STORAGE DEVELOPER CONFERENCE

A SNIA Event

SDC22 | Fremont, CA
September 12-15, 2022

BY Developers FOR Developers

Next-Generation Storage Will Use DPUs Instead of CPUs

Jai Menon

Chief Scientist

Agenda

1. DPU Introduction
2. Fungible Data Processing Unit (DPU)
3. A DPU-based Storage (DBS) Implementation
4. Comparing Architectures:
   Traditional CPU-based Storage (CBS) vs. DBS
5. Comparing real-world implementations:
   CBS vs. Fungible DBS
   - Performance
   - Storage efficiency
   - Power and Rack Density

DPU Introduction

DPUs will be an Essential Part of Next-Generation Cloud Data Centers

FUNGIBLE

DPUs Have Emerged to Address Two Data Center Mega-Trends

1 Rise of Data-Centric Tasks
(general purpose processors are inefficient at this)

Networking, Storage, Security
Big Fast Data, AI/ML, data analytics

DATA - CENTRIC CLOUD

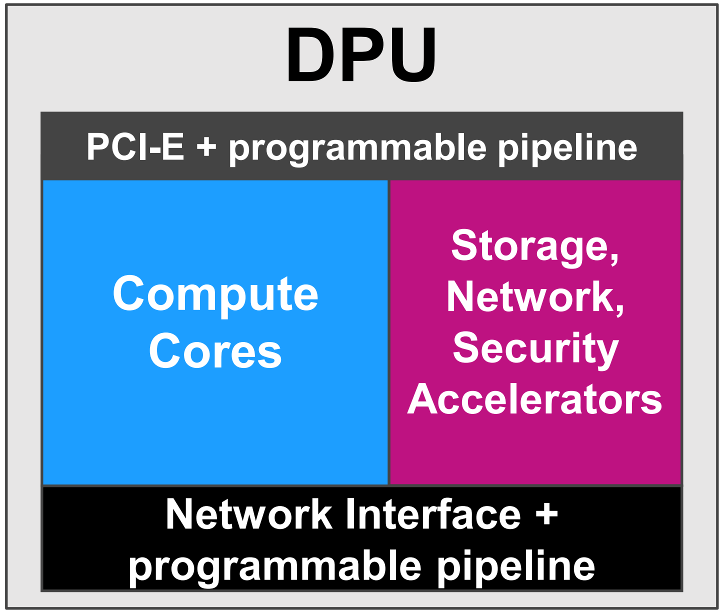
Agility, flexibility, reliability of cloud

2 Data Center Cloudification
(existing data center networks and data center architectures are inefficient at this)

*** Stateful processing of multiple high bandwidth streams of packetized data as needed for networking, storage, security, AI/ML**

What is a DPU?

- **A DPU or data processing unit is a specialized programmable processor tailored to efficiently execute data-centric tasks**
  - **they integrate general-purpose cores & h/w accelerators**
- **Data-centric tasks involve stateful, multiplexed processing of high bandwidth streams of data**
  - **Storage, network and security processing are data-centric**
- **DPUs complement CPUs & GPUs and will be a 3rd socket in data centers**



Two Classes of DPU-Based Products in the Industry

Offload DPU Products

Offload Networking

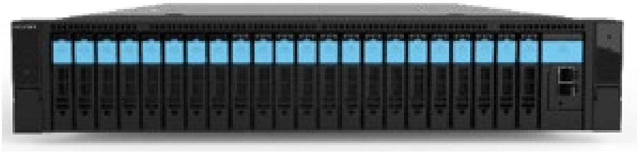
(vRouter – save 7 cores, improve latency by 10X, throughput by 2X)

Offload Storage (previous talk)

(NVMe/TCP Storage Initiator - e.g - save 7-8 cores per million IOPs, improve perf by 2X)

Standalone DPU products

Focus of this Talk

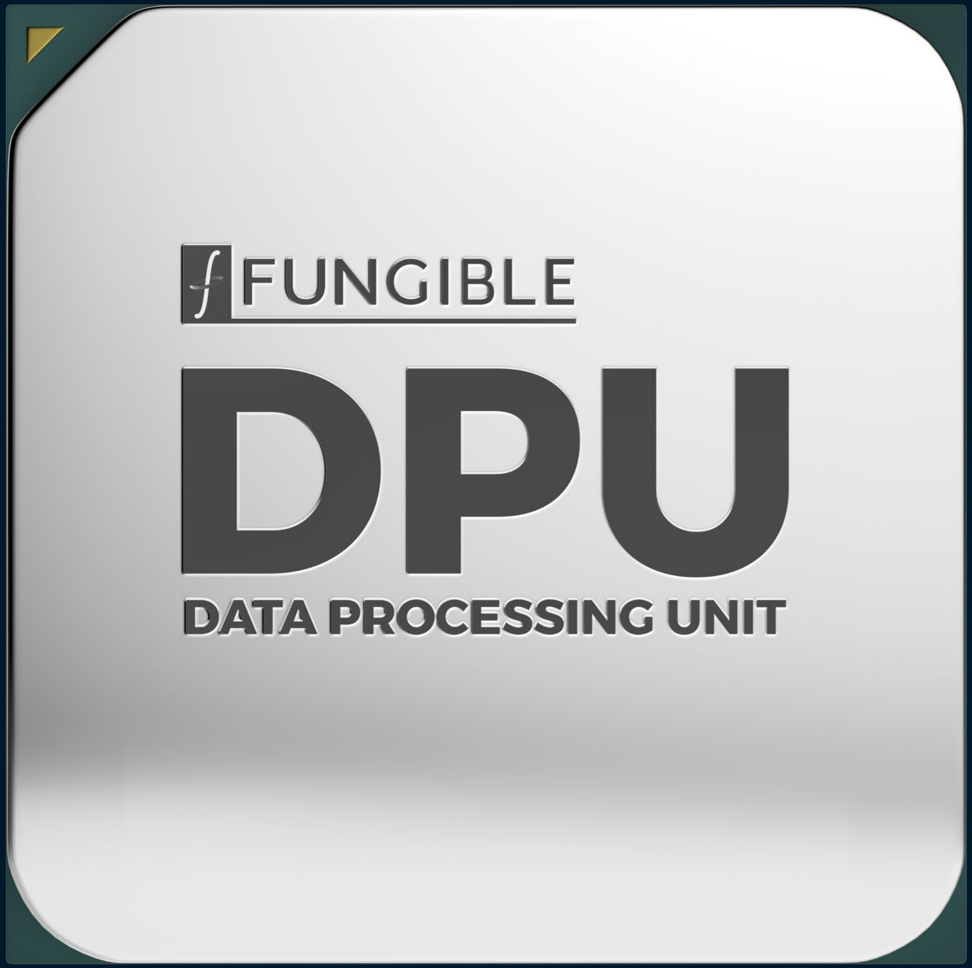


High performance storage
Computational storage
10X speedup for storage targets; 20x speedup for some TPC-H

