Open-Channel SSDs for Host-Based Optimization

Yu Du, Feng Zhu, Sheng Qiu, Shu Li
Alibaba Group
A brief of AliFlash

AliFlash V1
- Host-Based PCIe SSD
- Deployed since 2016
- > 50k pcs

AliFlash V2
- Device based NVMe SSD
- Deployed since 2017

AliFlash V3
- Open Channel SSD
- Volume ramping up
- Targeting DB/RDS/Search/EBS etc.
First Productionized OC-SSD

- Alibaba’s home-developed Open Channel SSD - AliFlash V3
- Deployment ongoing in data centers
- Major milestone since the announcement of Alibaba’s Open Channel SSD Architecture in FAST’2018
- Collaborating with multiple SSD vendors to build an ecosystem
Block FTL Driver Overview

- Simplified FTL design using Alibaba Open Channel (AOC) Command Set
- FTL Core: LOC < 50K
- A single code base to support kernel/user modes
- Accelerate regression test with simulation mode
Non-Contiguous Read Optimization

- 8KB/16KB read request
- LBAs are mapped to the same multi-plane page, but not contiguous
- Non-contiguous vector read is not support yet (HW limit)
- Read extra dummy data to avoid multiple 4KB reads
Write Coalescing and Padding

- Reduce # of write commands for better IOPS and latency
- **Write Coalescing**: multiple 4KB/8KB host write requests are combined into a single write commands
- **Write Padding**: periodical padding to ensure host writes are translated into PU-aligned write commands
Host / GC Write Stream

- Up to 3 host streams and 1 GC stream
- Host and GC writes are scheduled independently
- Quota-based GC policy to balance host and gc writes
- Fewer free space → faster GC reads

Optimized for low-QD random write latency
Sequential Prefetching

- Sequential read IO patterns in some production workloads
- State machine based detector
- Host DRAM prefetching buffer
- Up to 5X read bandwidth improvement.
Diagnostic Support

- 300+ runtime diagnostic parameters:
  - IOPS
  - Latency/QoS
  - GC/WL
  - Media Error
  - FTL driver parameters
  - FTL key data structures
Application Mode Support

- Fine-tune FTL driver parameters and policies for different usage scenarios
  - Database
  - Distributed Block Storage Service
  - ...

- Dynamic Configuration
  - Runtime adjustment support to a subset of driver parameters.
Conclusion Remarks

- AliFTL: FTL driver implementation based on Alibaba open channel command set
- Read/write optimizations to reduce # of commands
- Multiple write stream optimized for low-QD random write latency
- Diagnostic support and application-based tuning
- Alibaba is open to industry collaboration
THANK YOU