SDC 20

BY Developers FOR Developers

Storage Developer Conference
September 22-23, 2020

Emerging Data-centric Storage Architectures

Pankaj Mehra, Ph.D.
VP Storage Pathfinding
Samsung Semiconductor Inc.

Challenges of Data@Scale

Bottlenecks

Processing power and processing bandwidth

Metadata inefficiency of object storage & retrieval

Wire protocol termination for disaggregated flash

Inefficiencies

Inability to deliver both performance and scale

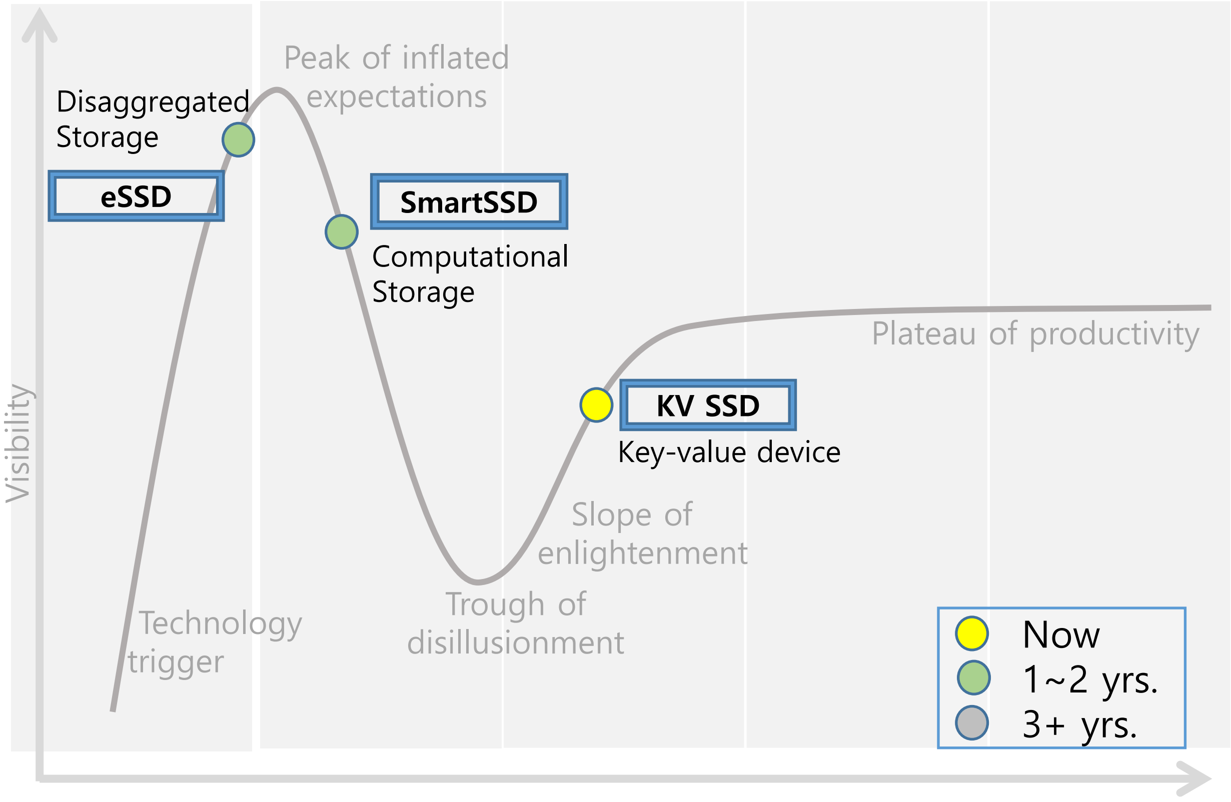
Wasted endurance

Wasted memory BW

CPU overhead of I/O

CPU overhead of I/O virtualization

Good Ideas, Already In-Play



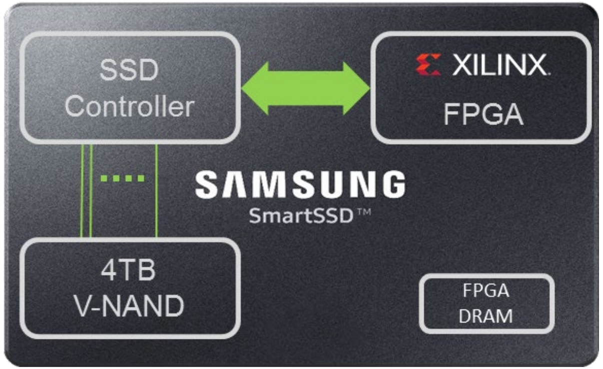
- Virtualization offload
- SMRDB (since HDD days)
- DB filtering acceleration
- Storage NW conv (since FC)
- Active Disk (since HDD days)
- OSD (since HDD days)

Why Revisit?