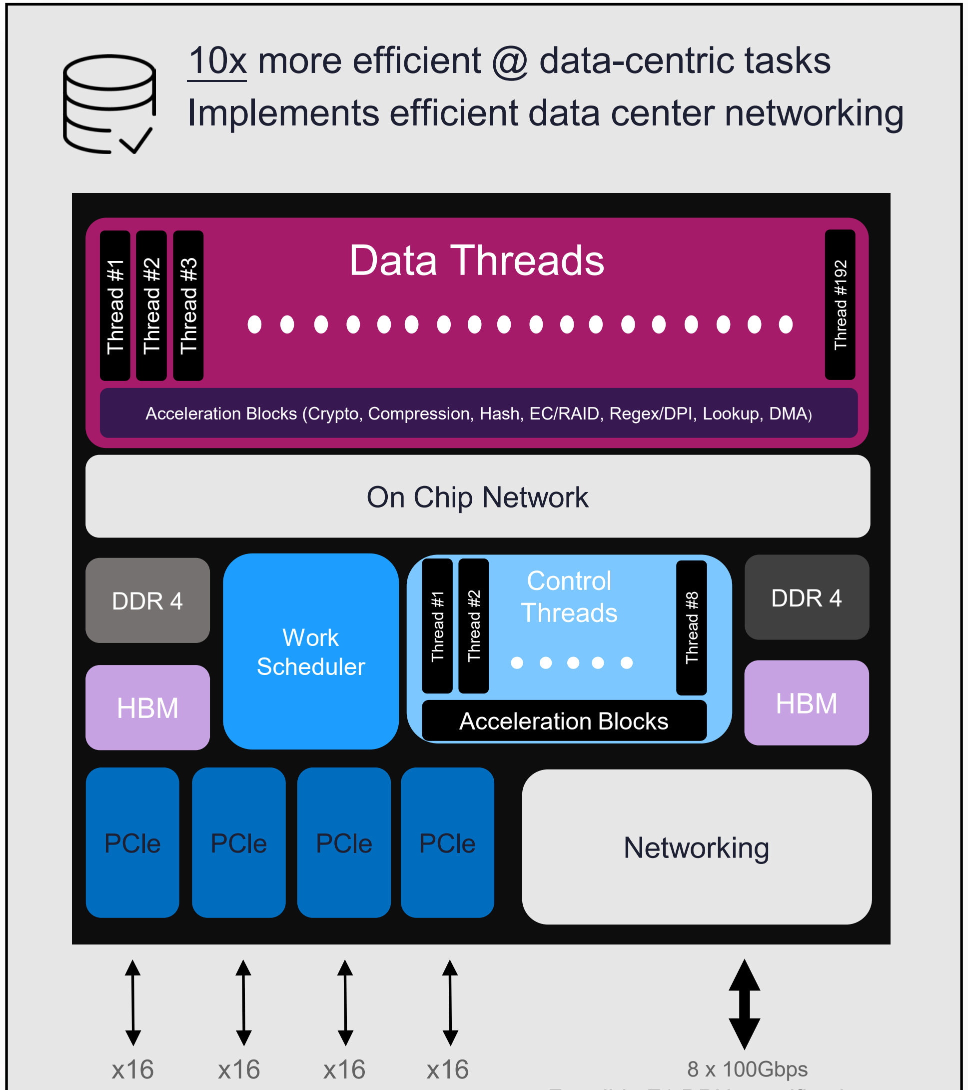
SDN Appliance
Overlay network processing @3X scale & 5X CPS

Fungible DPU

Built for Storage



The Fungible DPU is Built for Storage



Fungible F1 DPU specifications

Network

- **Efficient TCP**
- **TrueFabric™**
- **Transit & Endpoint functionality**
- **P4 programmable Transit Path**

Compute

- **Scheduling of run to completion handlers**
- **Interrupt free**

PCIe

- **Expose multiple personalities – root complex, end point, switch**
- **High performance DMA**

Specialized memory systems

- **HBM – 8 GB @ 4 Tbps**
- **Buffer memory for payload**
- **DDR4 - Upto 1 TB@300 Gbps**

Specialized Flexible Hardware Accelerators

- **EC, compression, regex encryption, Lookup, DMA**
- **Accessible in 10s of nsecs**

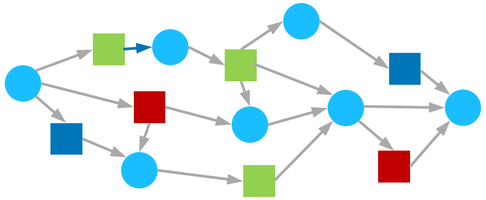
High-Speed Accelerators in the Fungible DPU

Used for Storage Target →

800 Gbps

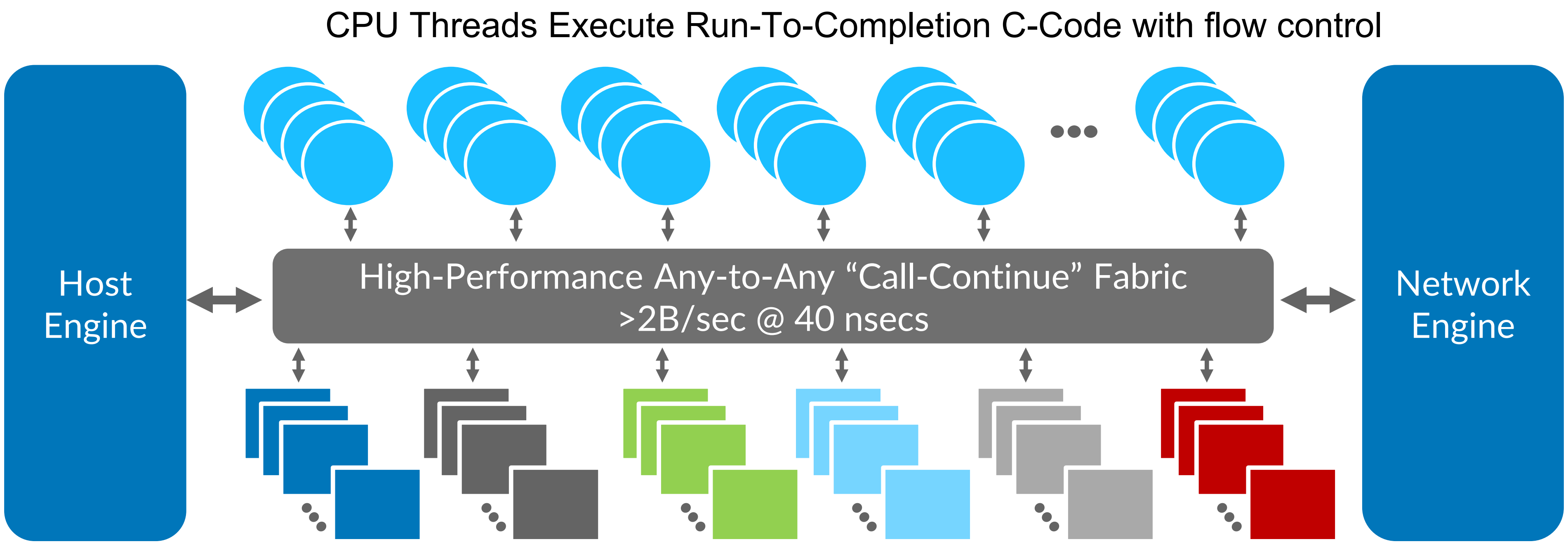
ACCELERATOR	F1 DPU
Flexible DMA	4 Tbps
Crypto (AES-GCM/XTS)	1 Tbps
SHA1, 2, 3 Hash	1 Tbps
Lookups (per sec)	320M
Compress/Decompress	512 Gbps
EC/Raid	800 Gbps
Regex Engine	100-400Gbps

High-Performance Programmable Data Path



100s of concurrent active flows

Millions of dormant flows



Heterogeneous Accelerator Threads
Fungible DPU is unique in the tight coupling of cores to accelerators

