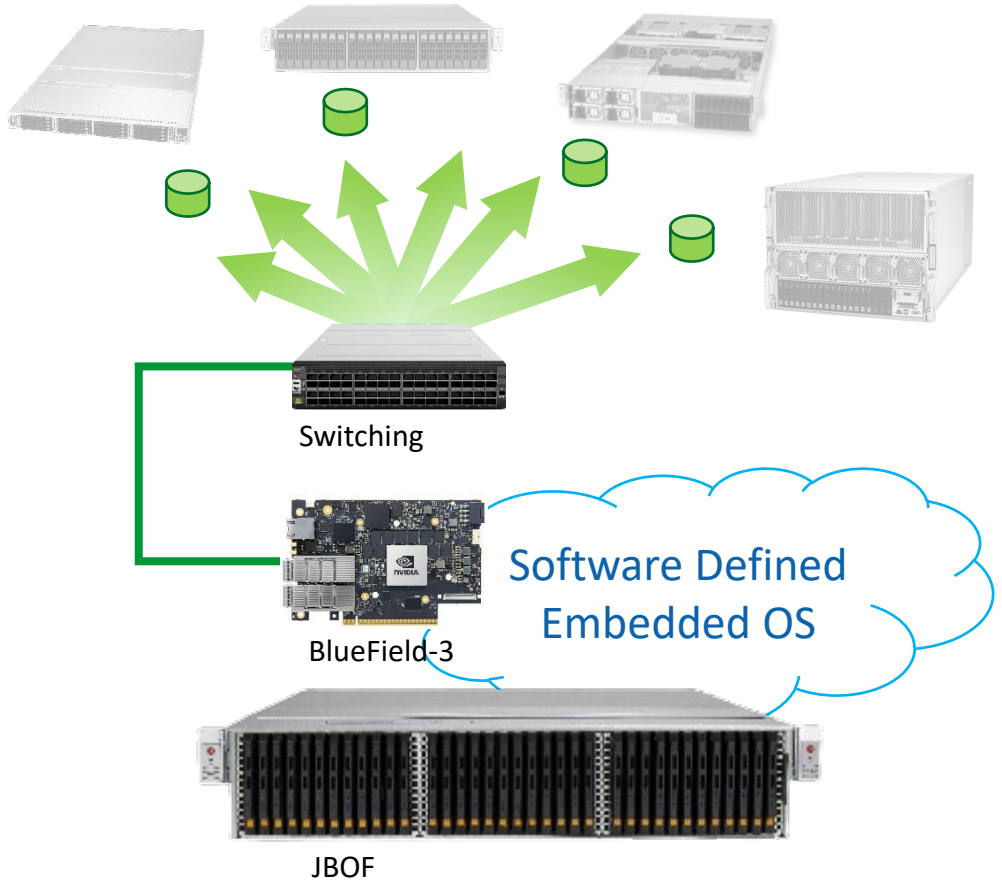
JBOD vs JBOF

Confidential

SAS/JBOD – DAS Expansion Chassis



NVMe-oF / JBOF– *Software Defined Connectivity*



Modular Architecture = Flexibility

Confidential

DPU

Optional expansion:

GPU

NIC for BF-3 in slot 1
Or
Second BF-3, running independent of Slot1
(Separate, Server Node)