Because in 2020, three distinct 25-y.o. ideas meet the SSD!

SmartSSD® CSD Scales to Accelerate Data-Rich Workloads

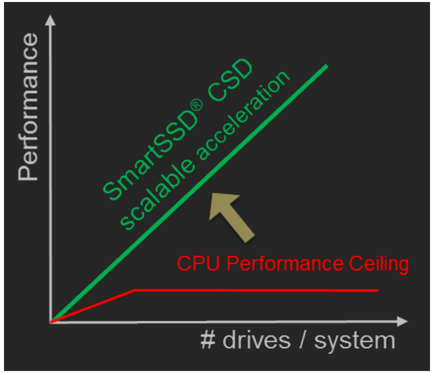
SmartSSD U.2 Platform



Computational Storage

- ✓ **3 & 6 GBps internal BW per device:** Minimize external data movement
- ✓ **FPGA:** Each device has 3x~10x core equivalents for offload/acceleration
- ✓ **4TB storage, 4 GB FPGA DRAM:** For Inline and Data@Rest processing

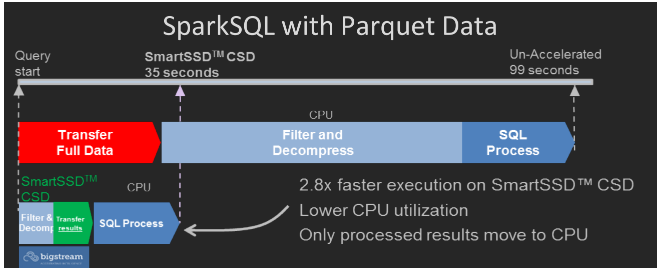
Acceleration Concept



Scalable Performance

- ✓ **Near Data Processing:** Data form conversion, Filtering, Metadata management, DB Analytics, Vide processing
- ✓ **New Services:** Secure content, Edg acceleration