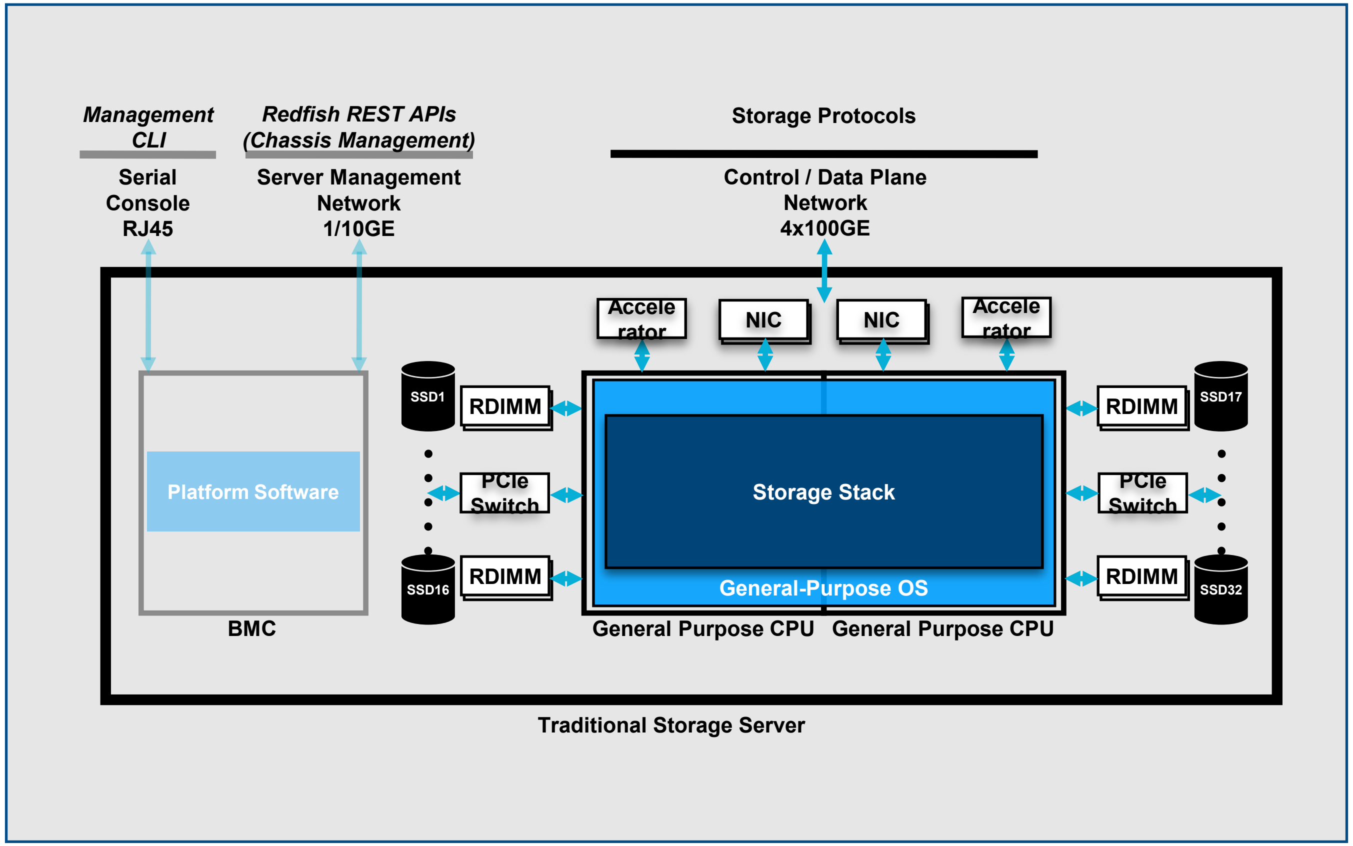
STORAGE DEVELOPER CONFERENCE SDC 22

Comparing DPU-based Storage (DBS) and Traditional CPU-Based Storage (CBS) Architectures

Traditional Way to Build Storage (the old way)

Main Components & Software

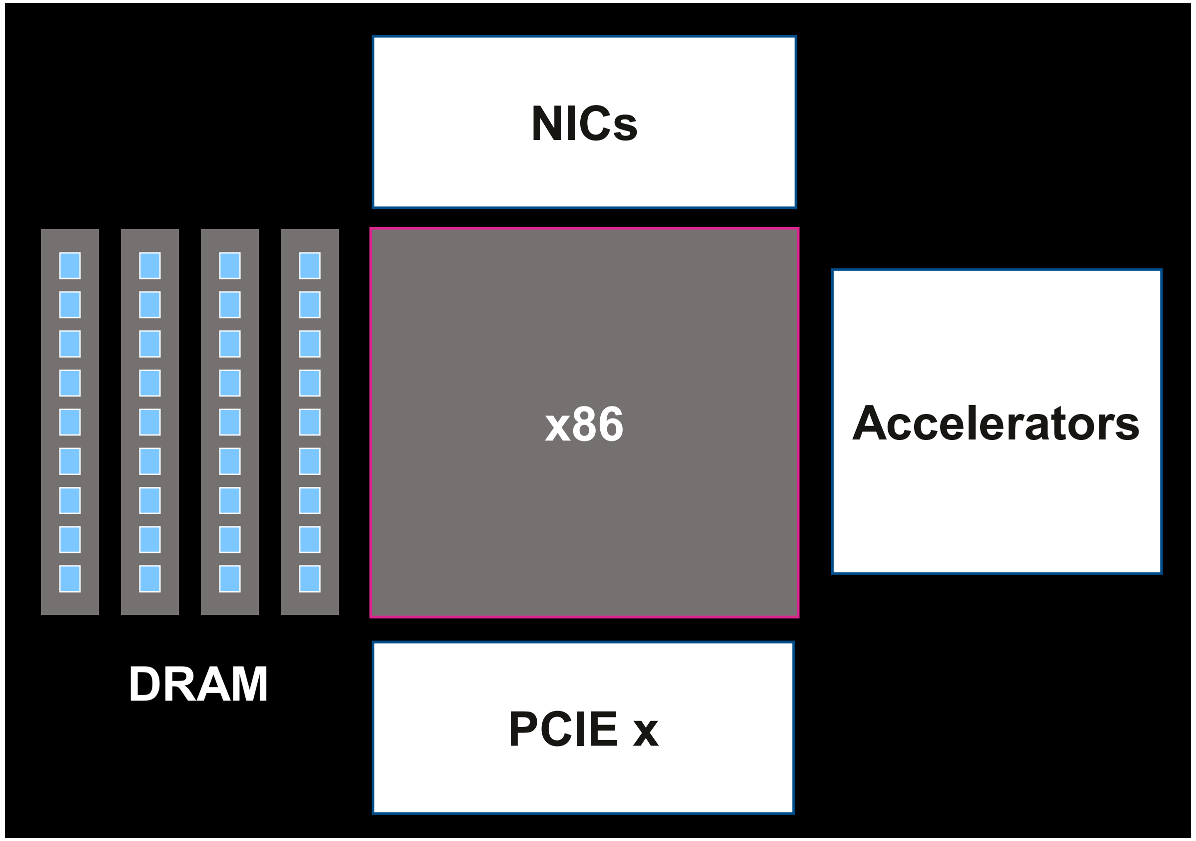
- **2 High End x86 Processors**
  - General Purpose OS
    - Storage Stack
- **Many physical IO devices**
  - 4 100Gbps NICs
  - 2 100Gbps Data Security & Data Reduction Accelerators
  - PCIe Switches to connect to many SSDs



STORAGE DEVELOPER CONFERENCE SDC 22

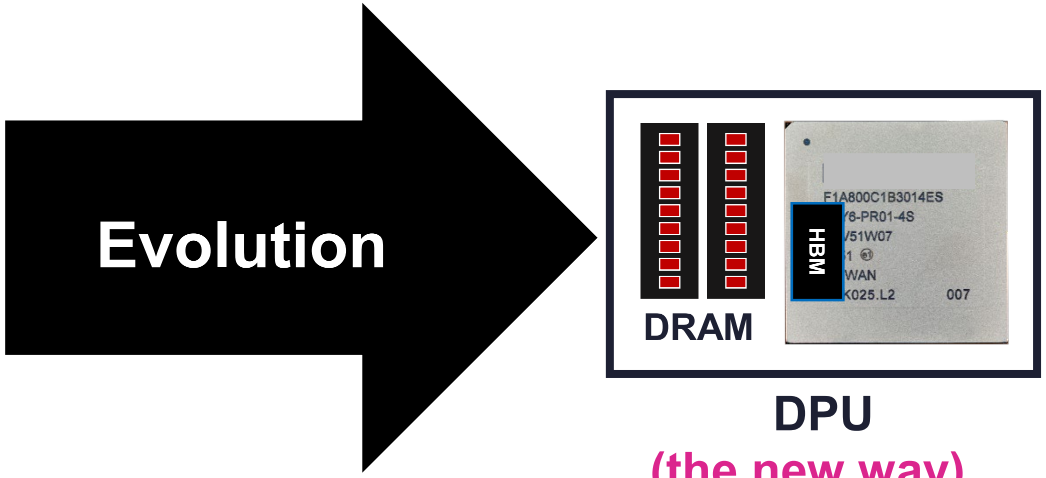
The New Way to Build Storage - DPU-Based Storage (DBS)

550 W, $4000 (without SSDs or DRAM or Accelerators)



Typical CPU Based Storage uses discrete parts
(the traditional way)