Slot1

Slot2

Slot3

JBOF Architecture

PCIe Switch

Canister 1

Canister 2

Single Port

E3.S
0-17

Single Port

E3.S
18-36

Dual Port

U.2
0-23

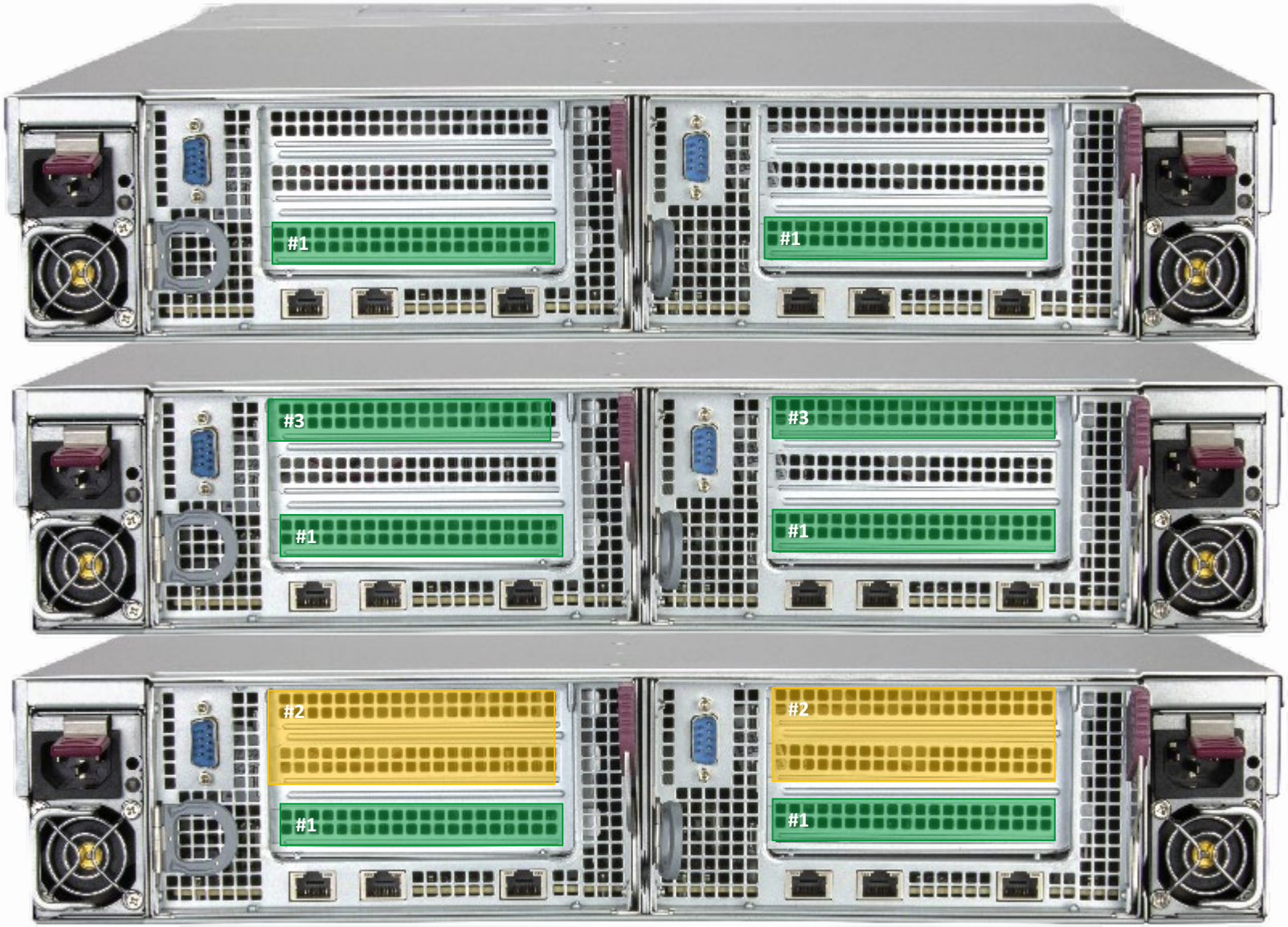
