NVRAMOS 2019

Towards Even Lower Total Cost of Ownership of Data Center IT Infrastructure

- a Solid State Storage Perspective

Joo-Young Hwang

Samsung Electronics

Oct, 25th, 2019

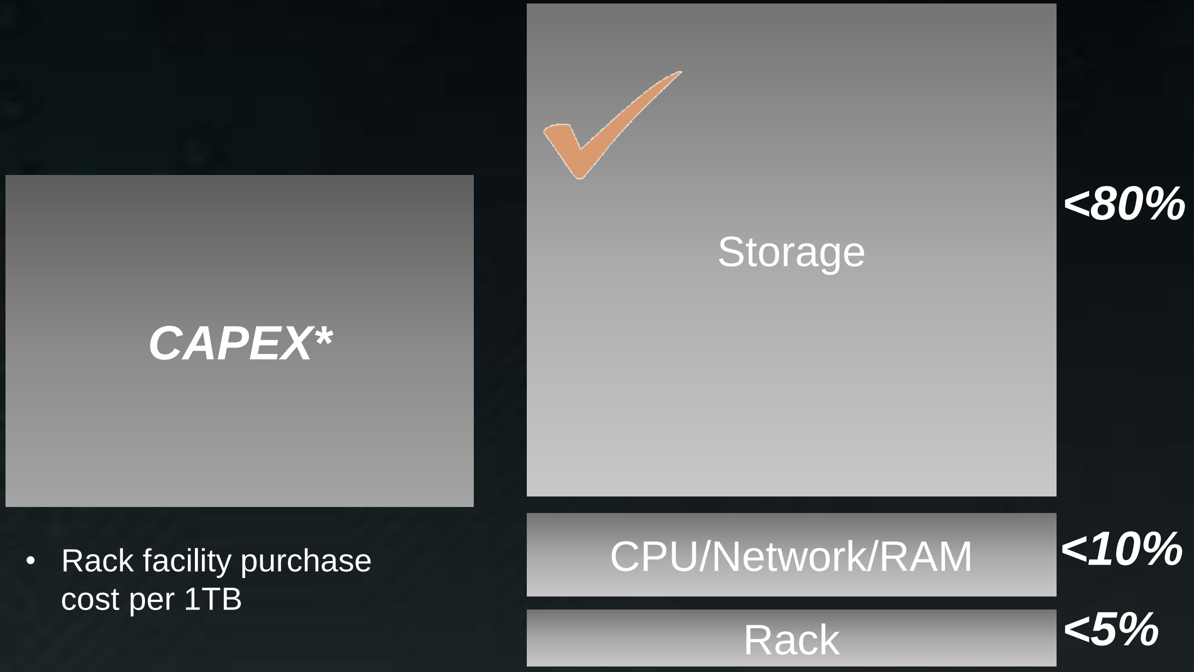
What is the TCO? (Total Cost of Ownership)

- ***TCO* = 1) CAPEX**
  2) OPEX
  3) Quality and Reliability Cost

<40%	***CAPEX***	***Rack purchase cost***
<10%	***OPEX***	***Operational Cost***
<50%	***Quality and Reliability Cost***	***additional capacities*** ***(for failure, peak response, and replica)***