85 W, $1000-$1500 (w/o SSDs, DRAM)



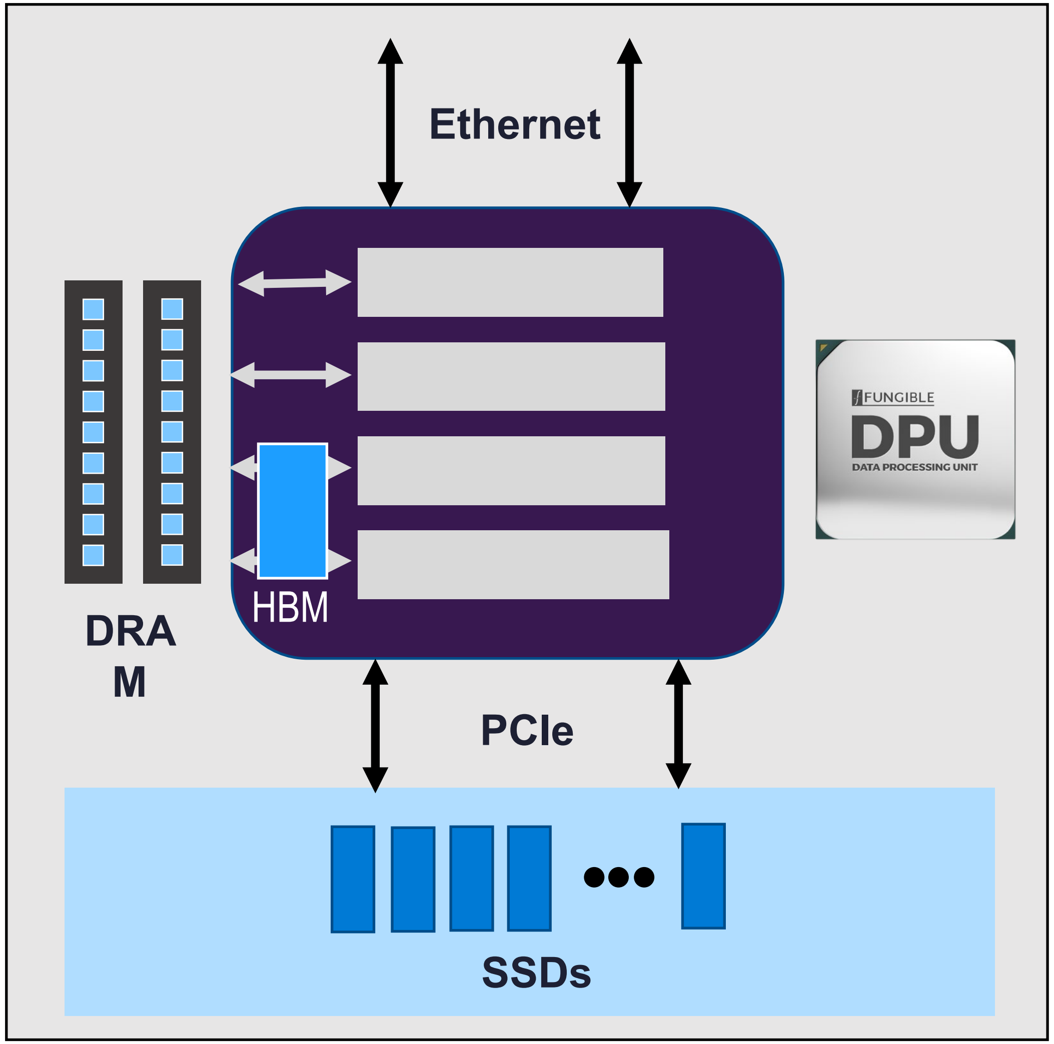
DPU
(the new way)

Lower cost, Lower power, Higher reliability, Higher performance

Why is Fungible DPU Important for Storage?

A Storage Workload has Special Requirements

- Handle Multiple (10s of 1000s) concurrent streams of data
    - CPUs have low IPC for multiplexed workloads
- Requires termination before processing
    - Packets to/from network; TLPs to/from PCIe
    - CPUs inefficient at termination handling
- Multiple passes needed over data
    - Compression, Encryption, Erasure coding
    - Stresses DRAM BW of CPU Based Storage (CBS)
- Needs separate memory for data & state to avoid cache pollution
- Needs accelerators for data reduction, security, protection



