Computational Storage by FPGA
ODSA Data Accelerator Workshop

David McIntyre
Director Product Planning and Business Enablement
Samsung Device Solutions America
August 27, 2021

The Infinite Data Ocean

Our Mission

- **Oceans of data, but how to search effectively?**

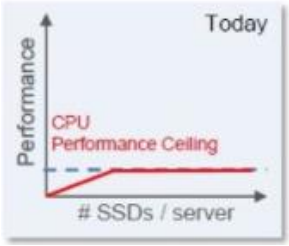
- **Data Search and Analytics: No Compromise**
  - Urgent
  - Accurate
  - Sustainable
  - Results- driven

<u>Scenario Planning</u>

- Medical (Genomics)
- Fintech (Colo Trading)
- Cyberforensics (Security)
- Edge (AI Surveillance)

Today's Compute and Storage Compromise

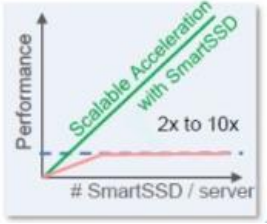
- **Latency**: Data transfer to cloud or on-prem host servers for all processing tasks
- **Restricted Bandwidth:** Network congestion to over-provisioned servers
- **Expense:** Cloud instances
- **Security and Privacy Issues**: Data exposure vulnerabilities and sharing across parties
- **Resource dependencies:** Misbalanced provisioning of compute and storage resources (see pic)



Performance Limit:
Storage to Host Server Ratio

Computational Storage Resolution

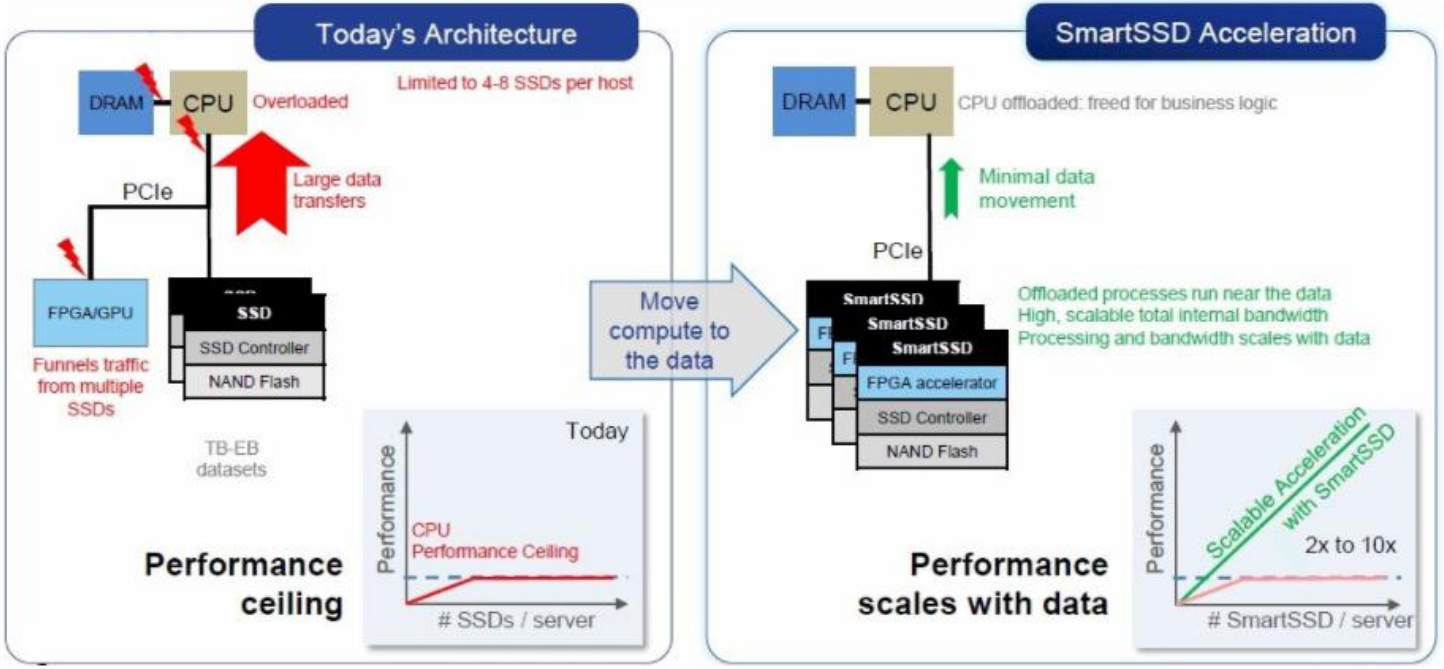
- **Reduced Latency**: Bringing compute to the data for improved response times with data analytics
- **CapEx and OpEx savings**: Less data center server resources required for reduced TCO
- **Security and privacy compliance**: Localized data processing provides better protection and control
- **Flexibility and scalability**: SmartSSD scale up based on customer application requirements