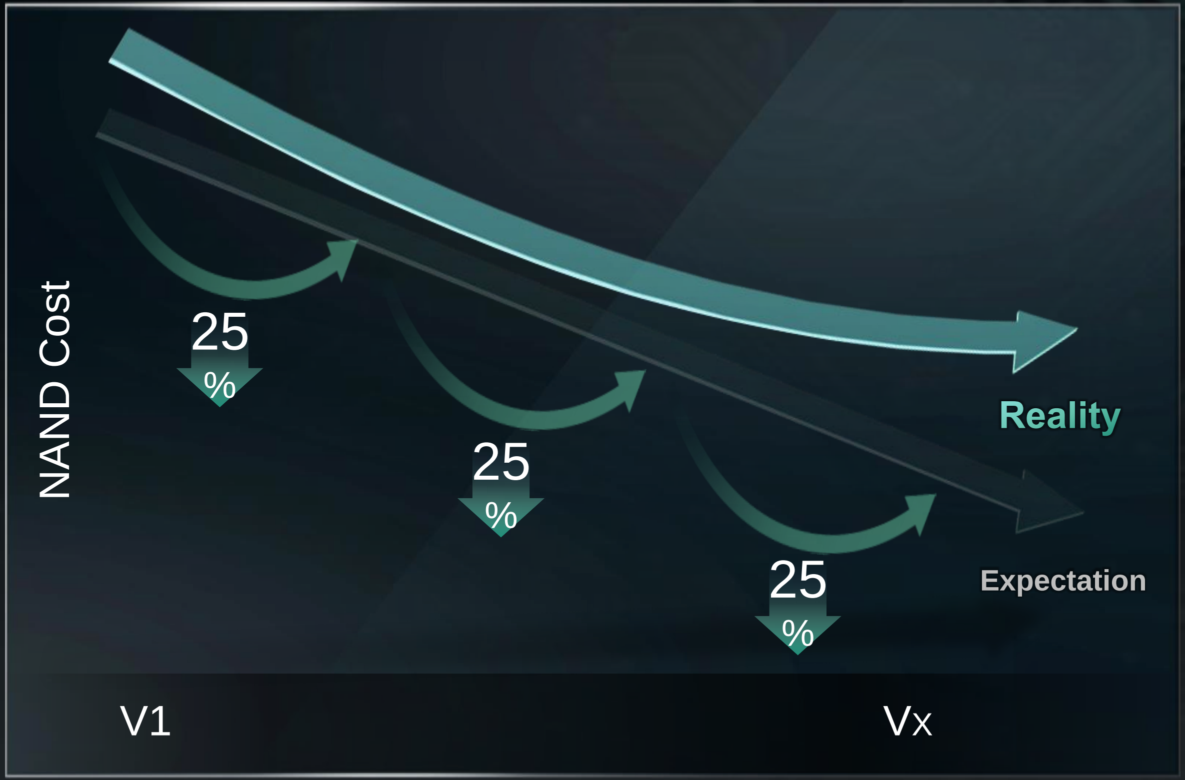
Datacenter Storage TCO

Difficulty for CAPEX Reduction



- Rack facility purchase cost per 1TB

Challenges in NAND Bit Growth



SSD with Next Gen NANDs

Performance

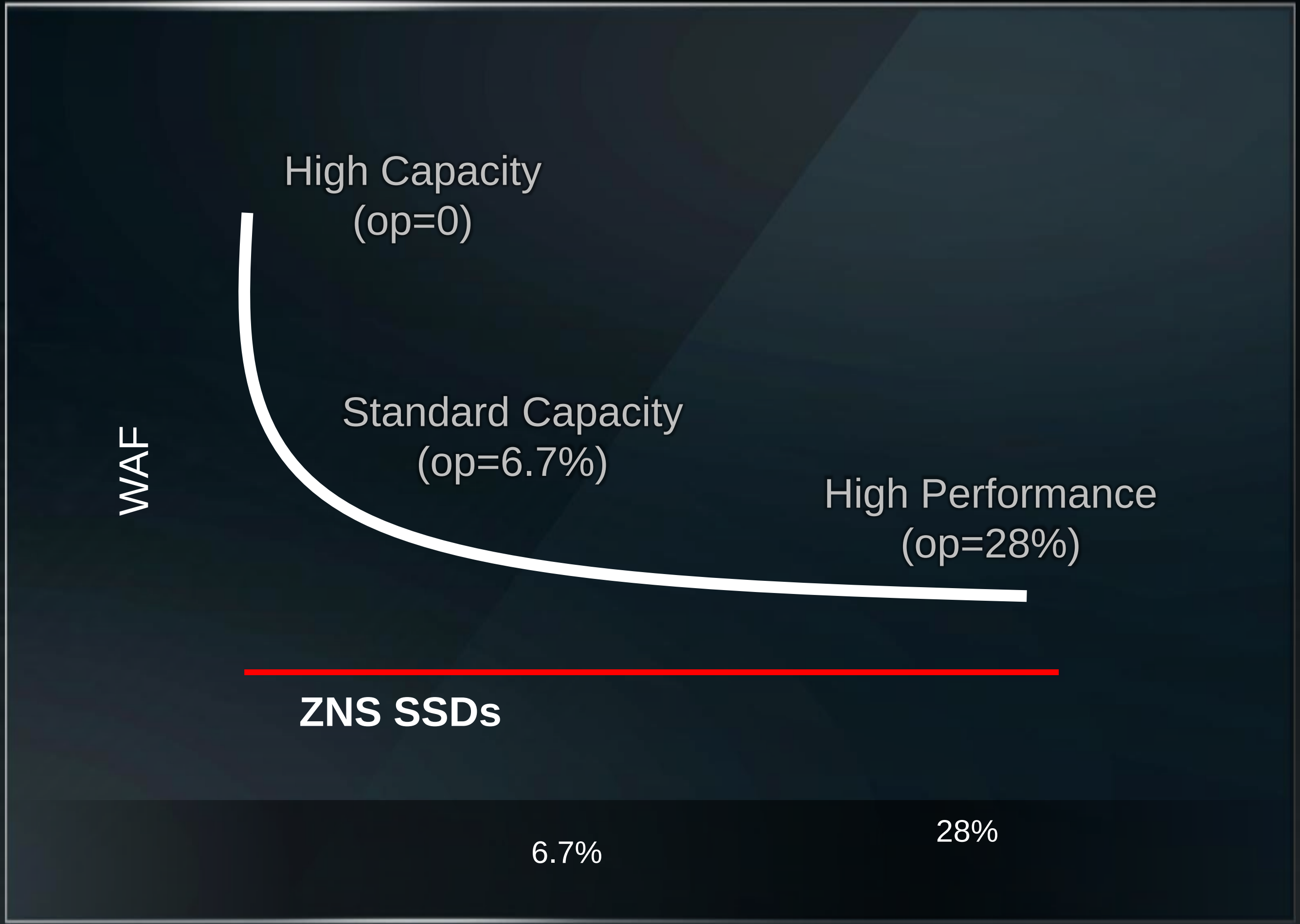
- Enhanced Program Technique
- Flexible Garbage Collection & Die Management

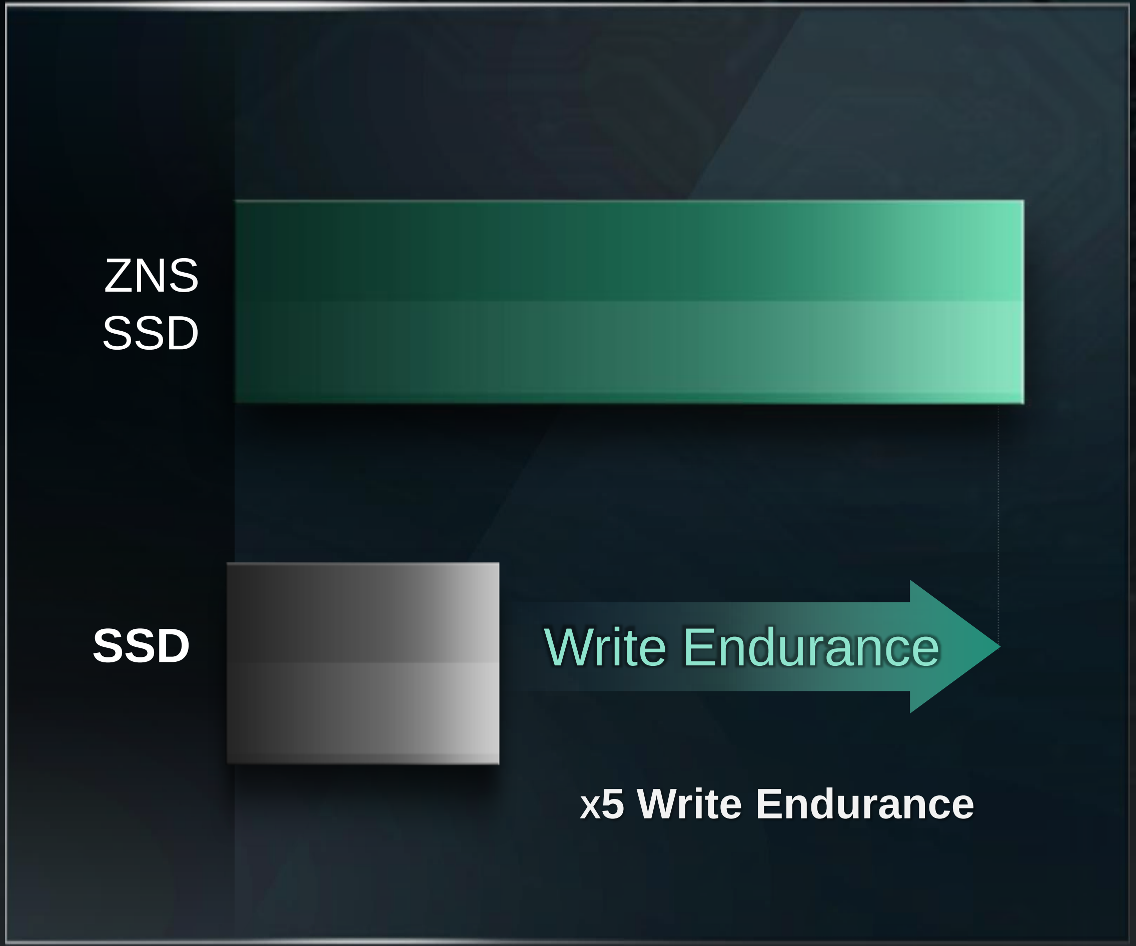
Reliability

- Enhanced Parity LDPC
- ML Algorithms adaptation

WAF Reduction

ZNS (Zoned Name Space):
an ongoing standardization effort to lower WAF and higher user space.