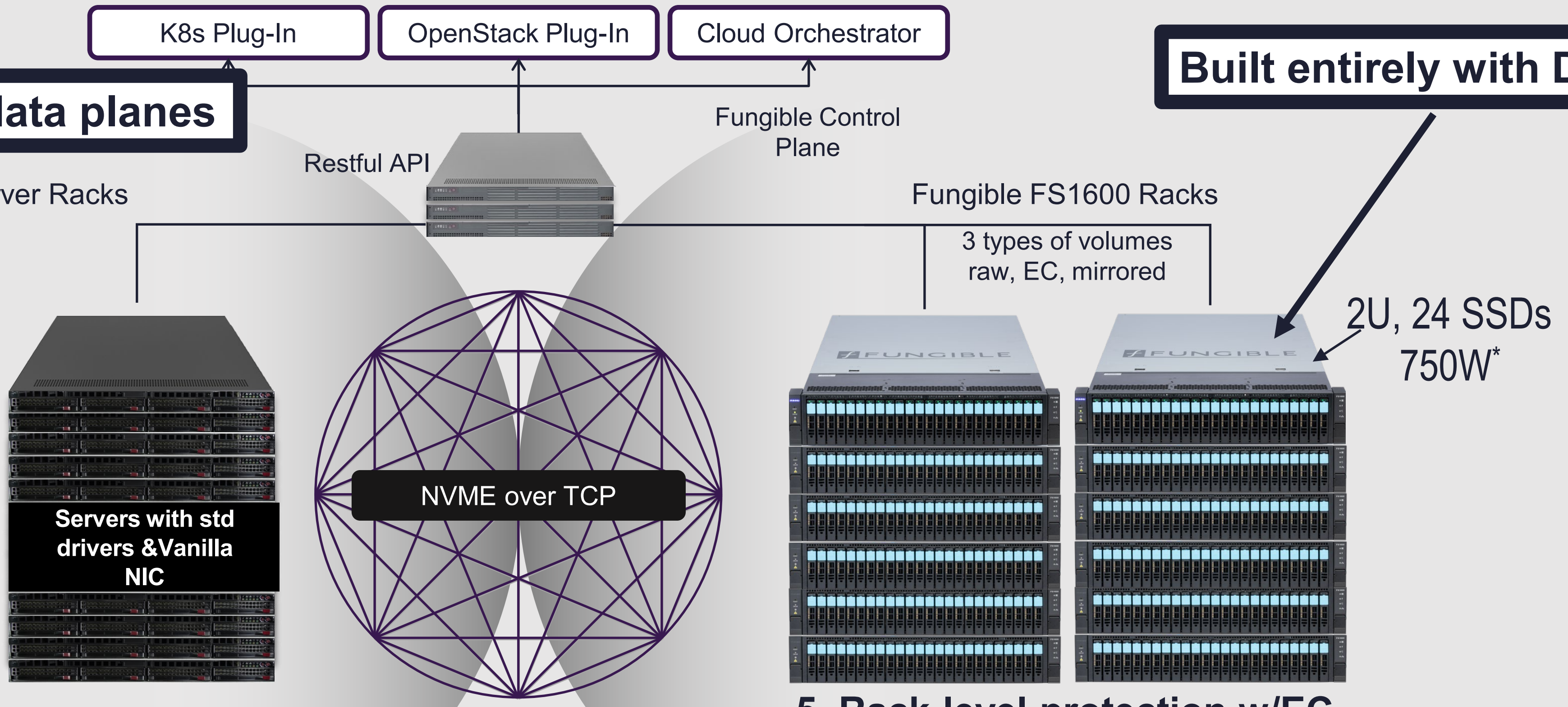
 |

DBS Implementation

Fungible Storage Cluster (FSC) – First DBS Implementation

Separate control & data planes

Built entirely with DPUs



5. Rack-level protection w/EC

1. High-Performance

- 13M IOPS / Node
- 120 µs
- 75 GB/s / Node

2. Scale-Out

- 70 TB → Petabytes
- 3 Nodes → Data Center
- Linear Performance Scaling

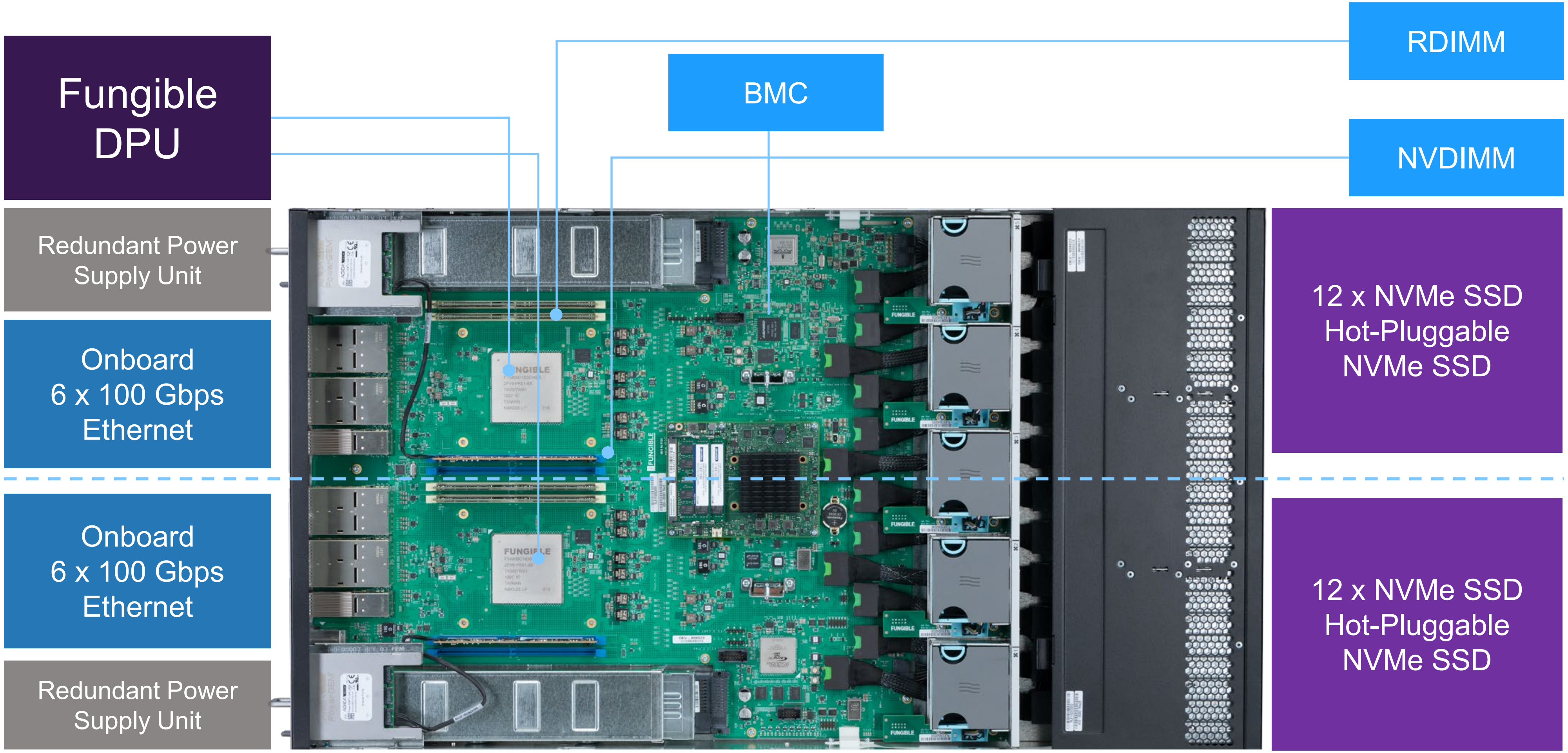
3. Full Set of Features Selectable by Workload

- Compression, Thin Provisioning, Snapshots, Clones, QoS
- Encryption
- Erasure Coding, Mirrored & Raw

4. Low Power, High Rack Density

750W vs. 1500W - 2000W for CBS

FS1600 Under the Hood – Only DPUs, no CPUs



FSC Performance - IOPS

	BLOCK READS
Raw Single Node	15M 4KB IOPS (4 KB) 75 GBytes/sec (16 KB)
Network Protected (RF=2) Two Nodes (SSD and node failure protection)	15M 4KB IOPS (4 KB) 120 GBytes/sec (16 KB)
Network Protected 4+2 EC 6 nodes (SSD & node failure protection)	20M KB IOPS (4 KB) 160 GBytes/sec (16 KB)

- Linear performance scaling measured up to 16 nodes, expect continued linear scaling beyond this
- Database performance – equal to DAS with EC
- <5% impact with compression and encryption turned on

IOPS and Latency - EC(4+2) with 6 FS nodes – 4K IOPS

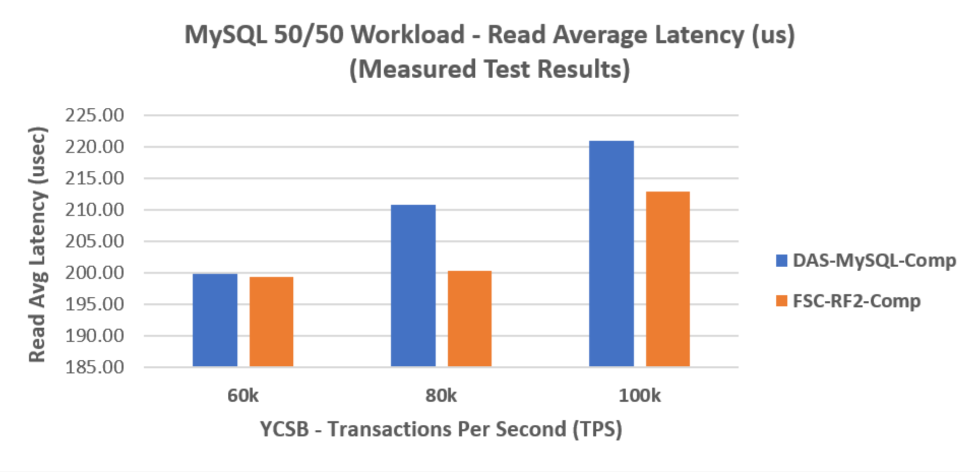
Operation	Workload	4K	8K	16K
No Compression & Encryption	Random Read	18.36M@485us	28.69M@469us	34.63M@533us 52M@519us, QoS disabled
	Random Write	5.12M@213us	6.65M@161us	6.11M@177us
With Compression & Encryption	Random Read	17.99M@496us	28.68M@470us	33.80M@549us
	Random Write	5.93M@178us	9.18M@235us	10.47M@204us

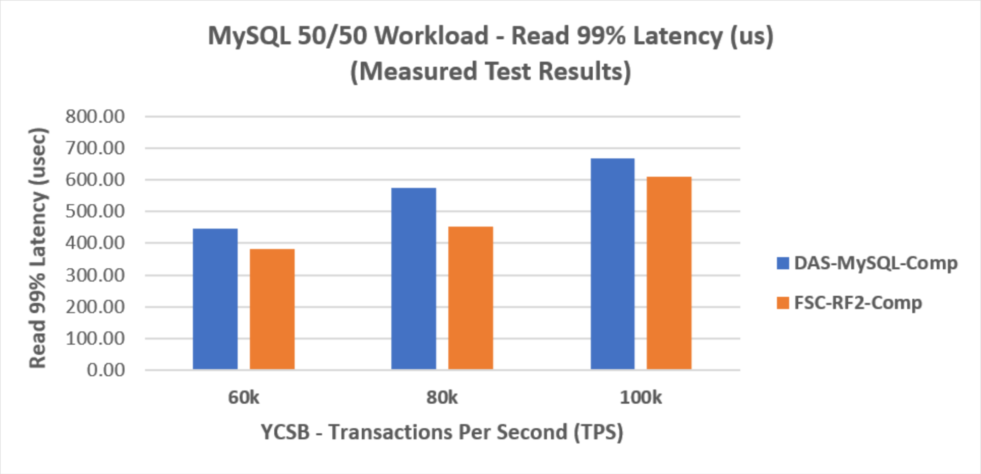
99th percentile latency near 1 msecs for reads; and 369 usecs for writes