Performance Enabled: SmartSSD local compute

Computational Storage Explained

- Improved Performance with local on-drive compute resources
- Minimal data movement required
- Secure data attributes



Computational Storage Deployments

Move Compute Closer to Storage

Current Compute/Storage Architecture

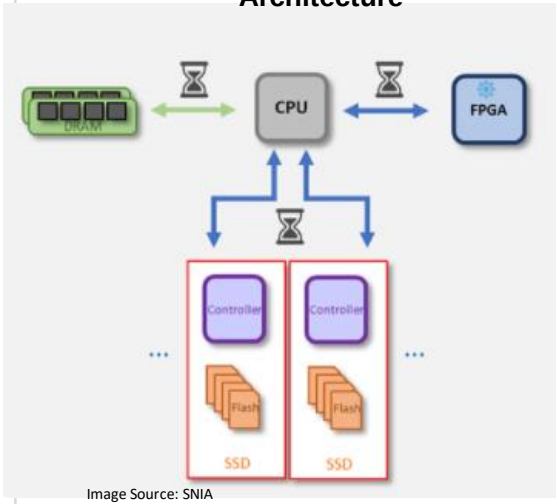
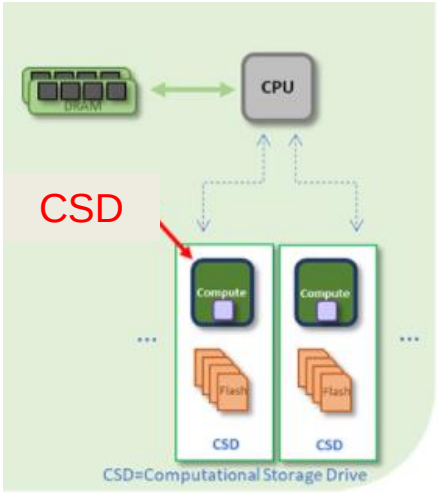


Image Source: SNIA

- Moving data between storage and host CPU creates performance bottlenecks for data-intensive applications

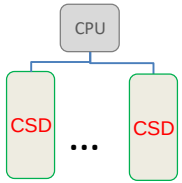
Computational Storage Architecture



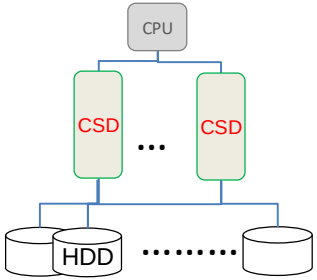
- Data processed directly on the CSD => no large data transfers, faster time-to-insight
- Adding CSDs adds processing power and internal bandwidth => scalable acceleration