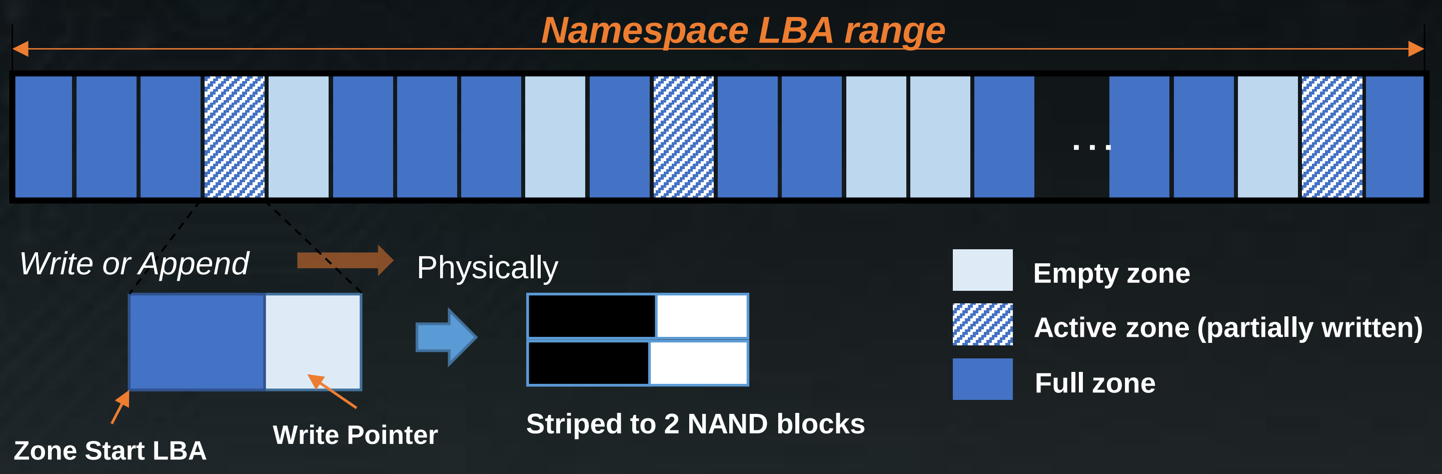


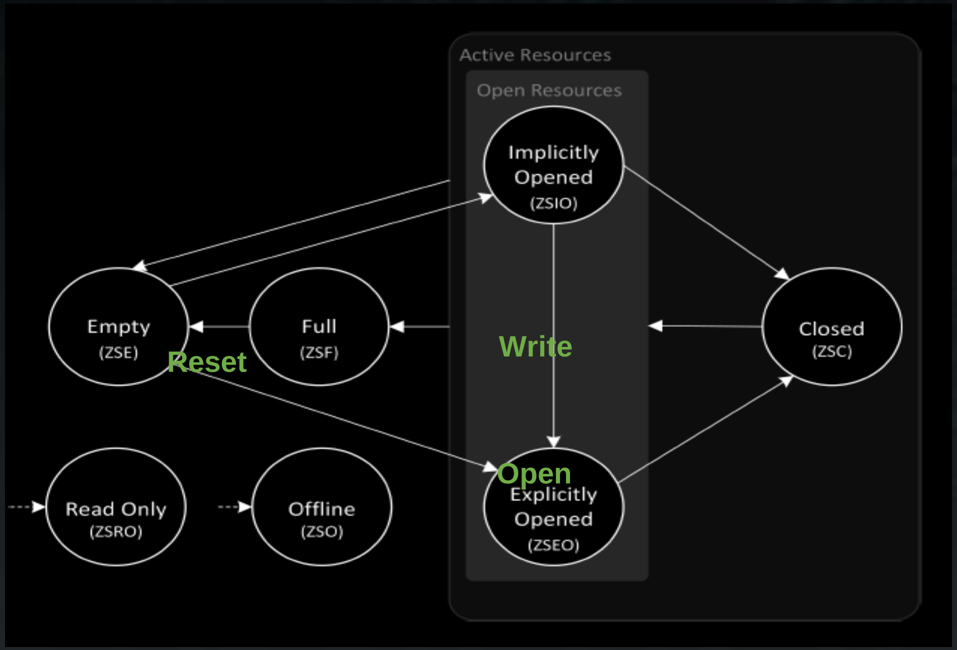


ZNS Concept

Adopting Zone concept from SMR HDDs (ZBC/ZAC)

- Storage capacity is divided into zones
- Each zone is written sequentially
- Nameless-write concept
  - Added zone append to allow out-of-order execution of multiple write operations to a zone
  - Support high queue depth writes
- Interface optimized for SSDs
  - Align with media characteristics (e.g. zone size aligned to NAND block sizes)

