

Fusion-io and MySQL 5.5 at Craigslist

Jeremy Zawodny

Jeremy@Zawodny.com

jzawodn@craigslist.org

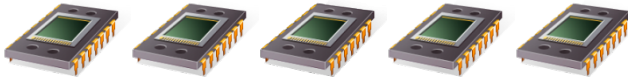
Story Time

(but first, some architecture
and numbers)

Some Numbers

- ~100,000,000 postings in live database
- Over 1,000,000,000 page views daily
- High churn rate (avg lifetime ~14 days)
- ~350-500GB on disk
 - MySQL 5.5.x and InnoDB Compression
 - Used to be ~100-150GB larger (or more!)
- All records touched multiple times
- 98% of queries are OLTP

The Posting Cache



Web Server Tier
(apache/mod_perl)



Posting Cache Tier
(memcached + perl)

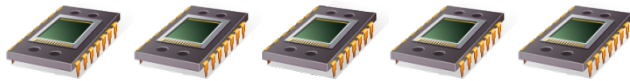


Database Tier
(MySQL)

The Problem

- Adding more memcached nodes
- Lots of cache misses initially
- MySQL boxes take a big query load
- (time passes)
- MySQL boxes pegged many hours later
- (time passes)
- Next day: WTF?!

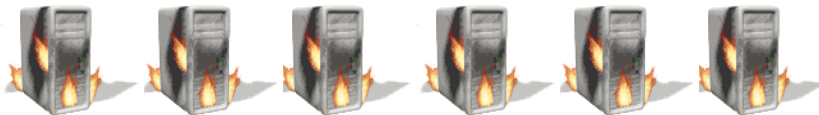
Fire!



Web Server Tier
(apache/mod_perl)



Posting Cache Tier
(memcached + perl)



Database Tier
(MySQL)

We were sending **nearly 30%** of requests all the way back to the DB tier instead of the normal 2-5%

Solution

- Let's put the New Hardware in the pool
 - Add 4 machines
- And it still sucked...
 - The 4 were fast but only took ~20% of the hits
- Remove all the Old Hardware
 - Remove 14 of 18 machines
- Sounds totally sane, right?

Old Hardware

- 3 years old
- 3U, Dual AMD 2218 HE
- 32GB RAM
- 16 15k RPM SAS disks
- RAID-10
- ~2,000 iops/sec
- ~325 watts

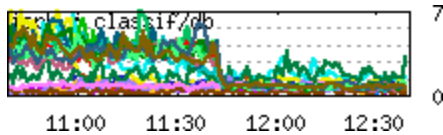
New Hardware

- HP DL-380G6
- Intel Xeon X5570
- Dual Proc, Quad Core, HT
 - 16 “cores”
- Dual Fusion-io 640GB SLC
- Software RAID-0
- ~80,000 iops/sec (conservative)
- ~200 watts

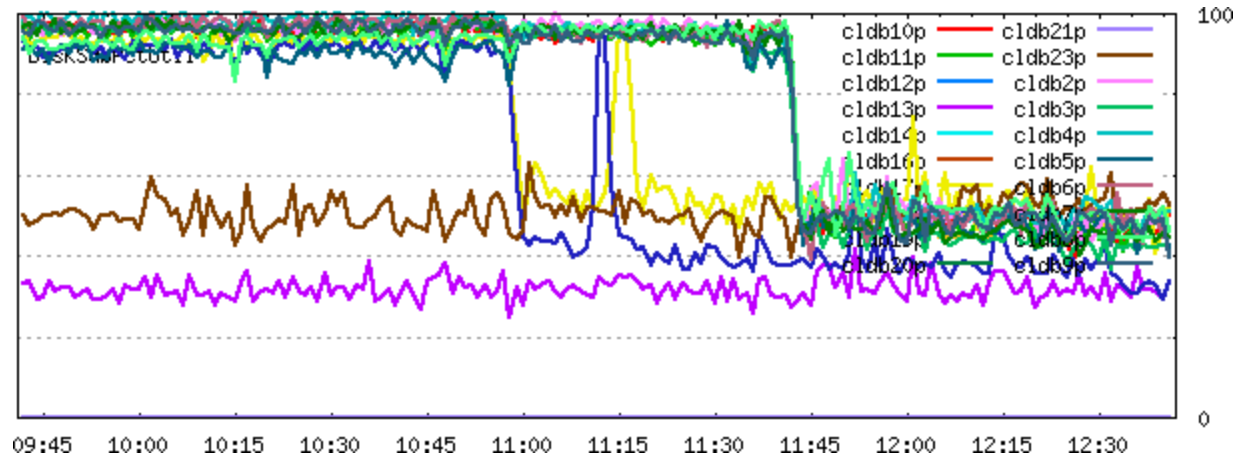
Before and After

- Night and Day!
- Old boxes return to “steady state”