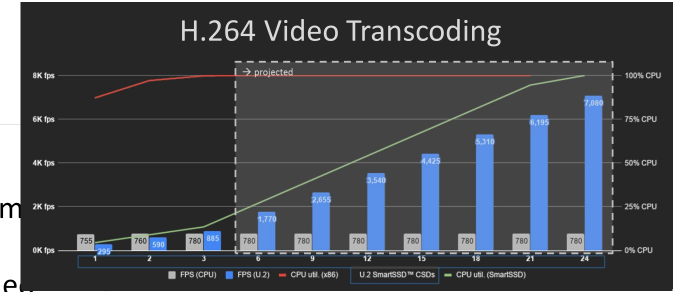
Partner Solutions



2.8x faster execution on SmartSSD™ CSD

Lower CPU utilization

Only processed results move to CPU



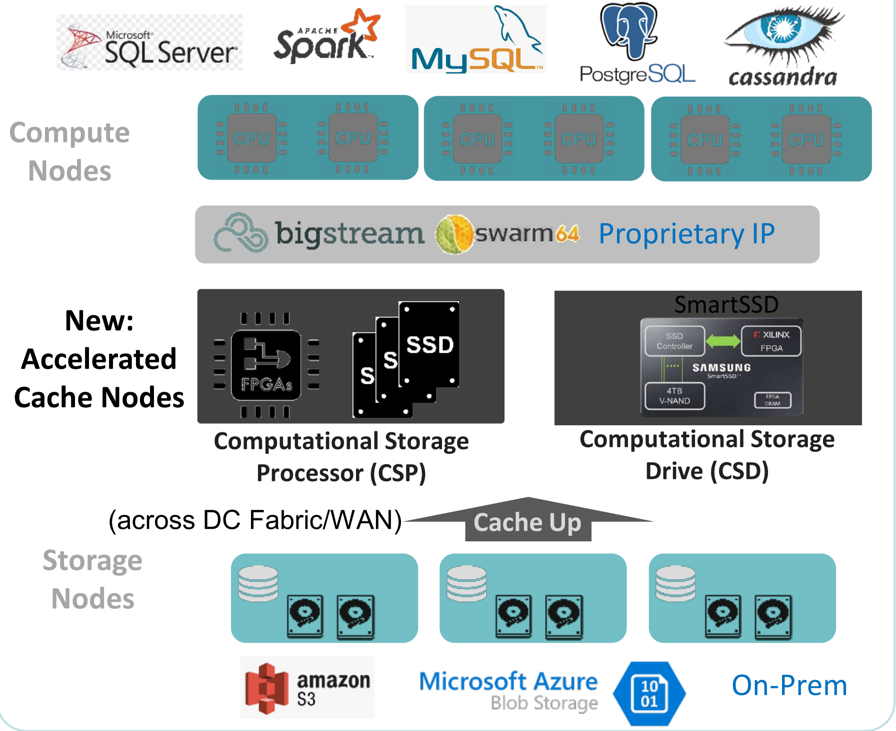
P2P Compression and Decompression

	CR	Throughput (MB/s)	CPU Eff. (MB/s/cpu)	PCIe Eff.	Mem Eff.
Compression: NoLoad-CSD	2.85	2173	2173	1.72	1.42
Compression: NoLoad-CSD w/ p2pdma	2.85	2862	2938	1.01	0.06
Decompression: NoLoad-CSD	2.85	1823	2989	1.71	1.81
Decompression: NoLoad-CSD w/ p2pdma	2.85	2022	3690	1.01	0.14

Computational Storage Use Cases Examples

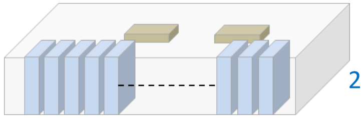
- 3^{rd} party and proprietary acceleration stacks run on Computational Storage to accelerate real-time analytics and regex searches for cybersecurity

Analytics Cache Node

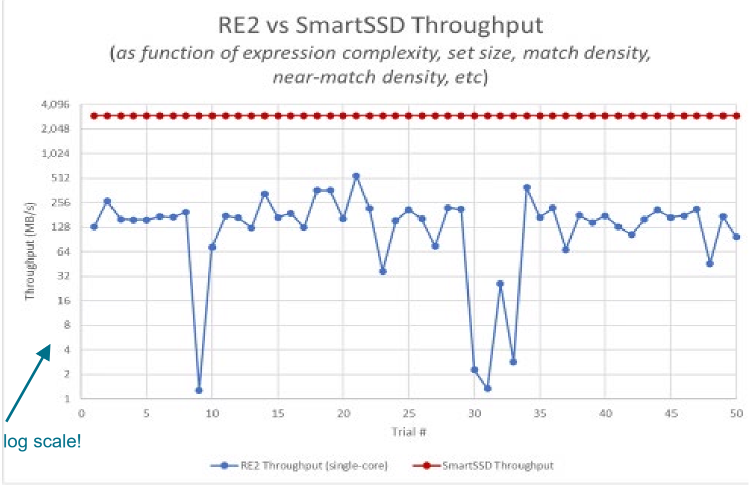


RegEx Appliance

- **Throughput scales to large datasets and complex searches**
- **>10x throughput improvement compared to x86**



RegEx Appliance, 24x SmartSSD, 48TB

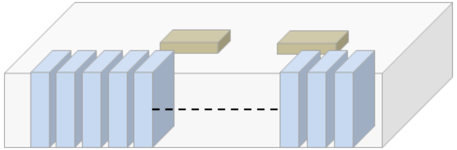


log scale!

Samsung SmartSSD® Technology Roadmap

- Samples, development tools, partners solutions available for immediate PoC
- Customer PoC Test&Dev systems/support available from Samsung and partners

