RAID4S: Adding SSDs to RAID Arrays

Rosie Wacha[1], Scott Brandt[1], John Bent[2], and Carlos Maltzahn[1]
[1]University of California, Santa Cruz
[2]Los Alamos National Laboratory

RAID (Redundant Array of Independent Disks)

RAID algorithms provide better performance and/or data reliability by combining several disk drives.

RAID4 and RAID5 store the parity computed across data blocks to protect against a single disk failure.

Computing parity for small I/Os significantly reduces performance. A common technique for mitigating this problem is to transform small writes into larger writes. However some small writes are inevitable, particularly as disk drives fill up.

Solid State Drives (SSDs)

SSDs compared to disk drives:
- Constructed of flash RAM thus not mechanical devices
- Faster random I/O
- Lower power consumption
- Wear out with each write
- More expensive

SSDs are not cheap enough to replace all disk drives. This work integrates SSDs into disk-based storage systems.

Acknowledgements

This work was funded by the Institute for Scalable Scientific Data Management (ISSDM).

Solving the Small Write Problem with RAID4S

Small writes limit RAID performance by requiring 4 I/Os per data block written. The old data and parity is read, the new parity is calculated, and the new data and parity is written.

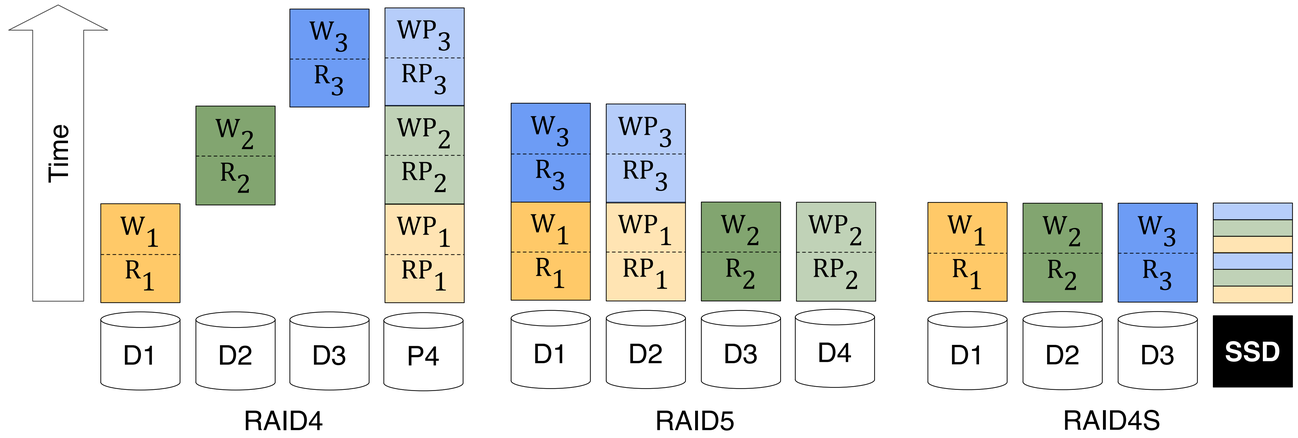

Figure 1. The time to complete three small writes is compared. RAID4S with three disk drives allows the complete parallelization of M=3 accesses to different disks.

Small Write Analysis

One small write = 4 I/Os
- Read old data and parity
- Write new data and parity
- Data and parity accessed in parallel

RAID4: parity drive is bottleneck
RAID5: data and parity accessed in parallel
RAID4S: parity offloaded to SSD and all data accesses are parallelized

For M small writes to N disks:
- RAID4: $2M$
- RAID4S: $2\lceil M/N \rceil$
- RAID5: $2*2\lceil M/(N+1) \rceil$

RAID4S speedup over:
- RAID4: $M/\lceil M/N \rceil$
- RAID5: $2\lceil M/(N+1) \rceil \ / \ \lceil M/N \rceil$

Calculated Throughput

- M outstanding I/Os at the controller
- N = 4 disk drives for RAID4S + 1 SSD
- N+1 = 5 disk drives for RAID 4 and RAID5

Assumptions:
Each small write incurs a seek and transfer time. Small writes are to different stripes. Thus, the I/O is parallelized as much as possible.

Throughput calculation:
- 64KB small writes
- Each I/O incurs a seek and some transfer time
- Outstanding I/Os are queued at the controller and are completed in parallel, if possible.

- Hardware:
  - Western Digital WD20EADS (low power disk drive)

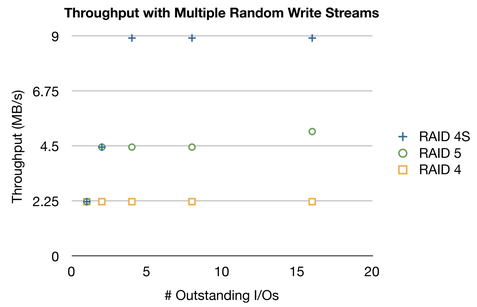

Figure 2. Throughput of small writes. RAID4S completes N small writes in the time that RAID4 completes one small write, as long as the SSD is N times faster than hard disks in the array.

RAID4S is up to 2X faster than RAID5 and 4X faster than RAID4

Conclusions

Flash SSDs provide better performance and new opportunities for data storage in distributed systems. This work replaces the parity disk drive in a RAID4 system with SSD. Initial results show improvements of up to (N/2)X speedup over RAID5 and NX speedup over RAID4.

Future Work

More general workloads
- Mix small and large writes
- Workload traces

Build RAID4S with hard drives and flash and measure:
- Throughput
- Latency
- Power

Investigate reliability impact of RAID4S