Software Defined Example Configurations

Confidential

1 NIVIDIA BF-3 per Node
TCO Optimized (Default)

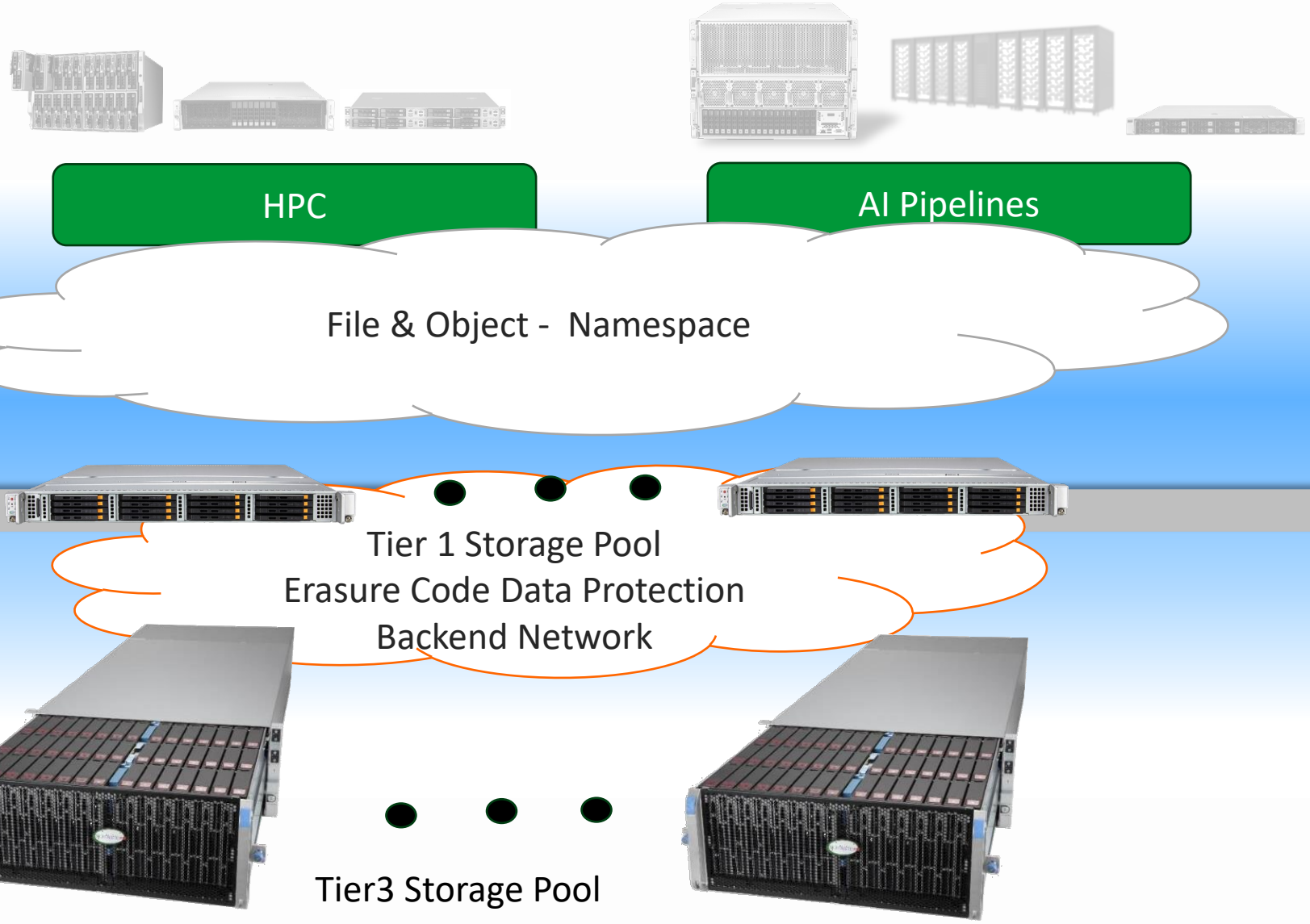
2x NIVIDIA BF-3 per Node
Single Port Performance
Optimized