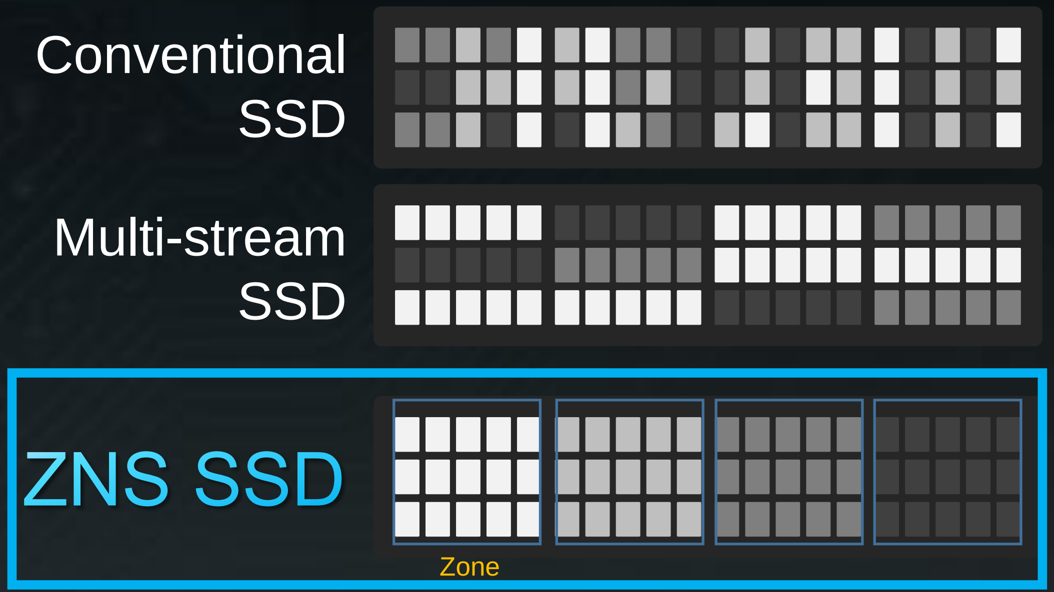




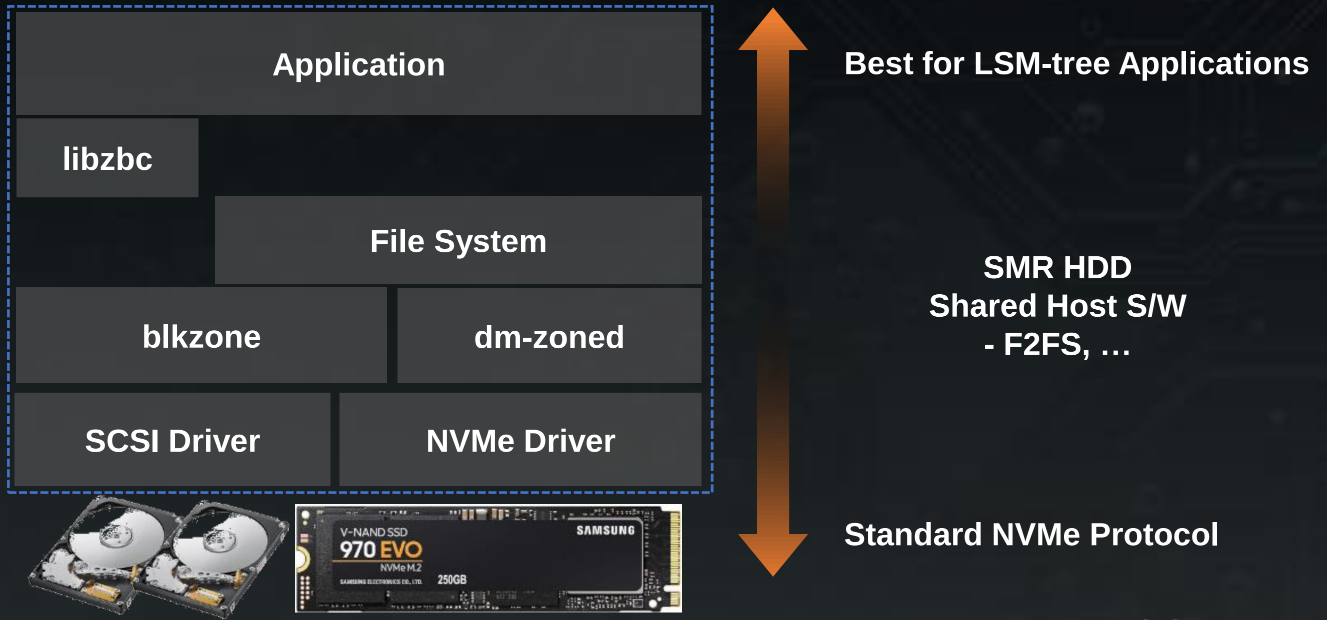
ZNS: Contracts between Host and Device

Advantage	Requirement or Restrictions
Many Stream	Page-size aligned write
Small FTL Mapping Table	Sequential write only
Better Performance w/o FTL Garbage Collection	Application-level Garbage Collection
Better Lifetime with Lower WAF	
More Capacity with Zero Over-Provisioning	

→ **Good for QLC SSDs**



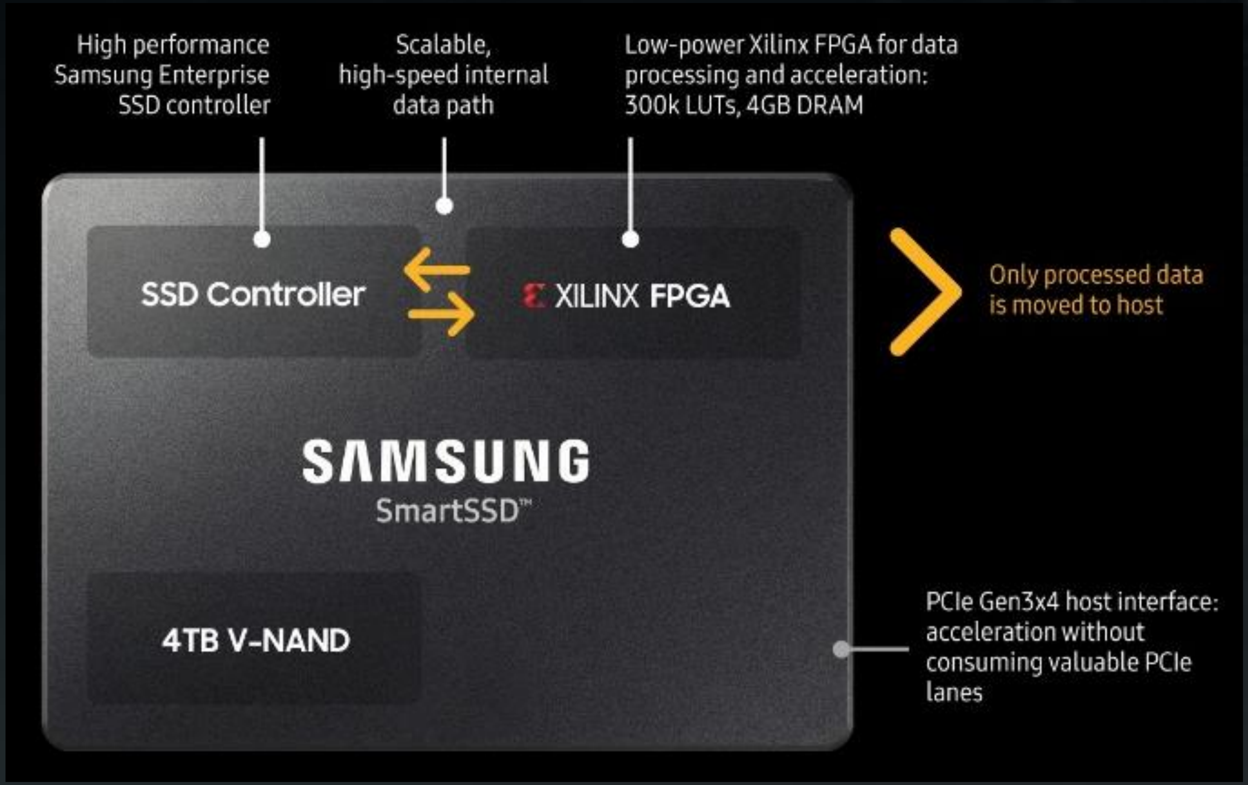
Host Layer Ready for ZNS Adoption



Offloading Computation to Storage

Reduce TCO by offloading computation to storage.

- Reduced Data Traffic between Storage and Host
- Higher Scalability & Lower TCO
  - Add more SSDs vs. Add more servers or upgrade CPUs