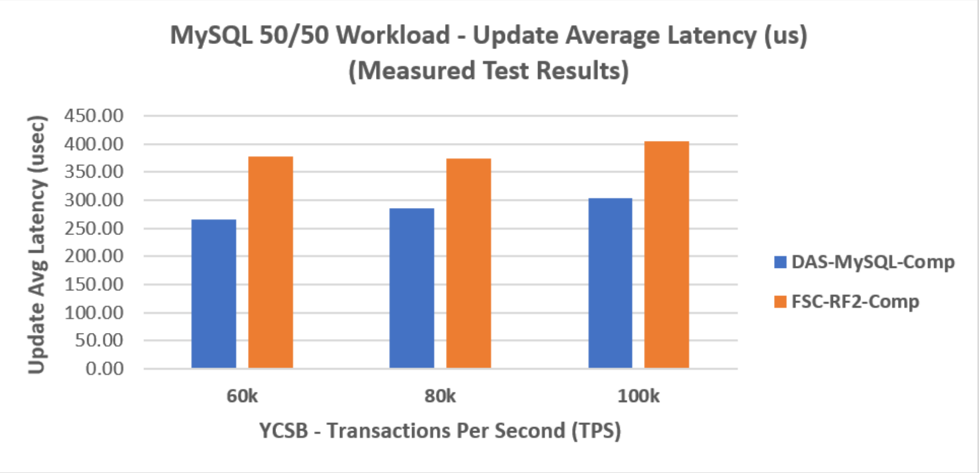
Fungible Storage Performance on MySQL Database

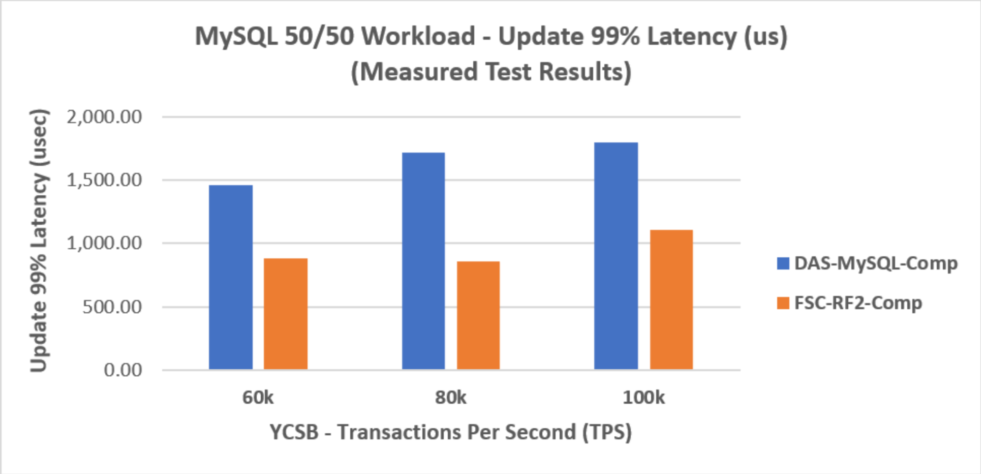
DPU-based storage can be as fast as locally attached storage

- MySQL 8.0
- XFS filesystem
- Innodb storage engine
- Innodb_buffer_pool_size= 16G
- DAS w/ MySQL table compression “zlib”
- FSC compression but no MySQL table compression
- Yahoo! Cloud Serving Benchmark (YCSB)
- 4KB record size
- 32,000,000 record count





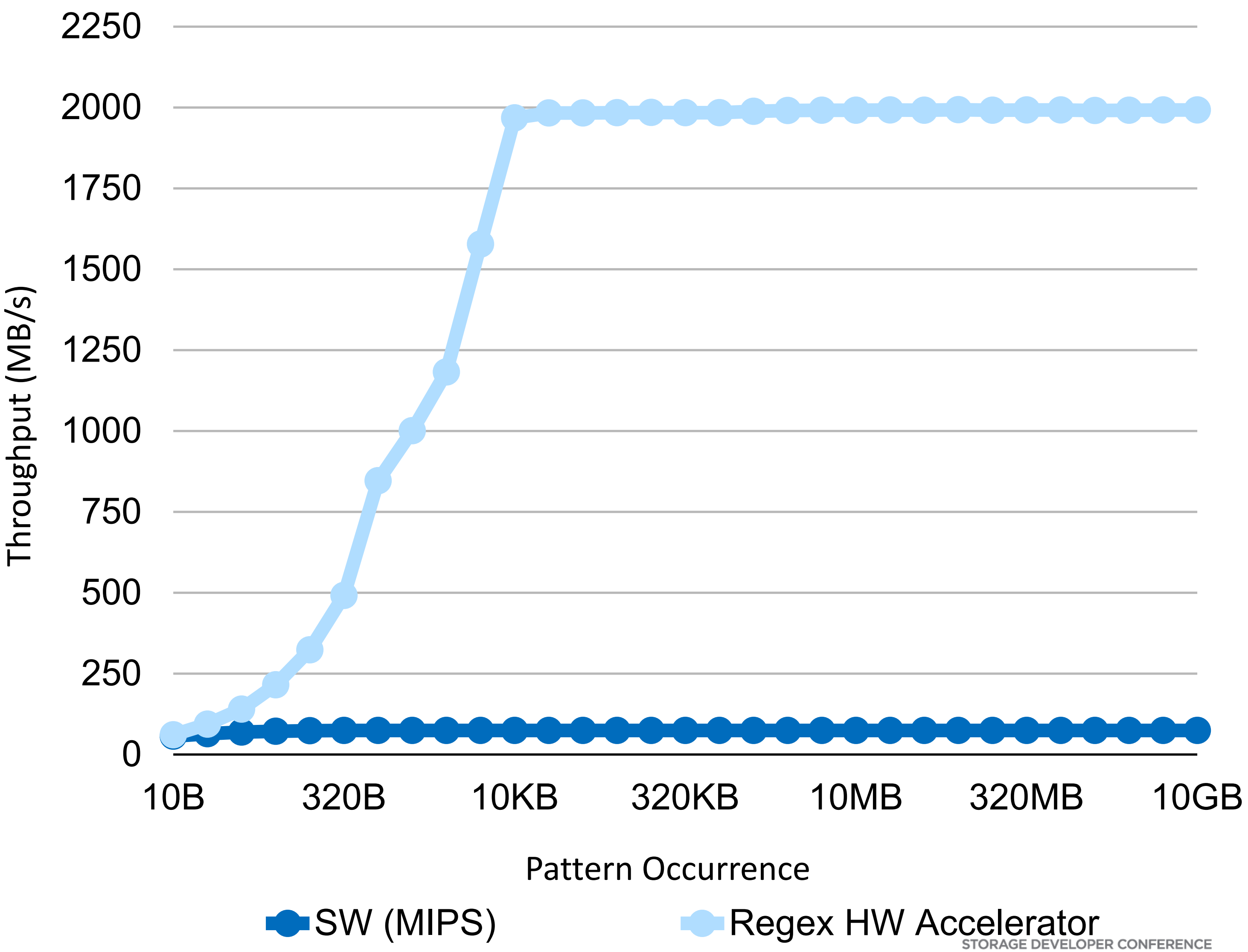




LOWER IS BETTER

Computational Storage --Regex Pattern Matching

- SW Pattern Matching (MIPS): ~75 MB/s
- Perl (x86): ~140 MB/s
- Grep (x86): ~200 MB/s
- Regex (DPU, single cluster): ~1900 MB/s
- Performance scales rapidly as frequency of matches drops
- Complexity of pattern has very minor impact



FUNGIBLE

DBS vs. CBS

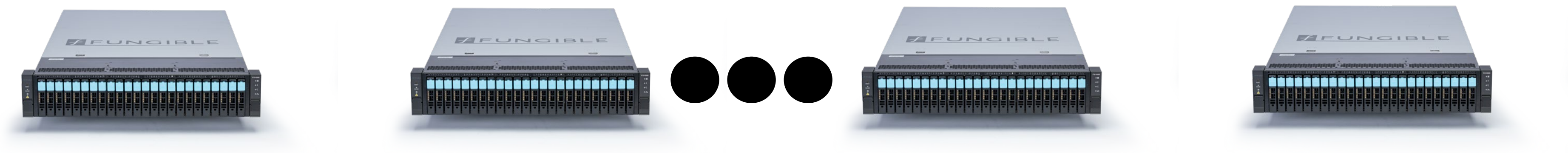
Comparing Storage Efficiency, Power, Rack Density & Performance