v1.0 SmartSSD® U.2 CSD

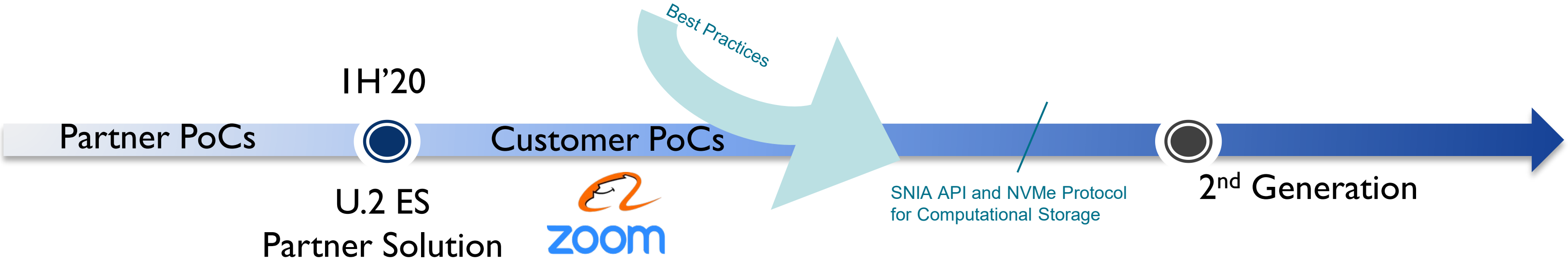


U.2 FF: Scale Processing to 24 ~ 48 devices

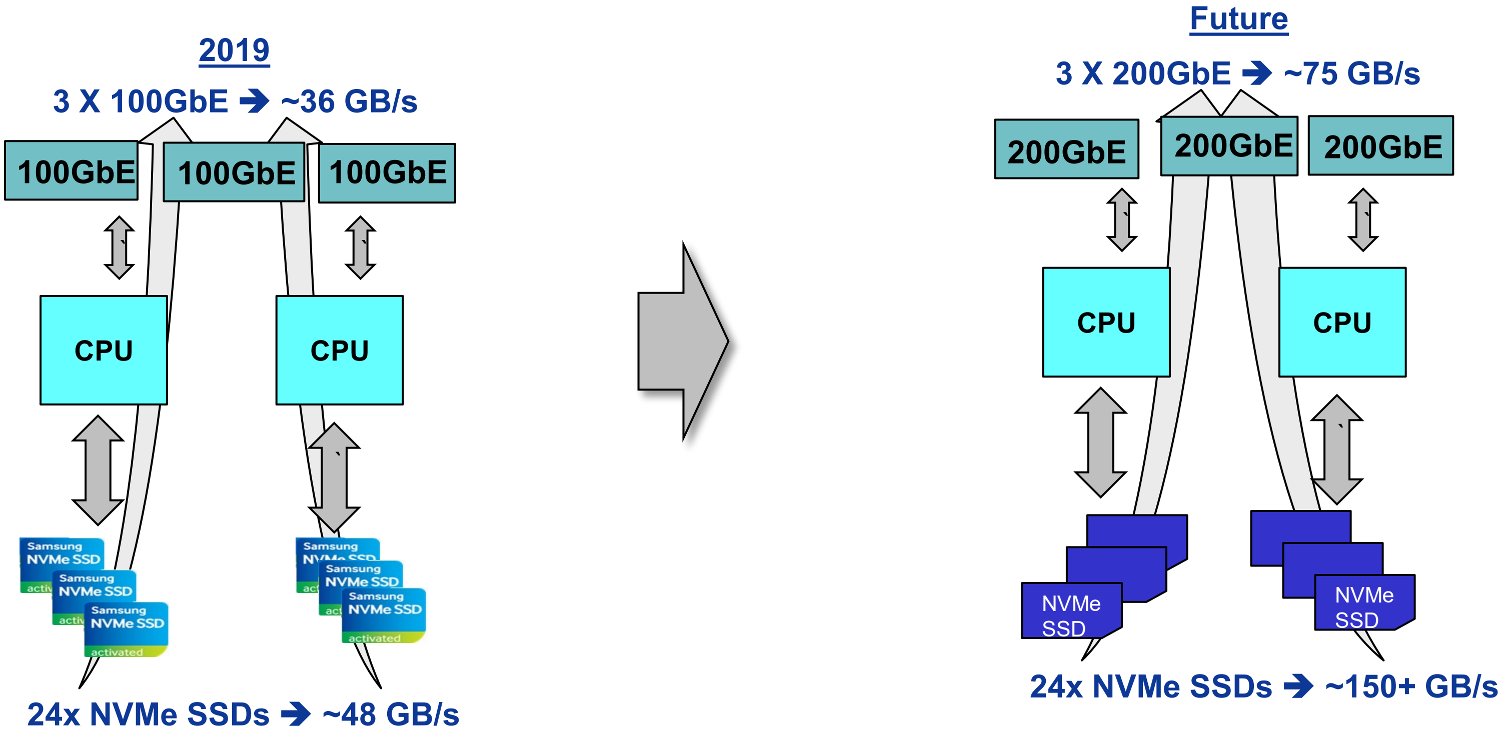
4TB, PCIe Gen3x4 External, ~530K LUTs,

Next Gen SmartSSD® CSD

Customers requirements: Integration, Interfaces, FF, workloads



Ethernet SSD targets IO bottleneck in Storage Chasses



2019

3 X 100GbE → ~36 GB/s

24x NVMe SSDs → ~48 GB/s

Future

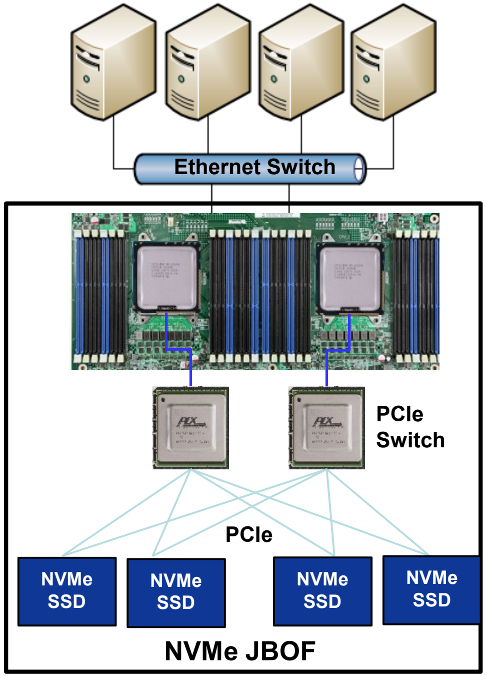
3 X 200GbE → ~75 GB/s

24x NVMe SSDs → ~150+ GB/s

CPU and IO bottleneck for storage throughput performance

NVMe-oF SSD based EBOF

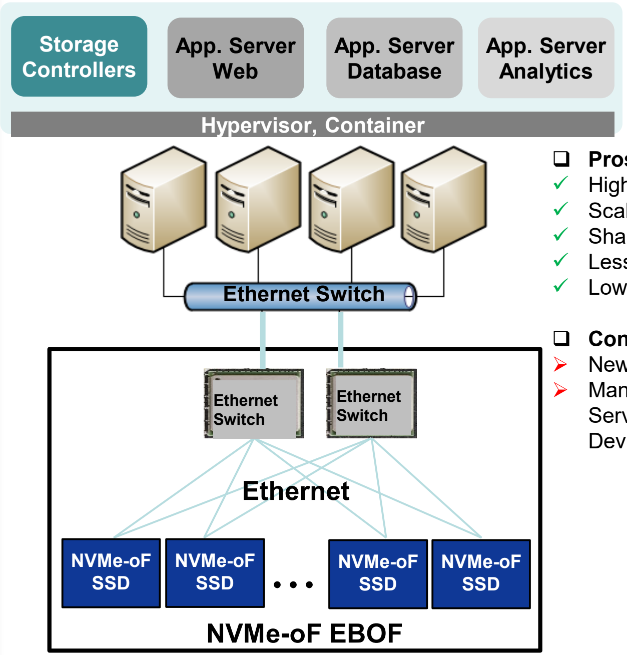
Conventional NVMe JBOF



