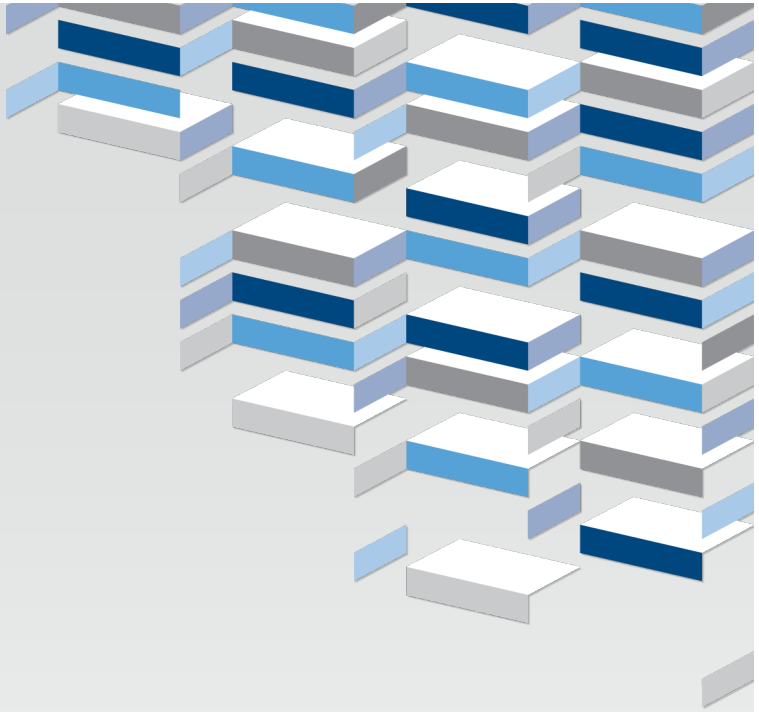
coraid
REDEFINING STORAGE



ZFS at Coraid

Richard Elling

About Coraid

- Invented ATA over Ethernet (AoE)
- Deliver cost-effective, scale-out storage products using the ubiquitous Ethernet interconnect
- ZFS-based products provide NFS and AoE
- Global customer base
- Petabytes and Petabytes served

coraid

Instrumentation

- Collectd agents for ZFS
  - Efficient – written in C, uses kstats when possible
  - Comprehensive – multiple OSes and variants
  - Insightful – see how things work
  - Trendy – record telemetry data for the long-term
- ARC, XUIO, Prefetch Kstat Analyzer
  - Replaces the venerable, but outdated arc_summary.pl
  - Retrospective – analyze kstats from yesteryear

coraid

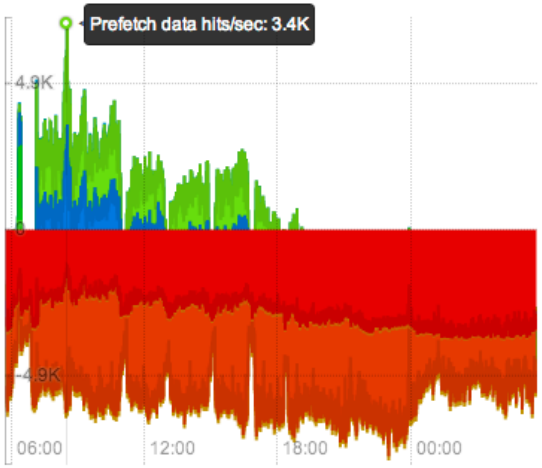
street ARC Statistics

Send Richard Elling your feedback.

coraid

Sun, 30 Mar 2014 08:32:00 GMT

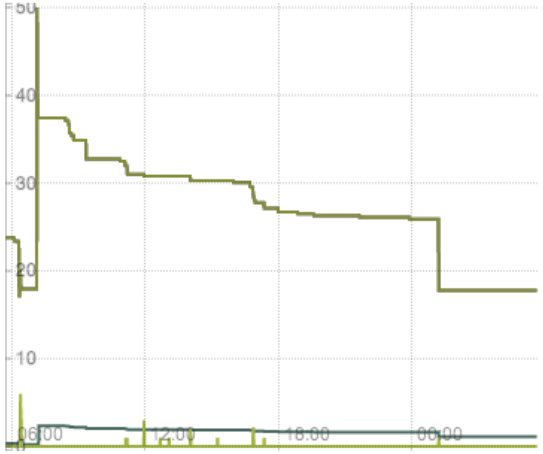
ARC Hits/Misses



Hits and misses by type

Prefetch metadata hits/sec	▼:0	▲:150	⬍:6.6
Prefetch data hits/sec	▼:0	▲:3.4K	⬍:1.0K
Demand metadata hits/sec	▼:0	▲:6.0K	⬍:2.3K
Demand data hits/sec	▼:0	▲:5.4K	⬍:1.9K
Prefetch metadata misses/sec	▼:-200	▲:0	⬍:-150
Prefetch data misses/sec	▼:-4.7K	▲:0	⬍:-2.6K
Demand metadata misses/sec	▼:-130	▲:0	⬍:-78
Demand data misses/sec	▼:-3.7K	▲:0	⬍:-2.7K