1x NVIDIA BF-3 + 1x GPU per Node



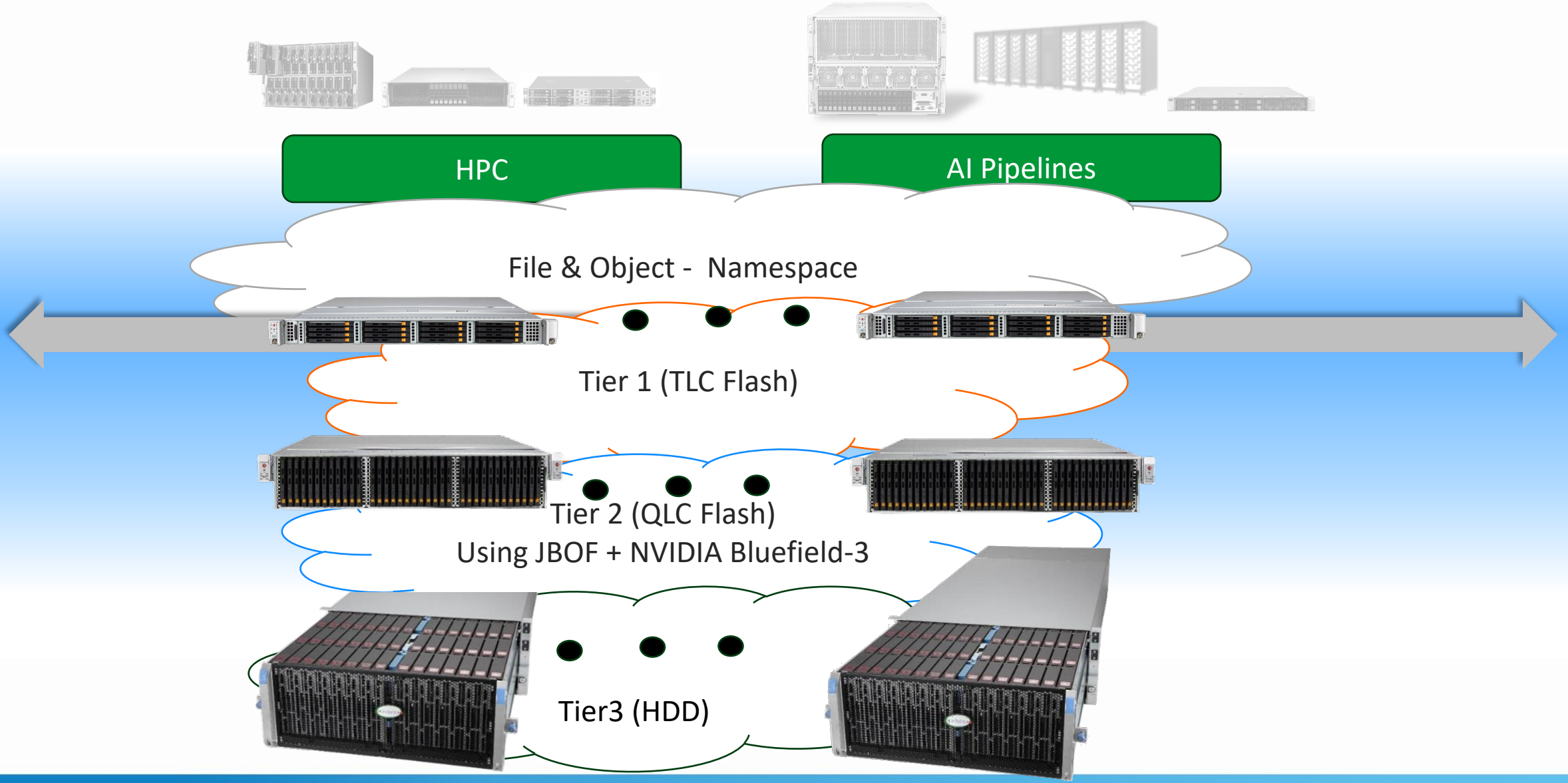
Scale-out share nothing with EC

Confidential



Scale-out share nothing with EC

Confidential



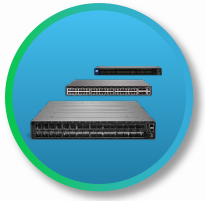
NVIDIA Spectrum-X

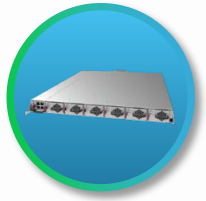
Discussion on scale-out networking efficiency
Spectrum-X in AIimproving flow
BF-3 Offloads
TCO considerations