Storage Head Nodes or Application Servers

- ❑ **Pros**
  - ✓ Enables disaggregation of NVMe SSDs
  - ✓ Management & Storage Services
  - ✓ Utilizing existing storage & server architectures
- ❑ **Cons**
  - ➢ Non-scalable Storage Controller - PCIe single root constraint
  - ➢ Bandwidth Limitation
    - CPU, PCIe, NW Constraints
  - ➢ Power and Thermals

NVMe-oF EBOF



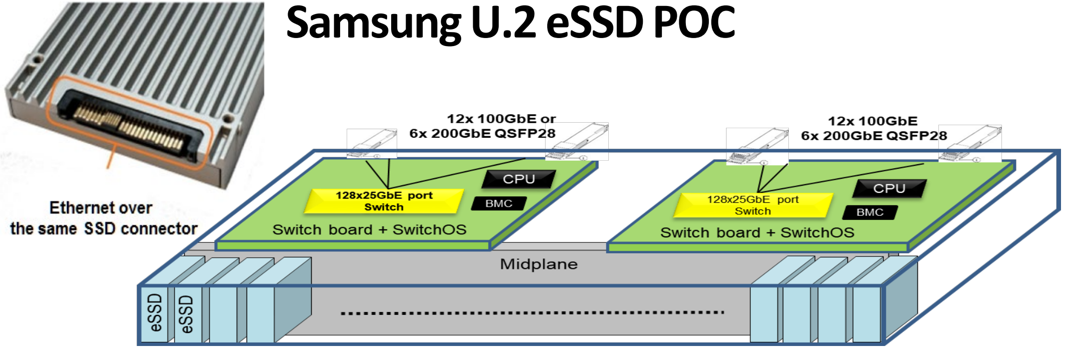
- ❑ **Pros**
  - ✓ High Bandwidth
  - ✓ Scaled Linearly (Ethernet)
  - ✓ Sharable via NVMe-oF
  - ✓ Less power
  - ✓ Lower latency
- ❑ **Cons**
  - ➢ New platform architecture
  - ➢ Management of Storage Services & Network Devices