ARC Sizes



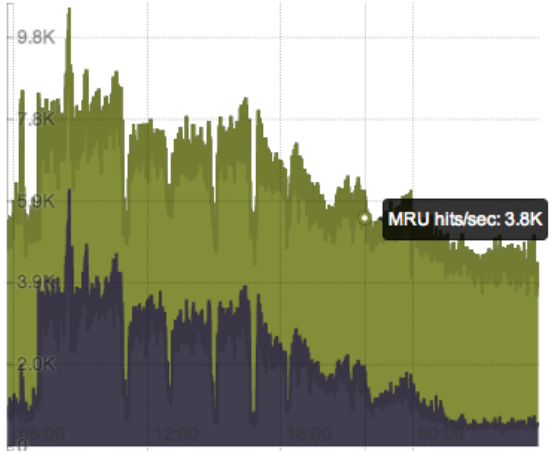
ARC Sizes in GiB

ARC reap	▼:0	▲:6.0	⬍:0.021
MRU target size	▼:0	▲:2.3	⬍:1.6
ARC target size	▼:0	▲:50	⬍:27
ARC size	▼:0	▲:50	⬍:27

Sun, 30 Mar 2014 21:

MFU/MRU

1,536



Most frequently and recently used cache hits

MRU ghost hits/sec	▼:0	▲:0	⬍:0
MFU ghost hits/sec	▼:0	▲:0	⬍:0
MRU hits/sec	▼:0	▲:7.1K	⬍:4.1K
MFU hits/sec	▼:0	▲:6.1K	⬍:1.9K

coraid

Why ZFS?

- Storage hardware breaks – all of it
- Firmware has bugs – everyone's
- Breakage arrives in strange and mysterious ways

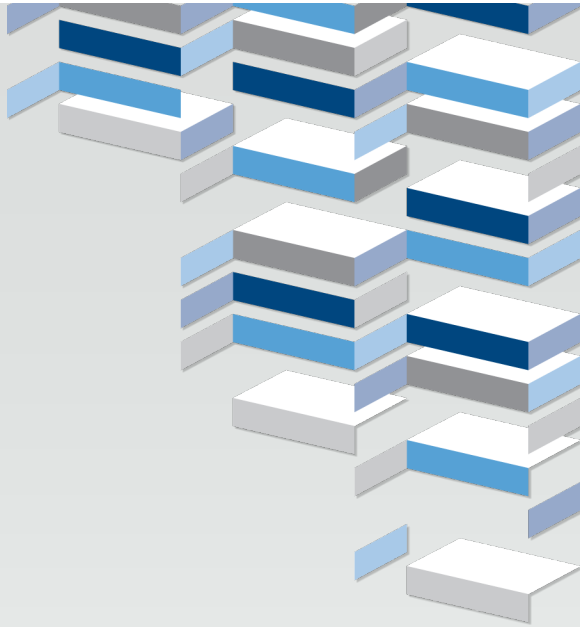
```
 pool: zpool-anonymous
state: DEGRADED
 scan: resilver in progress since Tue
   168T scanned out of 444T at 110M/s, (scan is slow, no estimated time)
         NAME                      STATE     READ WRITE CKSUM
         ...
         c1t2d5                  OFFLINE    199     0   449  (resilvering)
         c1t3d4                  DEGRADED   601     0 24.2K  (resilvering)
         c1t0d27                 DEGRADED   108     0 1.02K  (resilvering)
         c1t9d33                 DEGRADED 2.28K     0  112K  (resilvering)
         ...
         c1t0d34                 ONLINE       0     0     1  (resilvering)
         c1t2d6                  DEGRADED   200     1 5.84K  (resilvering)
         c1t2d12                 DEGRADED   189     0 2.38K  (resilvering)
         ...
```


coraid

```
errors: No known data errors
```


coraid

coraid



Thank you