DBS is More Cost Efficient for 3 Reasons

Low overhead durability (25% for 8+2 EC vs. 200% for 3-way replication)

- **EC uses Reed Solomon codes needing Galois field math which DPU is good at**
- **Durability needs efficient networking which DPU is better at**

8+2 EC



25% overhead versus 200% overhead
6.7% overhead with 30+2 EC

Superior compression (e.g. 3X vs. 2.5X) at line rate – minimal performance impact with in-line compression

Encryption without self-encrypting drives which are more expensive

DBS IS MORE STORAGE EFFICIENT
Actual Customer Example

Comparing Raw TB per Effective PB

Storage Requirements : 1PB

	DIRECT ATTACHED STORAGE (DAS)		DBS STORAGE		CPU BASED STORAGE	
	Method	Media Required	Method	Media Required	Method	Media Required
EFFECTIVE STORAGE (TB)		1000		1000		1000
UTILIZATION	60%	1667	80%	1250	80%	1250
2 FAILURE PROTECTION	RF3	5000	8+2 EC	1562.50	RF3	3750
COMPRESSION	1x	5000	3x	520.83	2.5X	1500

- **10X SAVINGS VERSUS DAS (Customer's current environment)**
- **3X SAVINGS VERSUS COMPETITIVE SDS solution that customer looked at**

DBS Has Lower Power

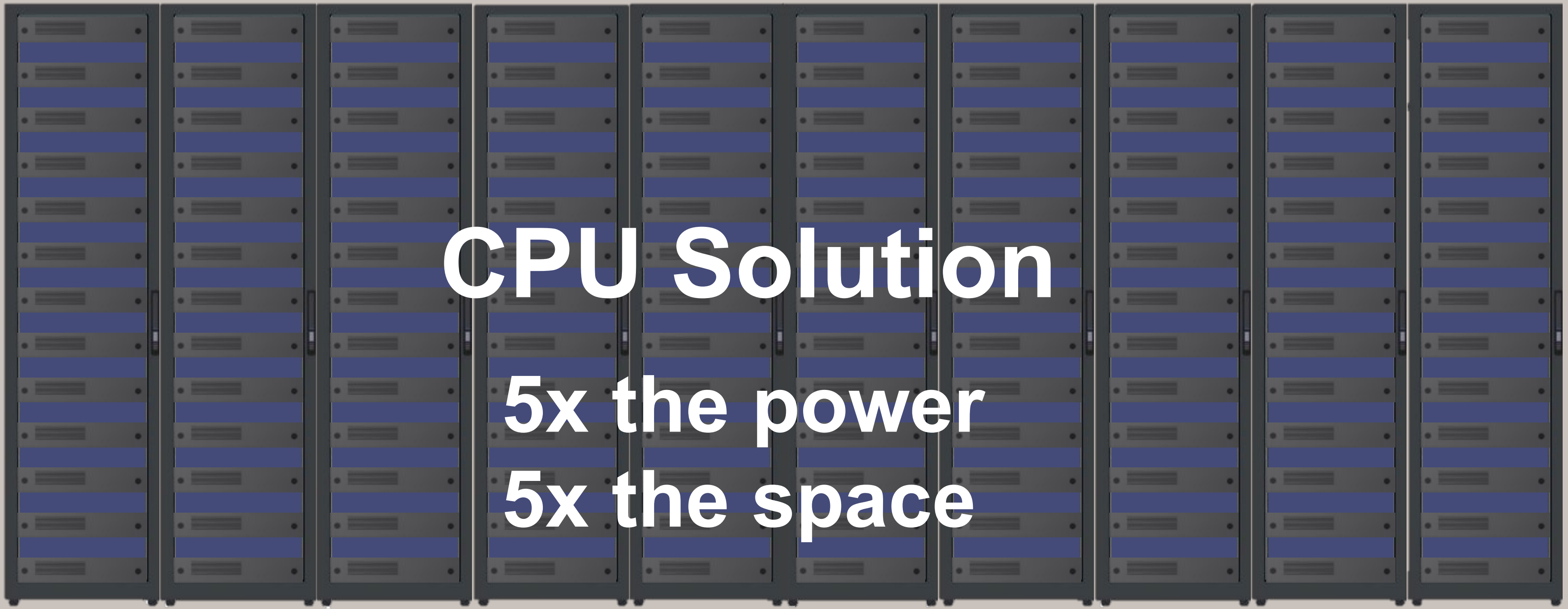
Yearly power savings at 10,000 storage boxes is 54000 MWH
$3M/year saved in power @ 6c/KWH

CPU assumptions
- Intel dual socket Icelake server, 2.5 Ghz Gold, 24 cores
- 2x100 Gbps NICs
- 128 GB DIMMs, 512 GB of NVDIMMs

DPU assumptions
- 2xF1 DPUs
- 8x100 Gbps integrated networking
- 256 GB DIMMs and NVDIMMs

	SDS	DBS	Ratio
Motherboard, DIMMs, networking	800	184	4.4X
2U chassis with 24 SSDs	1400	784	1.8X

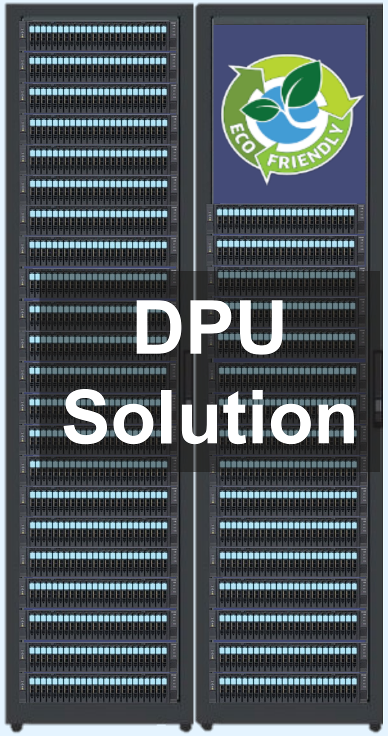
DBS Has Lower Power, Better Rack Density –
Customer Example with 12 KW Racks



