RAID4S: Improving RAID Performance with Solid State Drives

Rosie Wacha

UCSC: Scott Brandt and Carlos Maltzahn
LANL: John Bent, James Nunez, and Meghan Wingate

SRL/ISSDM Symposium
October 19, 2010
RAID:
Redundant Array of Independent Disks

- RAID0: striped
- RAID1: mirroring
- RAID4: dedicated parity
- RAID5: distributed parity
- RAID6: two parities
RAID: Redundant Array of Independent Disks

- RAID0: striped
- RAID1: mirroring
- RAID4: dedicated parity
- RAID5: distributed parity
- RAID6: two parities
Flash SSDs Replacing Disks

- Laptops
- Sensor networks
- Satellites
- Data centers (EuroSys '09)
  - Not cost-effective to replace hard drives
  - Caching tier only cost-effective for 10% of workloads
Our Solution: Replace Some Disks with Flash

- Flash SSDs are available, fast, and expensive.

- RAID 4 + SSD = RAID4S
Large, Sequential Writes (RAID4&5)
Large, Sequential Writes (RAID4&5)

- N write requests $\rightarrow$ N+1 writes to disk
  - N data writes and 1 parity write
RAID Small Write Problem

• 1 write → 2 reads + 2 writes
• Other solutions avoid small writes
  - Coalesce, log, NVRAM
• For remaining small writes
  - Use solid state drives!
  • Faster, lower power, but more expensive
RAID Small Write Problem

- 1 write → 2 reads + 2 writes
- Other solutions avoid small writes
  - Coalesce, log, NVRAM
- For remaining small writes
  - Use solid state drives!
    - Faster, lower power, but more expensive
RAID Small Write Problem

- 1 write → 2 reads + 2 writes
- Other solutions avoid small writes
  - Coalesce, log, NVRAM
- For remaining small writes
  - Use solid state drives!
    - Faster, lower power, but more expensive
RAID Small Write Problem

- 1 write → 2 reads + 2 writes
- Other solutions avoid small writes
  - Coalesce, log, NVRAM
- For remaining small writes
  - Use solid state drives!
    - Faster, lower power, but more expensive
RAID Small Write Problem

- 1 write → 2 reads + 2 writes
- Other solutions avoid small writes
  - Coalesce, log, NVRAM
- For remaining small writes
  - Use solid state drives!
    - Faster, lower power, but more expensive
RAID Small Write Problem

- 1 write → 2 reads + 2 writes
- Other solutions avoid small writes
  - Coalesce, log, NVRAM
- For remaining small writes
  - Use solid state drives!
    - Faster, lower power, but more expensive
RAID4S Solves Small Write Problem
RAID4S Solves Small Write Problem

RAID5 parallelizes some small writes
RAID4S Solves Small Write Problem

RAID5 parallelizes some small writes

RAID4S parallelizes N=4 small writes

D1 D2 D3 D4 D5
RAID5

D1 D2 D3 D4 D5
RAID4S

SSD
Experimental Setup

- Hardware experiment using Linux RAID software mdadm
- Intel X25-E 64GB
- 5 Western Digital Caviar Black 640GB 7200 RPM 32MB Cache SATA 3.0Gb/s 3.5"
- 4+1 arrays
  - RAID4
  - RAID4S
  - RAID4STUPID
  - RAID5
  - RAID5S
Performance is Equal for Sequential Write

- Ran `dd` to write files
  - 1MB IO size
  - 4GB total IO
- Same performance
  - Large writes fill stripes
  - No small write problem
Random Writes Setup

• XDD 6.5 benchmark
  - 100% random write
  - Repeat 3 times and plot average

• Two different IO sizes:
  - 4KB to 1GB (powers of 2); 1GB total
  - 1KB to 16KB (every one); 256MB total
RAID4S 1.6X Faster Than RAID5
RAID4S 1.6X Faster Than RAID5

128KB
Smaller Small IOs

• 64KB and lower
Smaller Small IOs

- 64KB and lower
Smaller Small IOs

- 64KB and lower

\[ P = P' \oplus D_1' \oplus D_1 \]
Smaller Small IOs

- 64KB and lower

\[ P = P' \oplus D_1' \oplus D_1 \]

N data disks
Smaller Small IOs

- 64KB and lower

\[ P = P' \oplus D_1' \oplus D_1 \]

N data disks
Larger Small IOs

- 128KB and above

\[ D_1 \quad D_2 \quad D_3 \quad \ldots \]

\[ D_1 \quad D_2 \quad D_3 \quad \ldots \quad D_N' \]

N data disks
Larger Small IOs

- 128KB and above
Larger Small IOs

- 128KB and above

\[ P = D_1 \oplus D_2 \oplus D_3 \oplus D_N' \]

\[ D_1 \quad D_2 \quad D_3 \quad \ldots \quad D_N' \]

N data disks
Larger Small IOs

- 128KB and above

\[ P = D_1 \oplus D_2 \oplus D_3 \oplus D_{N'} \]

\[ D_1 \quad D_2 \quad D_3 \quad \ldots \quad P \]

N data disks
Larger Small IOs

- 128KB and above

\[ P = D_1 \oplus D_2 \oplus D_3 \oplus D_N' \]

\[ \text{N data disks} \]
4KB-Unaligned Writes

![Graph showing throughput (MB/s) for different IO sizes (KB) across various RAID configurations.]

Throughput Normalized to RAID5

![Graph showing throughput normalized to RAID5 for different IO sizes (KB) across various RAID configurations.]
4KB-Unaligned Writes

Throughput (MB/s)

Throughput Normalized to RAID5

IO Size (KB)

"RAID4S"
"RAID5S"
"RAID5"
"RAID4STUPID"
"RAID4"
Conclusions and Future Work

• RAID4S speeds up small writes
  - 3.3X over RAID4
  - 1.6X over RAID5

• Status/Future
  - Experiments driven by I/O workload traces; mixed benchmarks
  - Verification of results with tracing
Questions?

rwacha@cs.ucsc.edu
SSD Reliability

- 64GB Intel SSD - 2PB random write lifetime
- RAID4S
  - 100MB/s constant writes: lifetime is 7.7 months
  - 25MB/s: 30.7 months or 2.5 years