iFlash

Shailendra Tripathi

Western Digital
Agenda

- Background
- Resource Utilization
- iFlash
- L2ARC Model
- Results and Analysis
- All Flash and Application Use cases
- Conclusion
Spa – Tegile Flash Meta
Resource Utilization

- Dedicated Devices
  - Inflexible
  - Utilization Problem
    - Capacity
    - IOPS/Bandwidth
- Partitions
  - Somewhat Flexible
  - Static size partitioning
  - Shared Device – “Noisy” neighbor
- Shared Device
Shared Devices

- Shared Capacity
- Shared IOPS / Bandwidth
- Varying Data Protection Need
  - Cache failure protection vs meta
- Widely different Characteristics
  - Throughput vs IOPS
- Latency vs Efficiency of Allocation
- Quality of Service
  - Priority / Weight
  - Enforcement
Priority

Slice

LOG
LOG
Data/Cache

Priority Queue

LOG
Meta
Data/Cache
iFlash - Implementation

- Per Class Allocator
- Per Class Queue Priority
- Per Class Metaslab Load Policy
- Per Class Data Protection Type
  - Meta/Data – Mirror, Triple Mirror
  - Cache – No redundancy
  - Log – Mirror
- Utilize 3 copies of blkptr, Strict device Enforcement
Dynamic Resource Sharing

- Policy
- Change Detection Infrastructure
- Action
- Maintain
  - Capacity
  - Fragmentation level
  - Weight / Priority Assignment
  - IOPS and MBPS rate
Dynamic Capacity Sharing
Enabler Changes

- Independent ARC and L2ARC
  - Independent Write Scale
- Compact 64 byte in-core structure
- Indexing similar – extra CPU cost, massive memory saving
- L2ARC stored as 1 MB page
- Buffers stored byte-compressed.
- Header Page contains 48 byte mapping on the 1 MB page
- 1 MBs stored as per-device linked-list per rooted in uberblock
Device – Page Array
Page and Indexing
Performance
Performance

Latency 4K

RR150P  RR200P  RR400P  SR400P

Latency iFlash  Latency Legacy
Performance

Latency 32K

RR50P  RR20P  RR0P  SR0P
Latency iF lash  Latency Legacy
Conclusion

- Efficient Resource Sharing
- Adapts dynamically
- Exploits the bi-modal patterns
- Aggregate better performance (both IOPS and lower latency)
- Extensible – All Flash
- Benefits extensible to application
  - DB /FS journal, Index