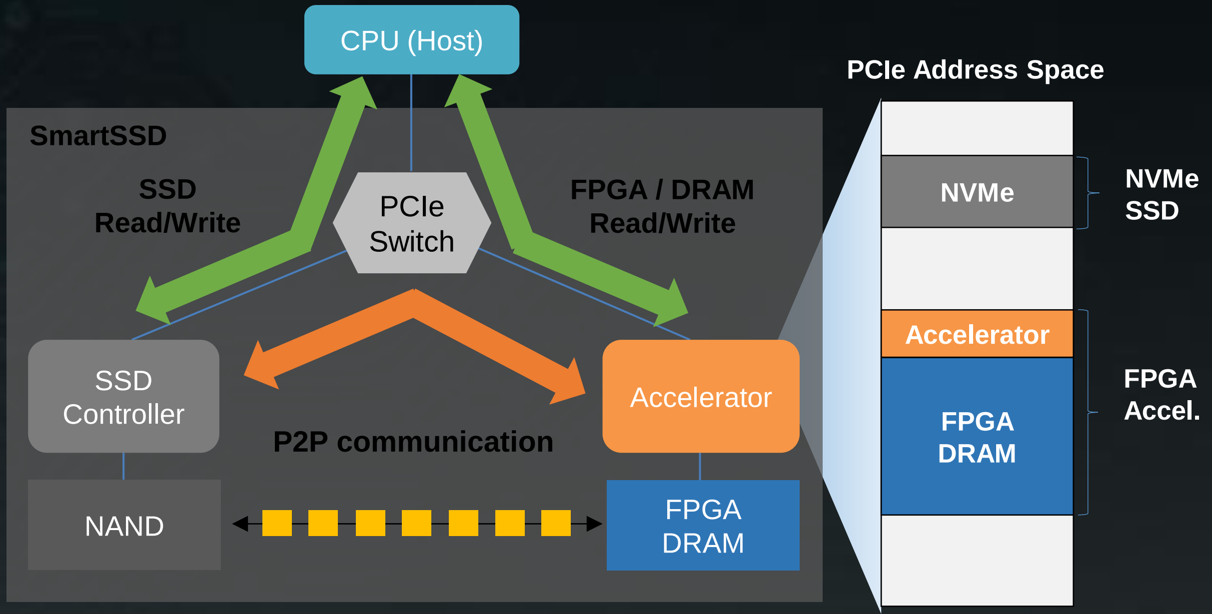
Samsung SmartSSD Architecture

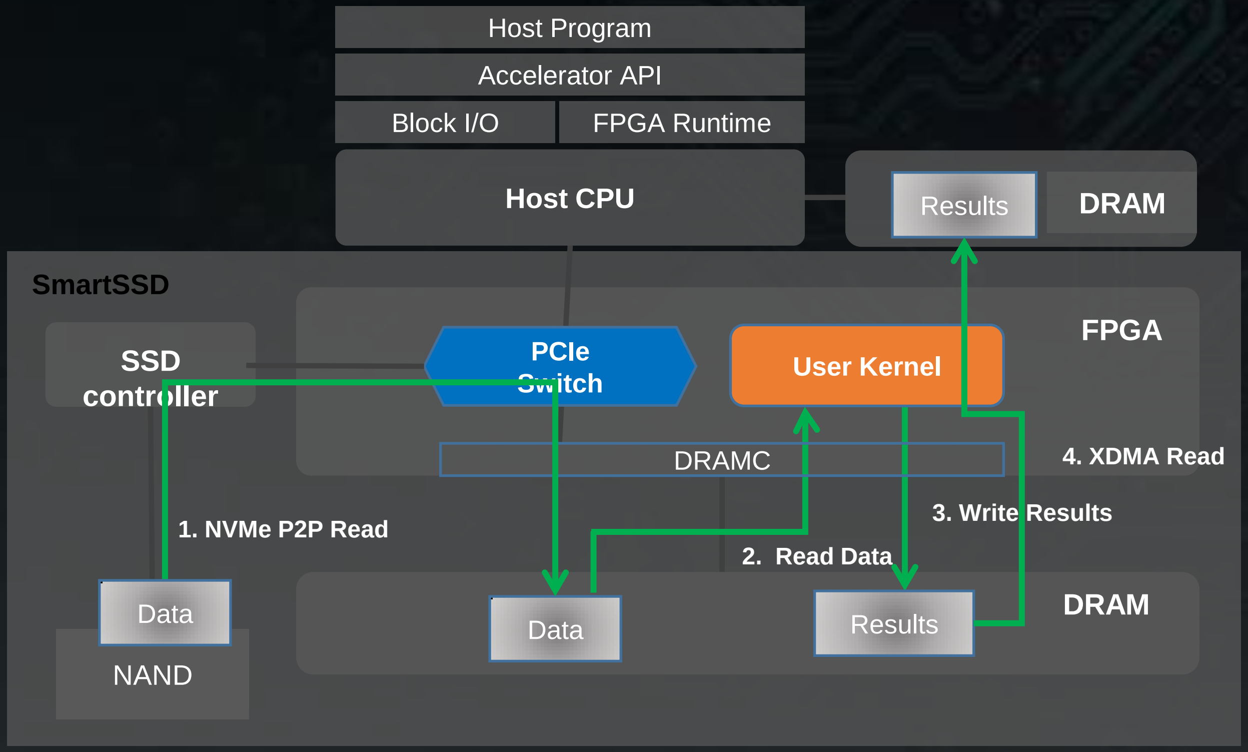
Peer-to-Peer (P2P) Communication

- P2P communication via internal PCIe switch
- FPGA DRAM is exposed to PCIe address space

Direct transfer to/from the FPGA buffers

- Bypass host DDR transfer to enable faster and efficient acceleration kernel access
- Host CPU manages the data path explicitly using Acceleration API





SmartSSD Applications

SmartSSD computing framework connects applications with SmartSSD.

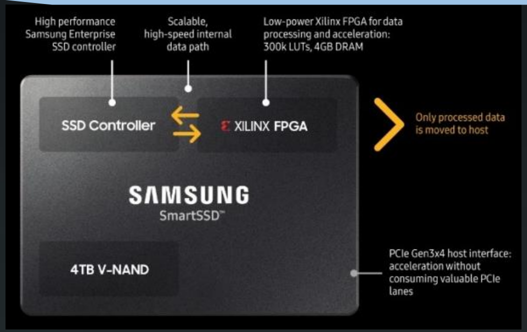
Applications

Databases	Video	AI and ML	Storage & Virtualization

SmartSSD Computing Framework

Acceleration Kernel Library

Databases	Video	AI and ML	Storage & Virtualization
DB Scan / Filter Spark acceleration Big data analytics	Video Transcoding Face/Object Recognition Real-time Processing	Classification Detection Segmentation	Compression Security Log search