NVMe-oF EBOF can address bandwidth, scalability, and flexibility

Samsung Ethernet SSD Technology Roadmap

- Samples, development tools, partners solutions available for immediate PoC
- Customer PoC Test&Dev systems/support available from Samsung and partners

Samsung U.2 eSSD POC

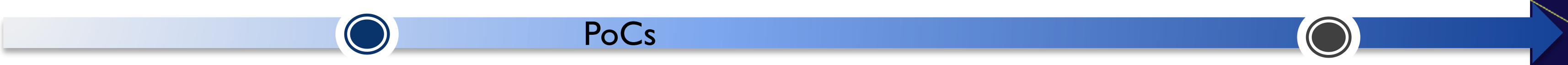


U.2, 2x25 GbE, RDMA/RoCeV2, iWARP NVMe-oF

Next Gen Ethernet SSD

Customers requirements:
NVMe-oF/TCP, HW Offloads
FF, Congestion Control (ECN/PFC),
Security, Management

2020

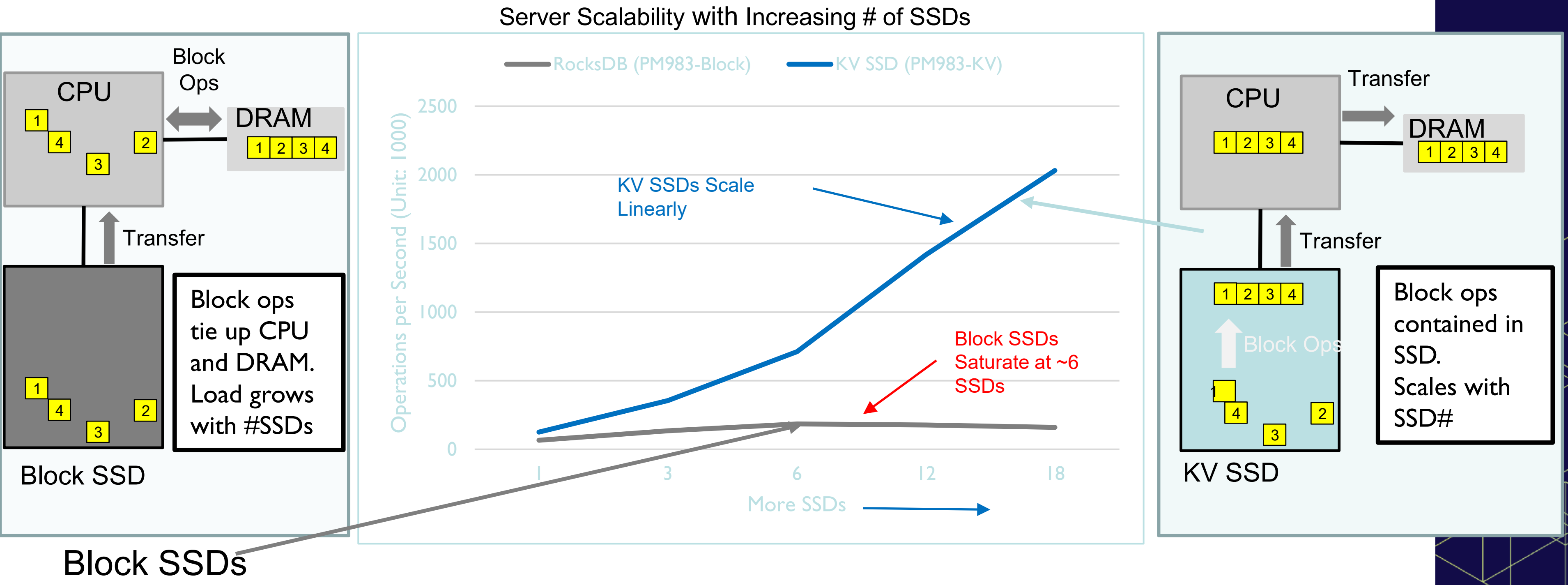


U.2 ES
POC

2nd Generation

KV SSD is about Efficient Block Operations

Block operations on CPUs ⇨ Bottlenecks, Scaling Inefficiency.
KV SSD offloads Block operations from CPUs.