Deployment Examples

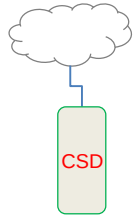
- Compute/Storage Server



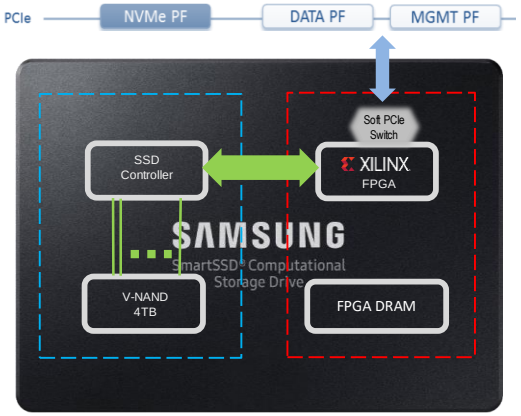
- Smart Cache Layer



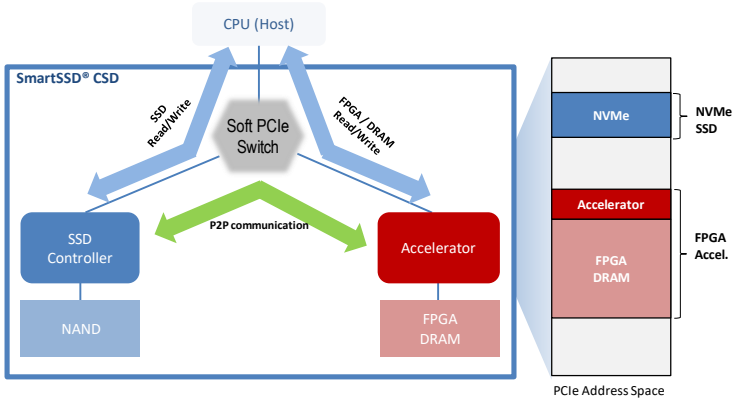
- Cloud to Edge compute



Computational Storage Product Example



Samsung SmartSSD® CSD



FPGA Accelerator, Flash Controller, 4GB DRAM and 4TB TLC NAND flash

- Peer-to-peer (P2P) communication enables unlimited concurrency

SSD-to-Accelerator data transfers use internal data path

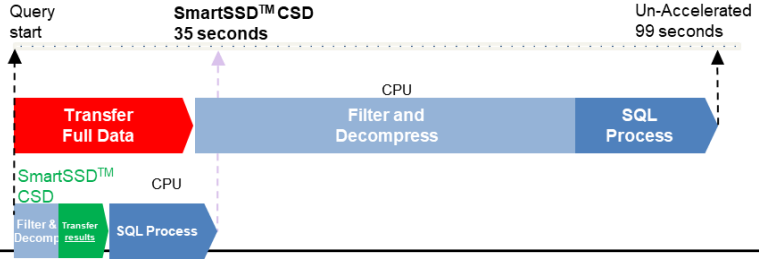
- Save precious L2:DRAM Bandwidth (Compute Nodes) / Scale without costly x86 frontend (Storage Nodes)
- Avoid the unnecessary funneling and data movement of standalone accelerators
- FPGA DRAM is exposed to Host PCIe address space
- NVMe commands can securely stream data from SSD to FPGA peer-to-peer

SmartSSD® Drive Performance and Use cases

◆ Scalable Across Many Applications: Data at Rest and Inline

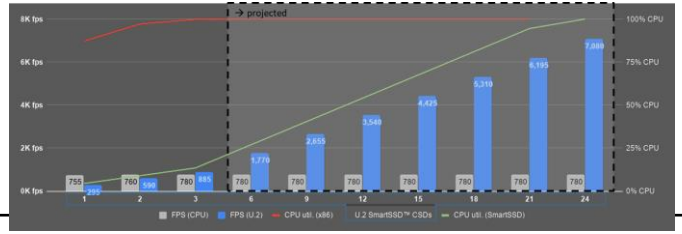
Database

SparkSQL with Parquet Data
5.3x overall performance improvement for heavy query
PostgreSQL 11
40x faster scan-heavy queries (7 -> 331 queries/hr)



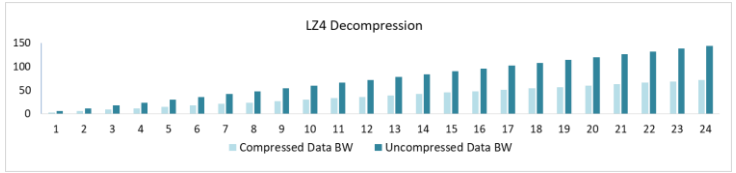
Rich Media

H.264 Video Transcoding
Multi-stream transcoding: 20% higher 1080p frame rate
Offloading CPU workload : 87% lower CPU usage