Building Block Approach, Best TCO for Today's Datacenters

Confidential


Chilled Door


Switches


Power Shelf

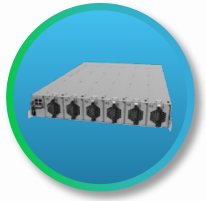

In row DLC

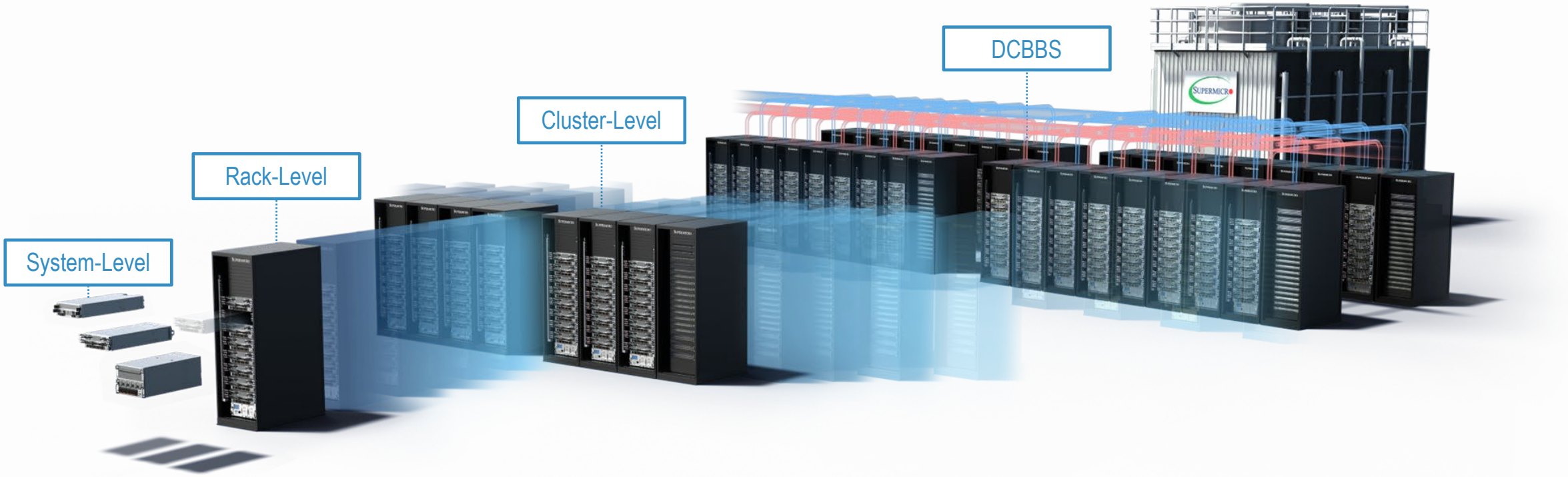

Storage


Cooling Tower


Dry Tower


In-Rack DLC


Battery Back Up



DCBBS provides everything required to fully outfit liquid cooled AI data centers

DISCLAIMER

Super Micro Computer, Inc. may make changes to specifications and product descriptions at any time, without notice. The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions and typographical errors. Any performance tests and ratings are measured using systems that reflect the approximate performance of Super Micro Computer, Inc. products as measured by those tests. Any differences in software or hardware configuration may affect actual performance, and Super Micro Computer, Inc. does not control the design or implementation of third party benchmarks or websites referenced in this document. The information contained herein is subject to change and may be rendered inaccurate for many reasons, including but not limited to any changes in product and/or roadmap, component and hardware revision changes, new model and/or product releases, software changes, firmware changes, or the like. Super Micro Computer, Inc. assumes no obligation to update or otherwise correct or revise this information.

SUPER MICRO COMPUTER, INC. MAKES NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE CONTENTS HEREOF AND ASSUMES NO RESPONSIBILITY FOR ANY INACCURACIES, ERRORS OR OMISSIONS THAT MAY APPEAR IN THIS INFORMATION.

SUPER MICRO COMPUTER, INC. SPECIFICALLY DISCLAIMS ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR ANY PARTICULAR PURPOSE. IN NO EVENT WILL SUPER MICRO COMPUTER, INC. BE LIABLE TO ANY PERSON FOR ANY DIRECT, INDIRECT, SPECIAL OR OTHER CONSEQUENTIAL DAMAGES ARISING FROM THE USE OF ANY INFORMATION CONTAINED HEREIN, EVEN IF SUPER MICRO COMPUTER, Inc. IS EXPRESSLY ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

ATTRIBUTION

THANK YOU

www.supermicro.com