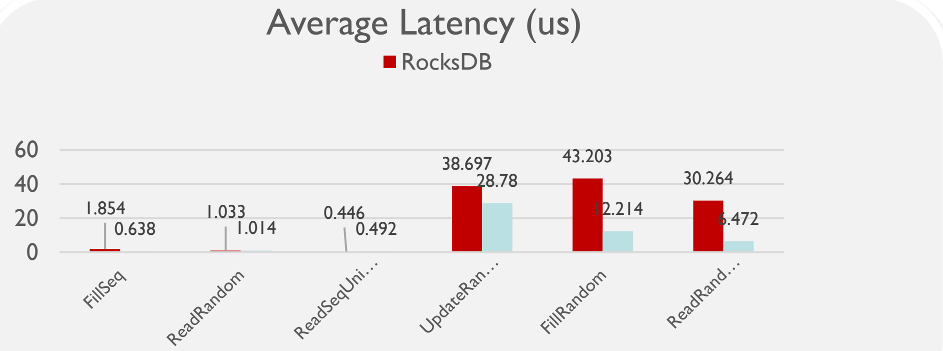
Server Scalability with Increasing # of SSDs

Block ops tie up CPU and DRAM. Load grows with #SSDs

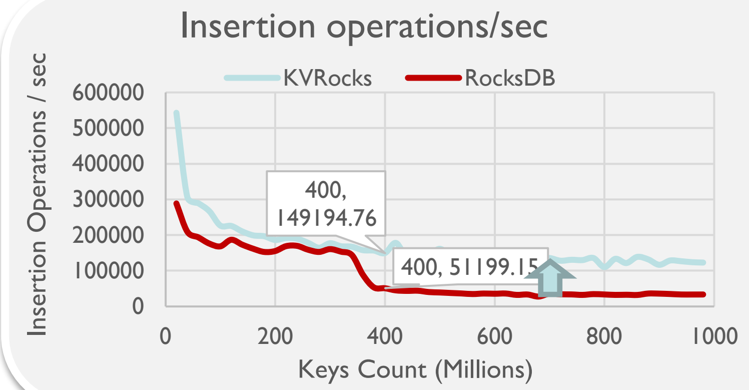
Block ops contained in SSD. Scales with SSD#

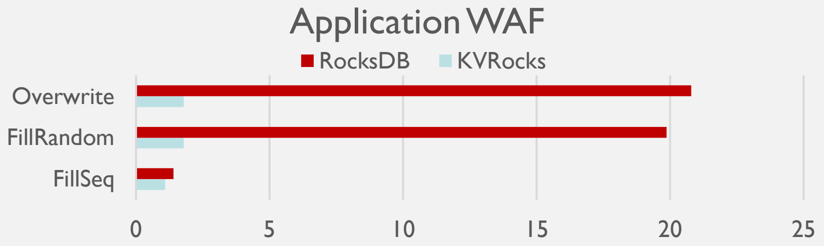
KV-optimized SW shows Multiple Efficiency Benefits

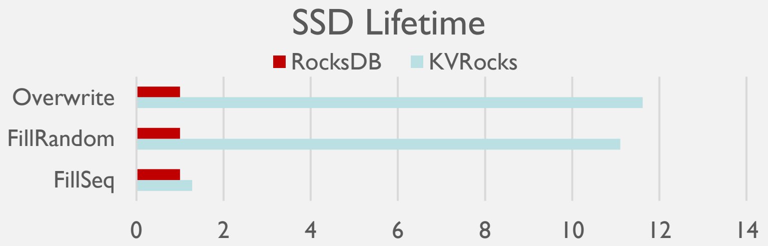
3x



Average Latency is up to 60% better on KVRocks





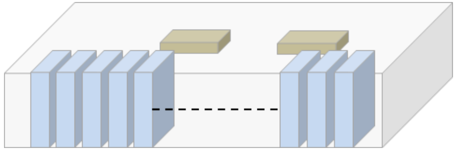


SSD lifetime is more than 11x for KVRocks compared to RocksDB ⇨ **11x less CapEx**

Samsung KV SSD Technology Roadmap

- Stack open sourced at https://github.com/OpenMPDK/KVSSD

U.2 KV SSD



U.2 FF: Scale Processing to 24 devices

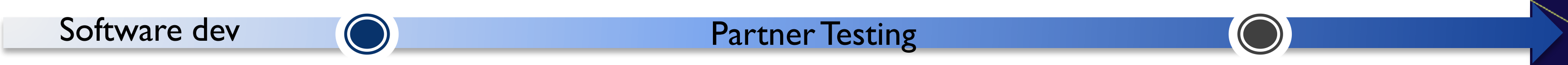
4TB, PCIe Gen3x4 External

Next Gen KV SSD

Your requirements?