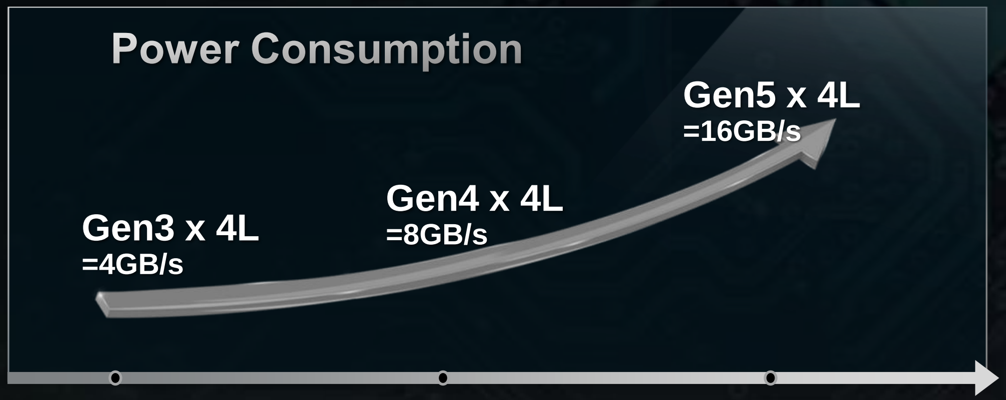


Datacenter Storage TCO

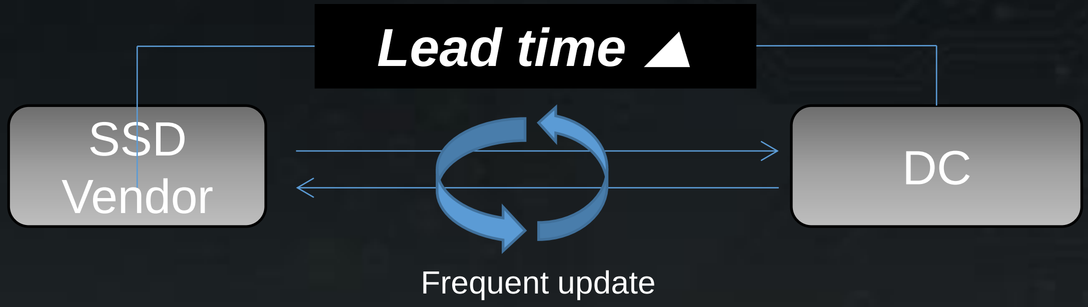
OPEX Reduction

OPEX

Electricity & Cooling Fee	*50~60%*
✓ Management	*10~20%*
Space	*10~20%*

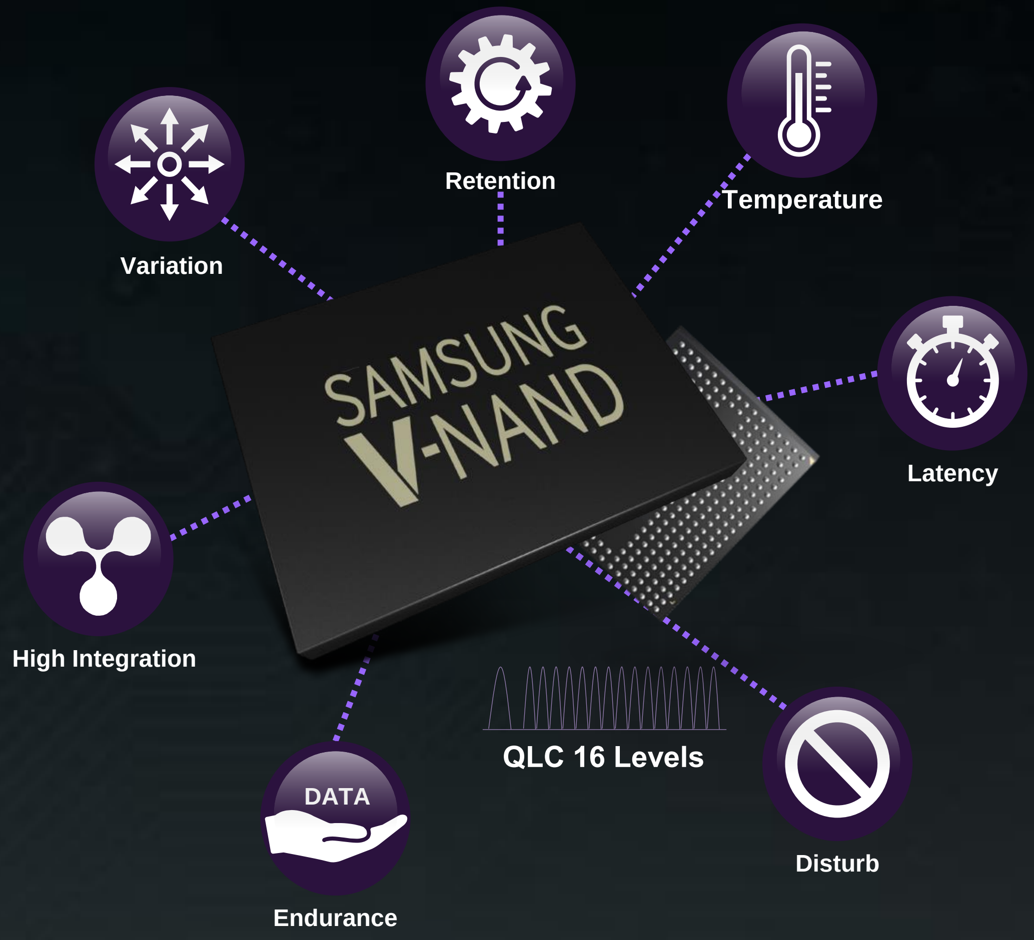


Failure Analysis Cost



Machine Learning for V-NAND

ML Technology to address Challenges of Next-Gen NANDs



ML based
Flash Management

Self-Diagnosing SSD

- Fast Recovery for Consistent Latency
- Lightweight V-NAND Monitoring Algorithm
- Error Protection for High Reliability

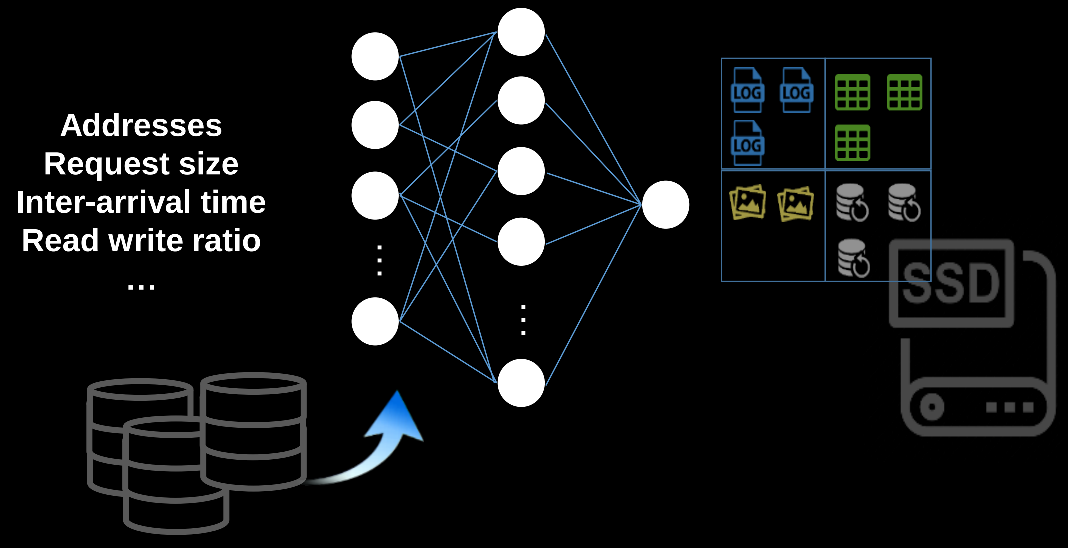
Machine Learning for Storage S/W

Storage management S/W can be even more enhanced by machine intelligence.

Human Intelligence

vs.

Machine Intelligence



Machine Learning for Storage S/W

Reducing GC Overhead based on Workload Prediction (HotStorage '19)

We solved stream classification with a *deep learning technique.*

1. **Analyze workloads and extract features**
2. **Capture feature information at run-time**
3. **Train each chunk's temperature model**
4. **Make a data placement decision based on the model**

Fig. 4. Correlation coefficient in RocksDB

Analyze Correlations

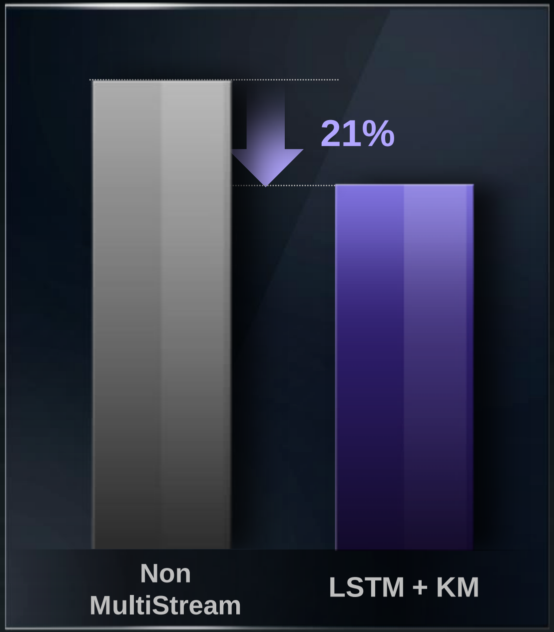
Capture I/O information in I/O Path

Train a LSTM Model

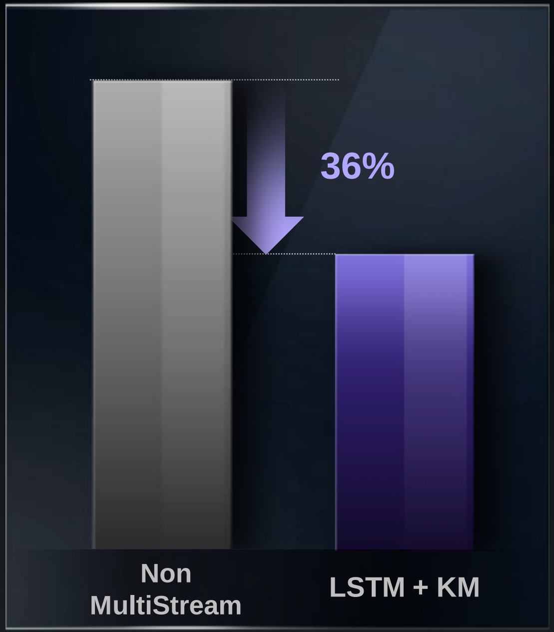
Predict temperature & use K-means clustering for aggregating chunks

