Open. Together.
Toward the large deployment of open channel SSD

Feng Zhu, Yu Du, Wentao Zhang, Sheng Qiu, Yan Liu, Shu Li
Alibaba Group
Agenda

Introduction to AliFlash

Replica SSD based on open channel

Multi-stream SSD based on open channel
AliFlash @ Alibaba

OEM
AliFlash V1
- Host-Based NVMe SSD
- Deployed since 2016
- > 50k pcs
- Serving DB / RDS / Finance / Search / CDN etc.

ODM
AliFlash V2
- NVMe SSD
- Deployed since 2017
- Serving DB / Search etc.

JDM
AliFlash V3
- Dual mode SSD, including open channel
- Ramping up deployment
- Targeting DB / RDS / Search / CDN / EBS etc.

Deployed since 2016
> 50k pcs
Serving DB / RDS /
Finance / Search / CDN etc.

• NVMe SSD
• Deployed since 2017
• Serving DB / Search etc.

• Dual mode SSD, including open channel
• Ramping up deployment
• Targeting DB / RDS / Search / CDN / EBS etc.
AliFSC V1

Alibaba’s first self-developed SSD controller AliFSC V1.

- Support Open Channel mode
- PCIe Gen 3 x 4
- NVMe 1.3
- 16 NAND channels, drive capacity up to 16TB
- 3 dual core Arm Cortex R5
- 4KB random read IOPS up to 800k
- Support all major vendors 3D TLC/QLC NAND

Kick off: 2016/10
RTL done: 2017/6
1st tapeout: 2017/10
1st chip back: 2017/12
AliFlash V3 early ES: 2018/2
Demand for HW/SW co-optimization
- Diversified, fast-changing workloads
- Continuous pressure for TCO and supply improvement
- “white box” of I/O path - more determinism
- Quick response to issues and requests
Replica SSD through open channel

- Build upon open channel SSD. Transparent to host software.
- Physical partitioning the NVM resources to two sets SSD-0 and SSD-1
- Each LBA has dual copies on the same drive
- Read and write are dispatched to different set
Replica SSD results

- Read-only like latency achieved
Multi-stream through open channel

- Multi-stream to improve WA and QoS: Separate the data with different life cycles into different streams
- Simplified FTL to handle the address collision
Multi-stream results

- Seqwriter1: sequential write (0, 32GB) --> log write
- Randwriter2: random write (64GB, 96GB) --> meta data update
- Randwriter3: random write (128GB, 640GB) --> data update
- Reader: random read (128GB, 640GB) --> data read
Multi-stream results

(normalized write amplification)

(normalized read latency)
Alibaba Open Channel Ecosystem

Alibaba is collaborating with major vendors in industry to build an ecosystem for Open Channel SSD

Share development & debug resources

Reduce time & complexity for SSD qualification in Alibaba

Massive deployment in 2019
What’s Next

Further I/O scheduling and QoS optimization

QLC deployment in open channel

On-line disk defect monitoring and prediction

...