What Flash Means to the Future of Storage System Architectures

Starboard Storage Systems
Kirill Malkin
Lee Johns
Flash in Storage Systems Today

• The Use of Flash in Storage Solutions:
  • Storage Tier
    • As a replacement for HDD
  • Storage System Writeback Cache
    • Absorbs rapid-fire writes
  • Storage System Read Cache
    • Extension of DRAM buffer cache
  • Server-side Cache
    • Eliminates networking latency
Different Shapes & Forms of Flash

- NAND chip types
  - SLC, e/MLC, TLC…
- Attachment type
  - SATA, SAS, PCIe
- Wear leveling algorithms
  - Rewrite cycles (longevity)
- DRAM front-end
  - Performance, read disturb mitigation
- Write completion guarantee
Basic Storage System Architecture

- NVRAM absorbs writes
- NVRAM assists RAID updates
- Buffer cache facilitates read-ahead
- NVRAM 1-8GB
- Cache 16-64GB
Storage System with Flash SSDs

- Replace HDDs with SSDs
- Avoid full RAID rebuilds
- Compression and deduplication
  - Reduce data footprint to mitigate write amplification
  - Speed up reads, improve caching
All-SSD Systems have Issues

- **Endurance & Longevity**
  - How long will it last given the load?
  - SLC is good, MLC not so much
  - Some devices don’t fail, just slow down
- **Lower capacity compared to HDD**
  - Especially SLC
  - Particularly on non-compressible, non-dupe data
- **Cost**
  - SLC is $$$$$$
  - MLC is $$
Active Data Footprint

• Modern Storage Systems feature large data capacity (30TB-1PB)
• Yet only about 5% of data is active at any one time: e.g. for a 30TB system, only 1.5TB
  • For a typical set of workloads
  • Except for initial loads and full backups
• If the active I/O is always directed to Flash, we can get Flash-like performance for the entire Storage System
Cached or Tiered? Accelerated!

- Absorb random writes with flash
  - Heavy write streams go directly to pool
- Large writeback area acts as a “tier”
- Frequently accessed stripes stored in flash
  - Heavy read streams go directly from pool
- Flash stratification:
  - SLC for the most critical loads and metadata
  - Disposable MLC for read caching
  - Redundant MLC/TLC for the dynamic pool
- Flash is added if the active footprint grows
- The pool acts as an archive tier
Hybrid Storage System Architecture

- **WRITES**
  - Writeback Accelerator
  - SLC, mirrored, 0.1-1TB
  - Circular buffer
  - Linearized flush
  - PCIe or SAS

- **READS**
  - Read Accelerator
  - MLC, linear, SAS
  - Adaptive Algorithm
  - Optional compression

- **Spillover SSD**

- **Pool**

- **Stripe Cache**

- **Redundancy**
  - Transactional writes
  - Optional compression & deduplication
  - HDD or MLC/TLC SSD
  - SAS or SATA

Streams
Starboard AC Series Systems for Mixed Workloads

- Unstructured Data
- Virtualized Data
- Structured Data
Summary

• High-performance Flash memory and SSDs can help accelerate reads and writes in mostly-flash and hybrid storage systems
• Multi-level caching architecture accommodates a broad spectrum of Flash devices available on the market
• With only 5% active data footprint, multi-level caching can deliver the performance of the best Flash device for the entire storage system
• Adaptive autonomic tiering, caching and linearization algorithms are required to fulfill this promise
Questions?

http://www.starboardstorage.com