Machine Learning for Storage S/W

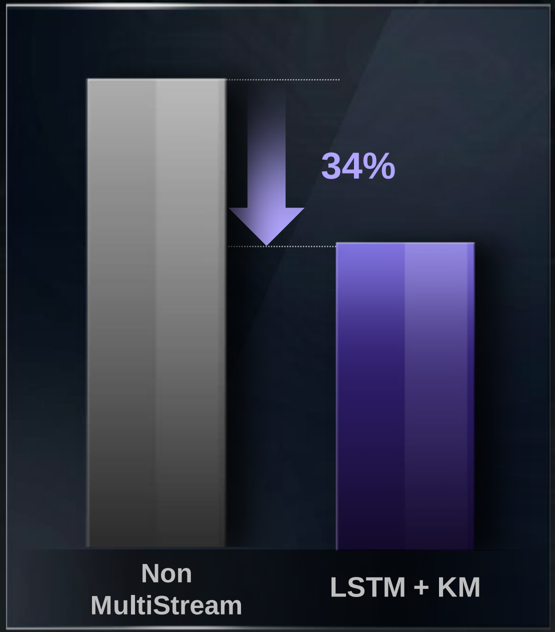
Reducing GC Overhead based on Workload Prediction (HotStorage '19)



WAF
(MySQL)



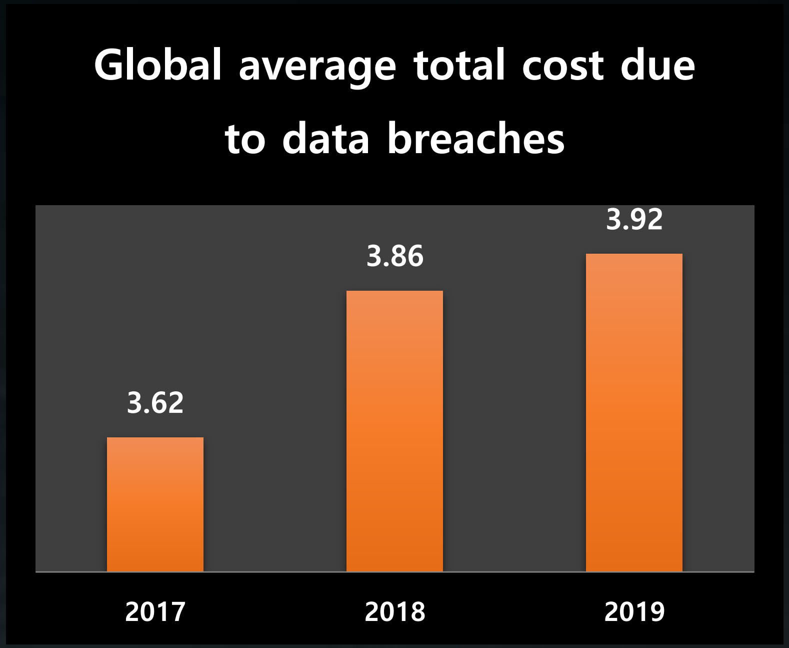
WAF
(RocksDB)



WAF
(Mixed RocksDBs)

Enhanced Security Function Needs

- Average cost of data leakage (2019) : **$3.92M**



According to EU GDPR, which came into effect in 2018, companies are fined the bigger of €20 million and 4% of total sales
if they fail to fulfill their customer data protection obligations.

Reference: "Cost of a Data Breach Report 2019" by Ponemon Institute LLC and IBM Security

The Impact of GDPR on Storage Systems

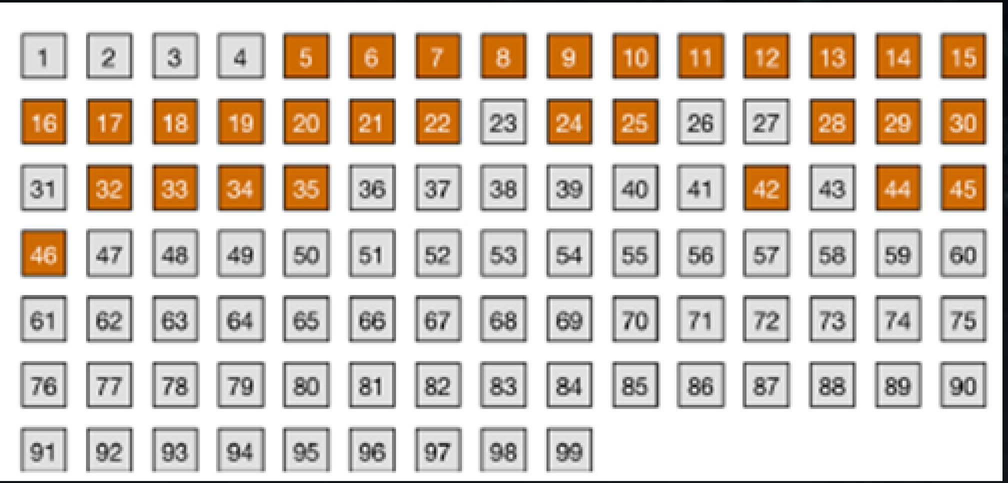
GDPR's goal of

data protection by design and by default

conflicts with the traditional system design goals of

performance, cost, and reliability

31 of the 99 GDPR articles directly pertain to storage systems



Key GDPR Articles concerning Storage Systems

- Rights of Data Subjects
- Responsibilities of Data Controllers

Reference: Analyzing the Impact of GDPR on Storage Systems, ATC '19

The Impact of GDPR on Storage Systems

- **What effort is needed to make a modern storage system, GDPR-compliant?**
- **What is the resulting performance impact?**
- **Is it possible to achieve strict compliance in an efficient manner?**

