The Failure of SSDs

How Integration into the Storage Hierarchy Will Advise SSD Design

Adam Leventhal
Senior Staff Engineer
Sun Microsystems / Fishworks

Who Am I?

- Engineer in Sun's Fishworks group
- Project to deliver unified storage
- Integrate existing components
  - Solaris, DTrace, FMA, SMF
  - x64 platforms
  - CIFS, NFS, iSCSI
  - ZFS

ZFS

- Complete storage system
- Incorporates volume manager and filesystem
- Obviates the need for hardware RAID controllers
- Designed to turn commodity parts into enterprise-class storage
- Problem: commodity HDDs are really slow

Flash SSDs

- Disks: $/GB, W/GB, $/MB/s
- SSDs: $/IOPS, W/IOPS
- Use flash to accelerate ZFS

Hybrid Storage Pool (HSP)

- Adapted ZFS to integrate flash
- ZFS intent-log (ZIL) device
  - Accelerate small, synchronous writes
- Second Level Adaptive Replacement Cache (L2ARC)
  - Larger caching tier than ARC (DRAM)
  - “Evict-ahed” cache
  - Accelerate reads

HSP Needs

- ZIL: Logzilla
  - Writes: low-latency / high-IOPS
  - Reads: not performance critical
  - Low capacity needs (8GB or more)
- L2ARC: Readzilla
  - Reads: low-latency / high-IOPS
  - Writes: just don't get in way of reads, please
  - High capacity / low $/GB
- No perfect SSD, but several close enough

HSP Results (1/2)

4 Xeon 7350 Processors (16 cores)
32GB FB DDR2 ECC DRAM
OpenSolaris with ZFS

Configuration A:

(7) 146GB 10,000 RPM SAS Drives

Configuration B:

(1) 32G SSD ZIL Device

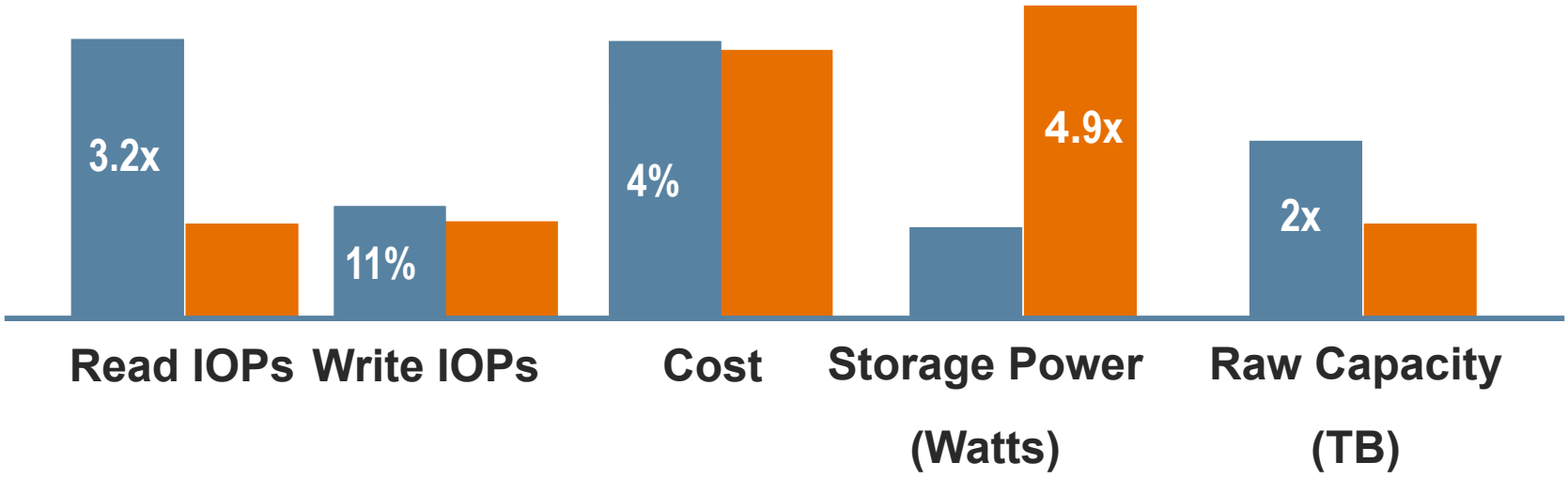
(1) 80G SSD Cache Device

(5) 400GB 4200 RPM SATA Drives

HSP Results (2/2)

Hybrid Storage Pool (DRAM + Read SSD + Write SSD + 5x 4200 RPM SATA)

Traditional Storage Pool (DRAM + 7x 10K RPM 2.5")



State of SSDs 2007

Compatibility for Laptops and Rugged Environments

- Replacements for low-end hard drives
- Relatively cheap
- Focus on low power, environmental tolerance
- Performance: terrible
  - > Devices with 36ms latency (!!)
- Reliability: awful
  - > Accelerated write/erase cycle exhaustion
- Lousy FTL, controllers gave flash SSDs a black eye

State of SSDs 2008

Performance and Reliability for the Enterprise

- Replacements for 15K enterprise drives
- Boutique vendors
- Absolute performance, not price/performance
- Increased complexity logic to accelerate devices
- Decreased observability into devices

State of SSDs 2009

Volume and Scale

- High quality, volume prices
- Intel, Samsung, Toshiba, et al.
- SLC for the enterprise
- MLC for the desktop
- Margin is disappearing

Failings of Flash SSDs

- Constrained by design goal of complete compatibility with hard drives
- Unique properties of flash obscured
- Better ways to interface than SAS/SATA/FC
- Closed solution makes it harder to integrate flash into the storage hierarchy

Interface / Connectivity / Form Factor

- SAS/SATA/FC was convenient
- Fine as a replacement device
- Inefficient and anachronistic
- SSDs will find a better interface

FTL: Good / Bad / Ugly

- SSD controllers present flash as disk to OS
- Don't need to change OS or application
- Not a direct map e.g. “RPM = 1”
- Spare sectors to accelerate writes
- Role of controllers will change
  - More collaboration with higher level software
  - Less complex in some ways
  - Enable broader innovation higher up the stack
  - TBD: what does the SSD do vs. what does the FS do?

Broadly Purposed

- Flash: great for reads, good for writes
- SSDs strain to get both right
- Better path: optimize for a specific use case
- HSP
  - Logzilla: small capacity, write IOPS, rarely read
  - Readzilla: huge capacity, read IOPS, slow writes
- Both devices simpler and cheaper to build than an SSD that aims to solve both

Conclusion

- SSDs gained adoption with hard disk compatibility
- Time to re-evaluate the other side
- Caveats
  - Economics is king
  - Some element of compatibility must be preserved
  - Partial solutions first until broader markets evolve
- Slow transition that ***we*** can guide
  - Promote compelling uses of flash
  - Work with vendors
  - Vote with your wallet

Thank You!

Adam Leventhal
blogs.sun.com/ahl