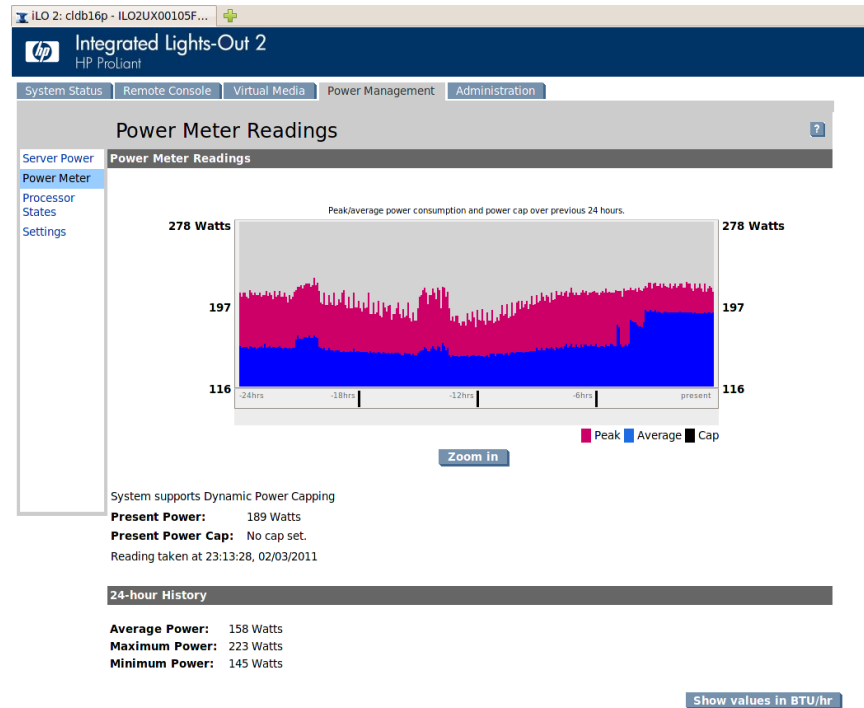
Load Average



I/O Capacity of Data Disks



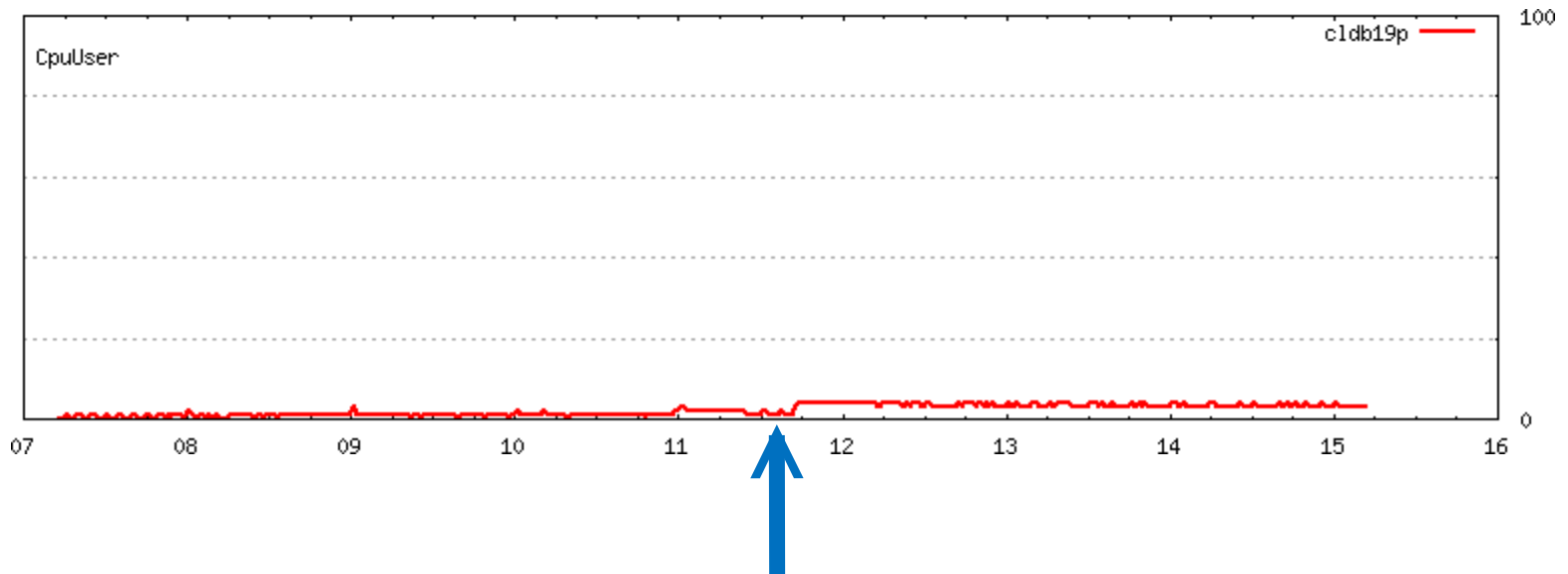
This Chart Should be Green



Average power for Fusion-io equipped server: **~200 watts**.
It was closer to 160 when replicating but not serving traffic.

Fusion-io FTW

- Can you tell when this machine started getting live traffic?
- OLTP means disk matters WAY more than CPU



Into the fire!

The Numbers

- Old: $2,000\text{iops} / 325\text{W} = 6.15 \text{ iops/watt}$
- New: $40,000\text{iops} / 200\text{W} = 200 \text{ iops/watt}$
 - Conservatively assumes a lot of degradation

33-66x performance/watt

But let's just call it 50x

Epilogue

- A week later, we re-purposed 1 Fusion-io box
- The cache eventually did fill
 - Poor slab size configuration had been causing early expiration of cached objects
- 14 “old” servers: 4,500 watts
 - 28,000 iops/sec capacity
- 3 “new” servers: 570 watts
 - 240,000+ iops/sec capacity
- What to do with 10+ spare “db class” boxes?