Features of GDPR-Compliant Storage

Timely Deletion

Metadata indexing

Encryption

Manage data Location

Access Control

Monitoring & Logging

Reference: Analyzing the Impact of GDPR on Storage Systems, ATC ‘19

Datacenter Storage TCO

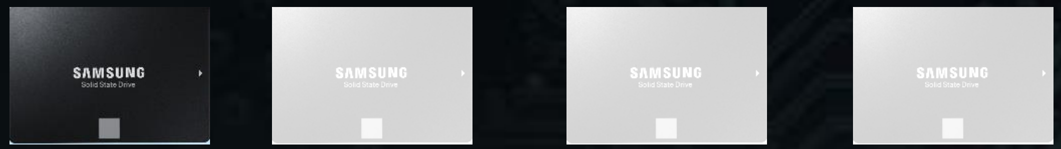
Quality and Reliability Cost Reduction

Quality and Reliability Cost

Failure Recovery & Replica

Peak Service

Additional copy for failure and Re-build time reduction



A B C

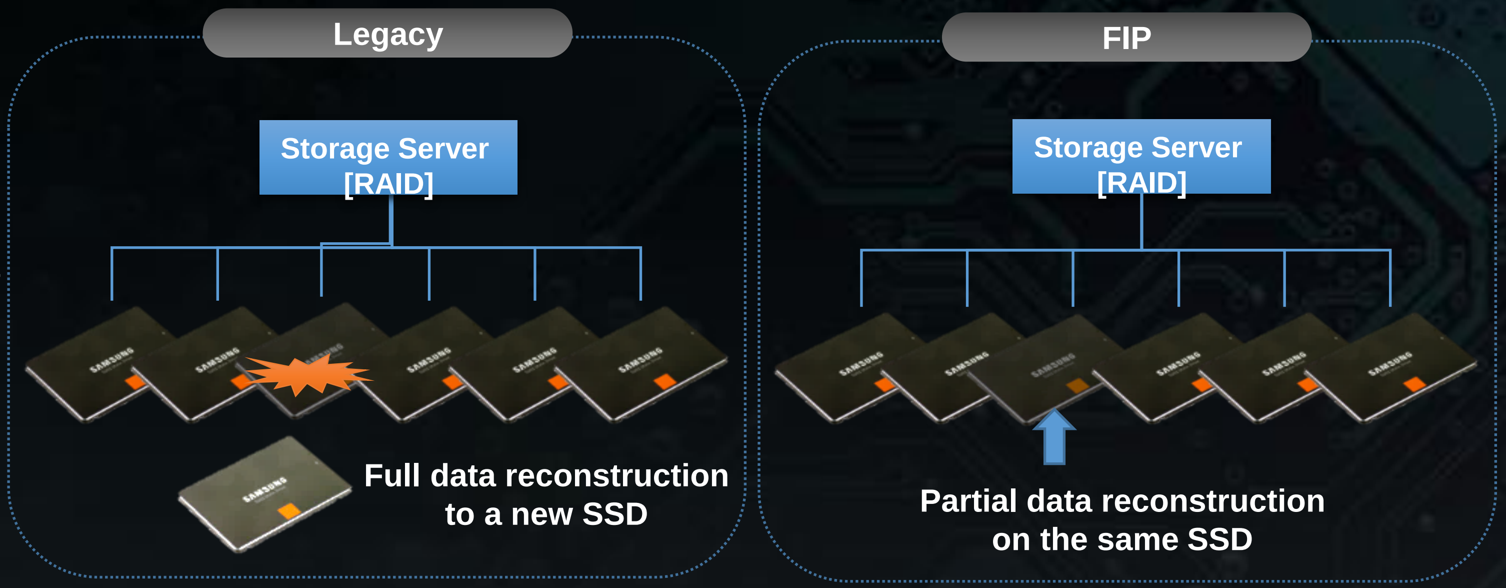
Over-provisioned servers to provide QoS at peak time.

Highly Reliable SSD

Fail in place (FIP)

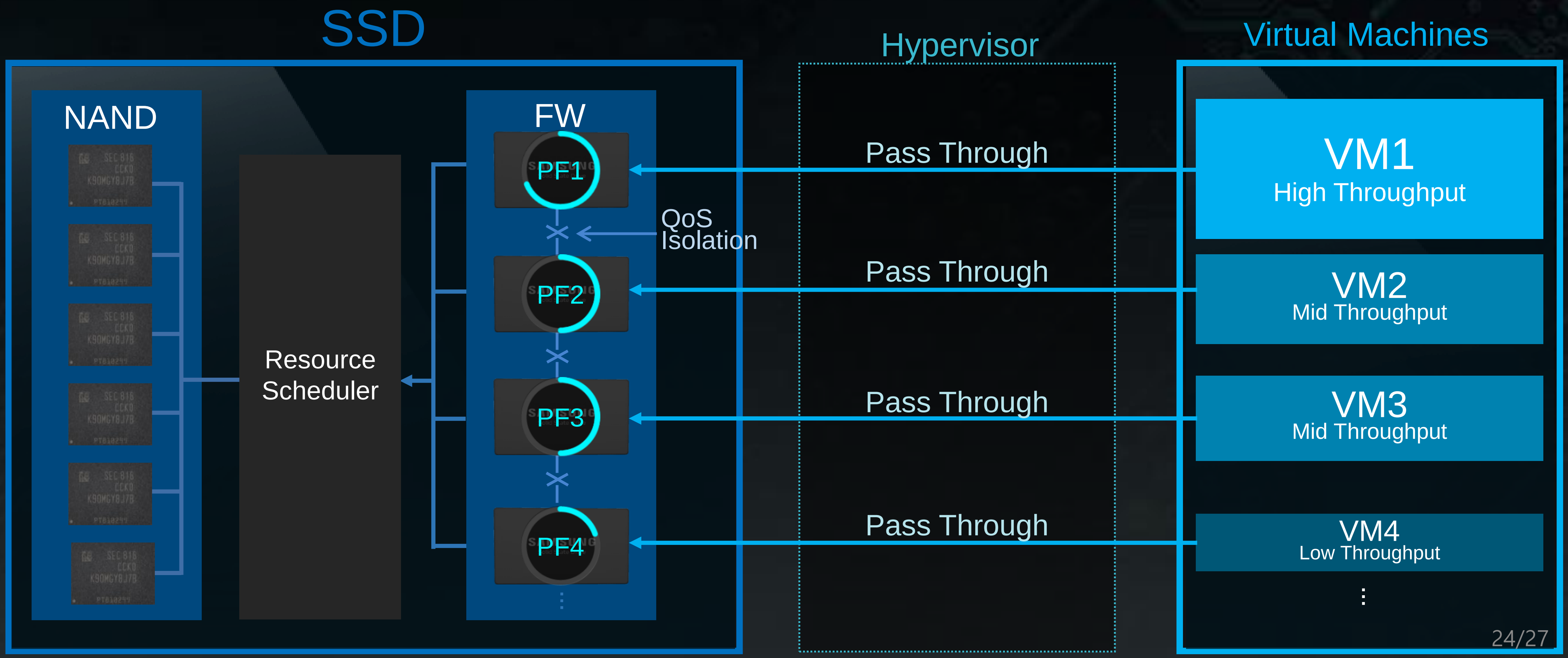
- Robust against chip failure
- Cost reduction by avoiding disk replacement
- Fast recovery by communication between

  Host system and SSD



Multi-Tenancy Support

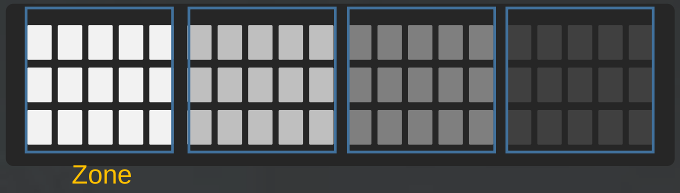
- QoS Isolation: Deterministic Read Latency w/ Noisy Neighbor
- QoS Differentiation: Configurable Throughput per Tenant



Summary: Technologies for TCO Reduction

WAF Reduction

- Adopting Zone concept from SMR HDDs
- Write Endurance Improvement



Computational Storage

- Samsung SmartSSD with FPGA

Machine Learning

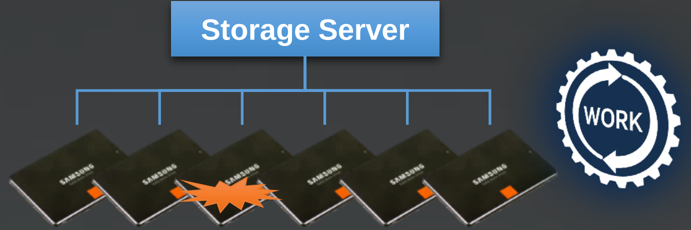
- Machine Intelligence for Storage Mgmt.

Security

- Technologies to make SSD secure
- Data privacy regulations

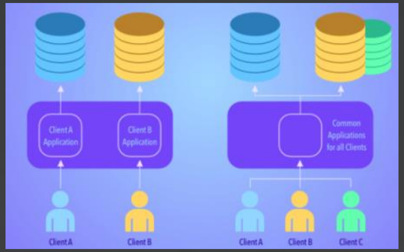
Reliability

- FIP : Never-die-SSD
- Error Protection for High Reliability



Multi-Tenancy

- Deterministic Latency with Noisy Neighbors



삼성 반도체 소프트웨어의 세계

https://www.youtube.com/watch?v=scRLlU_ZX1Y&t=65s


Thank
You