(Integration, Interfaces, FF, Workloads)

1H'20



Rocks DB repl.
KVCeph
Minio

Partner Solution

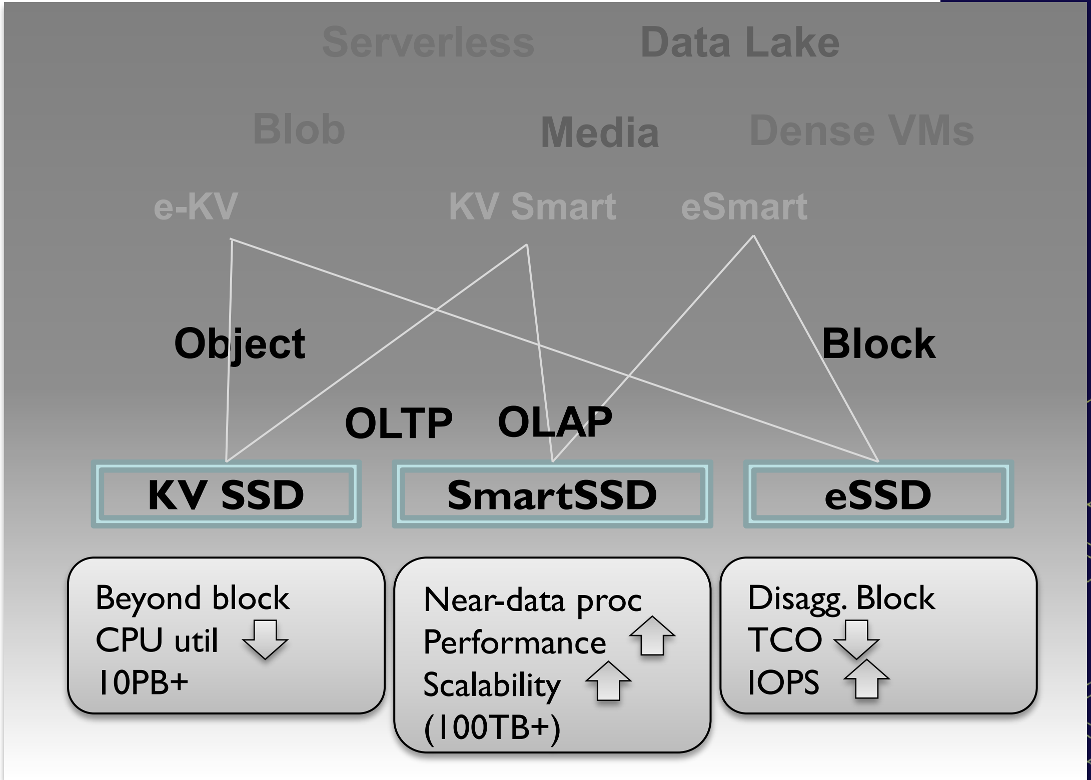
2nd Generation

Benefits Summary

	SmartSSD	Ethernet SSD	Key-Value SSD	Zoned Name Spaces
Application Awareness	✓		✓	✓
Acceleration	✓			
Reduce data-related CPU load	✓		✓	
Improved Write Endurance			✓	✓
Fewer protocol terminations	✓	✓		
Min device virtualization o/h				✓
Fewer stack translations		✓	✓	
Metadata Optimization			✓	
Scaling Data Bandwidth	✓			
Saving L2-to-Memory BW	✓		✓	
Control@Scale (IODT, QoS)				✓
Maximize #SSDs/chassis	✓	✓	✓	✓

Possible Convergence

Host Interface	Addressing	Accelerator
PCIe	Block	None
Ethernet	ZNS	FPGA
	Key-Value	



Serverless

Data Lake

Blob

Media

Dense VMs

e-KV

KV Smart

eSmart

Object

Block

OLTP

OLAP

KV SSD

SmartSSD

eSSD

Beyond block
CPU util ⇩
10PB+

Near-data proc ⇧
Performance ⇧
Scalability
(100TB+)

Disagg. Block
TCO ⇩
IOPS ⇧

Please take a moment to rate this session.

Your feedback matters to us.

- Bullet one
  - Bullet two
    - Bullet 3
      - Bullet 4

Dr. Pankaj Mehra

Vice President,
Storage Pathfinding

Samsung Semiconductor Inc.

- Subhead
    - Example 1
    - Example 2
- Subhead