CPU Solution

5x the power
5x the space

128 kW
10.7 racks



DPU Solution

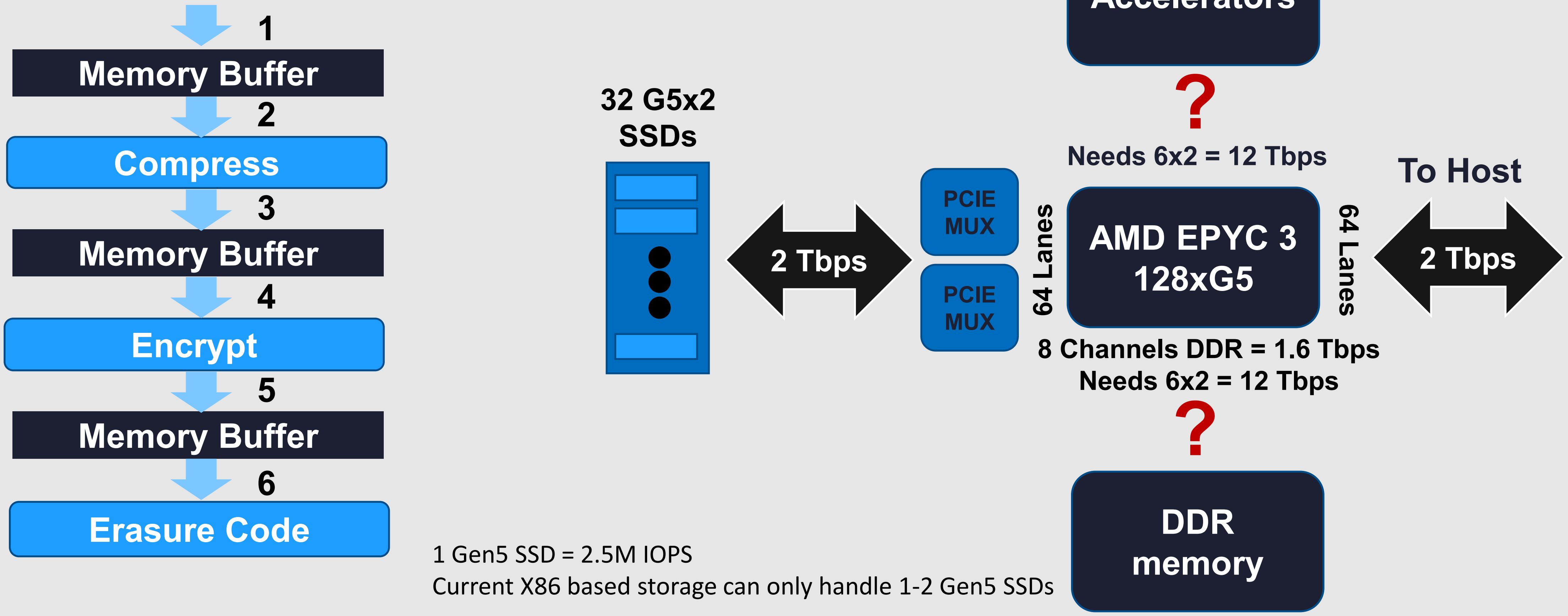
28 kW
2.3 racks

 |

DBS Has Better Performance

CPU Based Solutions have Insufficient Bandwidth and are Missing Accelerators

Storage Pipeline Needs 6-8x Memory BW vs. SSD BW
DRAM BW too slow, on-chip buffers not large enough
HBM is fast enough and large enough



1 Gen5 SSD = 2.5M IOPS
Current X86 based storage can only handle 1-2 Gen5 SSDs

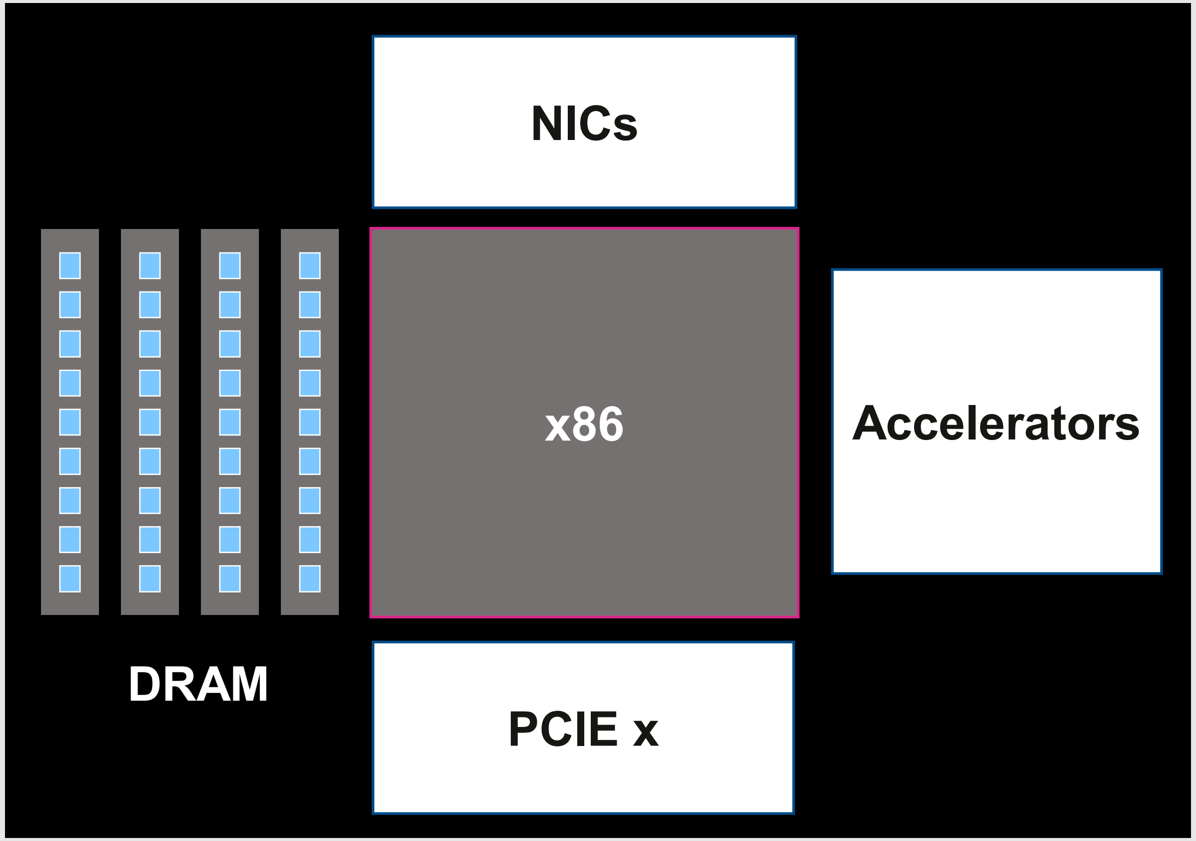
FSC DBS Has Better Performance – Real Examples

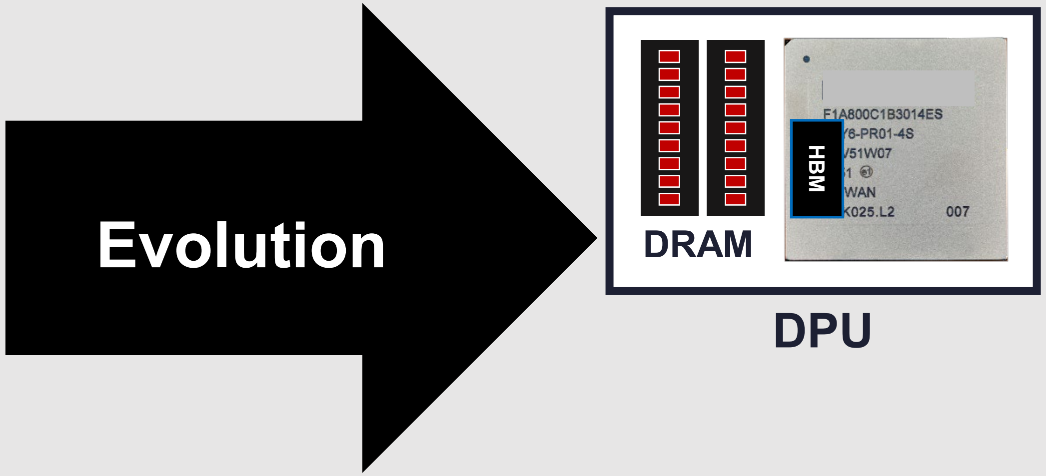
Per socket performance – DPU is 3X to 8X better

Attributes		Best CBS in Production	Fungible DBS in Production	Improvement Factor
Raw	4K IOPs	1.5M - 2.5M	7.5M	3X - 5X
	Bandwidth GB/s	12	37.5	3.1X
2-Way Replication	4K IOPs	1M - 1.5M	4M	2.7X – 4X
	Bandwidth GB/s	6	30	5X
Networked EC	4K IOPs	0.45M	1.8M	4X
	Bandwidth GB/s	1.9	15	8X

Summary -- DBS is the new way to build storage

(the new way)





Typical CPU Based Storage uses discrete parts

(the traditional way)

Comparing CBS vs DBS implementations

Attribute	Best CBS	DBS	Improvement Factor
Performance/W	**7.05 K IOPS/W**	**104.2K IOPS/W**	**14.8x**
Performance	**3M - 5M IOPS**	**15M IOPS**	**3x – 5x**
Power (w/o SSDs)	**800 W**	**184 W**	**4.4x**
Power, Rackspace (w/ SSDs)	**128KW** 10.7 racks	**28 KW** 2.3 racks	**5x**
Storage Efficiency (TB per effective PB)	**1500 TB**	**520 TB**	**2.9x**
Regex	**75 MB/s**	**2000 MB/s**	**26.7x**

High Performance, Low Power, Full Featured

Other Presentations from Fungible

- Next Generation Architecture For Scale-out Block Storage By Jaspal Kohli

- DPU as a Storage Initiator By Pratapa Vataka

Thank You!

Please take a moment to rate this session.

Your feedback is important to us.

FUNGIBLE

Thank you.

STORAGE DEVELOPER CONFERENCE SDC22