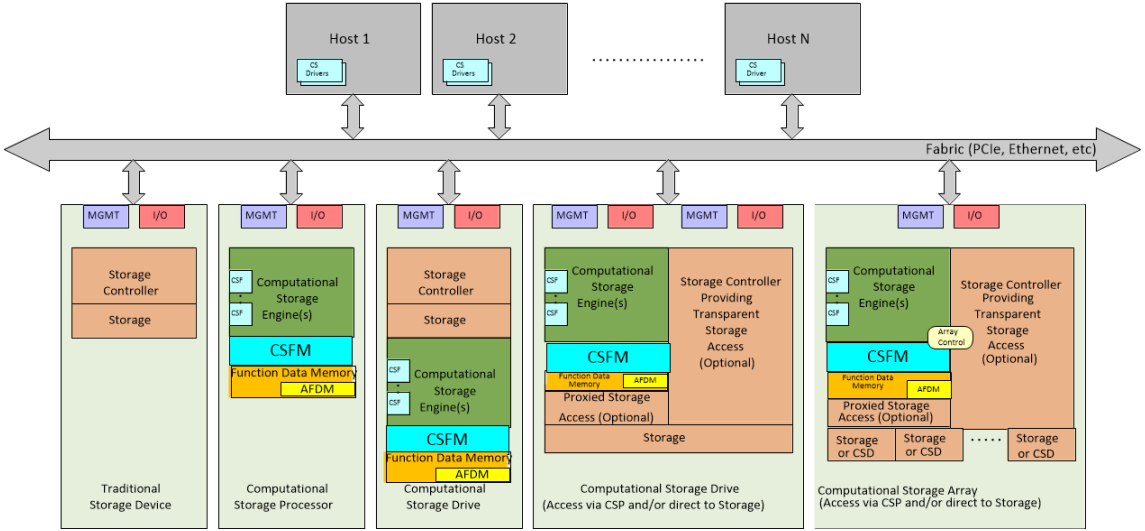


Storage & Big Data Services

LZ4 Decompression Scale-out
3x decompression bandwidth, scales to 24 SmartSSD drives
MPU search Scale-out
Same regex search time for 10PB as 4TB



Computational Storage and Security



The CSD may perform security functions:

- **Authentication**. Host agent to CSD and CSD to host agent
- **Authorization**. Mechanisms for secure data access and permissions control
- **Encryption**. Mechanisms to perform computation on encrypted data that was not encrypted by the CSD. Mechanisms that exchange information necessary for the CSD to encrypt/decrypt data.
- **Auditing**. Mechanisms that allow for generating and retrieving of a secure log

Risks vs standard storage:

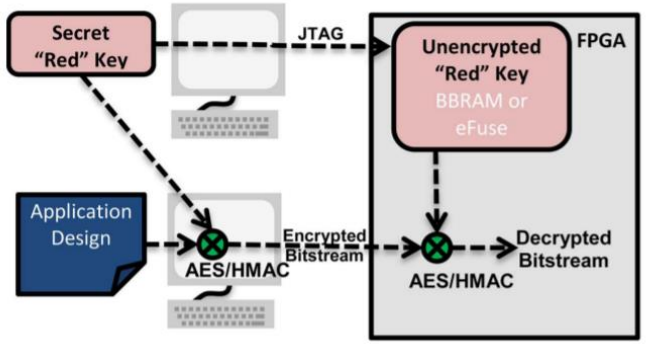
- **The CSD can delete/add/modify data on the drive**
- **The CSD functionality can be programmed**
- **Virtualization**

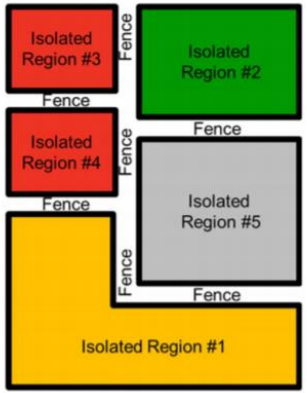
Risks vs external accelerator:

- **Direct access to storage**
- **FPGA programming**
- **Access to network infrastructure (NVMe-oF)**
- **Decryption of data prior to processing**

Computational Storage and FPGA security

- **FPGAs are SRAM based devices which are programmed by secure bit streams**
  - Key is programmed via JTAG port
  - Bitstream in encrypted with design tools
  - FPGA identifies encrypt/no encrypt for field testing
- **AES 256 secures bitstream programs**
- **Additional Security Measures**
  - Design Region Isolation
  - JIT Partial Reconfiguration
  - SOC and Bus Isolation
  - PUF files for device dependency
  - E-fusing





https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6849432

Computational Storage Summary

Mission Attributes		Solution Requirements
- Urgent	→	Low Latency
- Accurate	→	Data Analytics Solutions
- Sustainable	→	Product Leadership
- Results	→	Performance Value
- Security	→	Data Protection
- Leadership	→	Innovation and shared customer vision

Cloud to Edge Deployments

Source: Gartner
ID: 450640_C

Computational Storage: Your Invitation

- **Collaborate on workloads and use cases**
  - Data Analytics
  - Data Management
  - AI Inference
- **Solution development**
  - POC trials
  - Ecosystem partners
  - TCO comparisons to alternate technologies
- Reach out to learn more:

David McIntyre
d.mcintyre@samsung.com