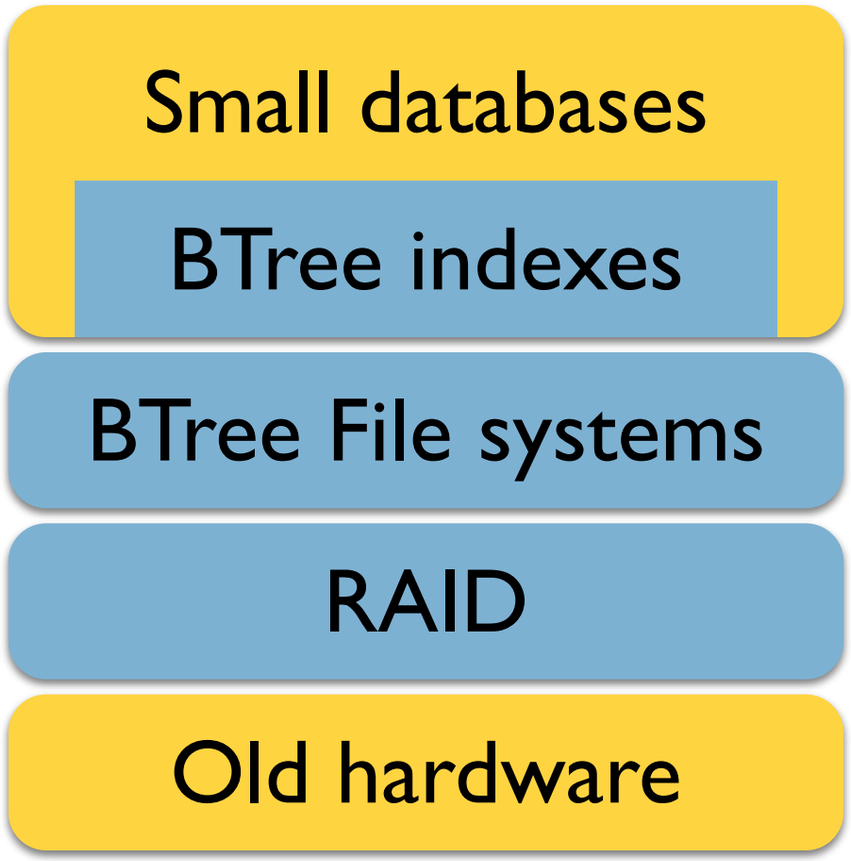
Supercharging Cassandra...

Tom Wilkie
Founder & VP Engineering
@tom_wilkie

Acunu

Before the Flood

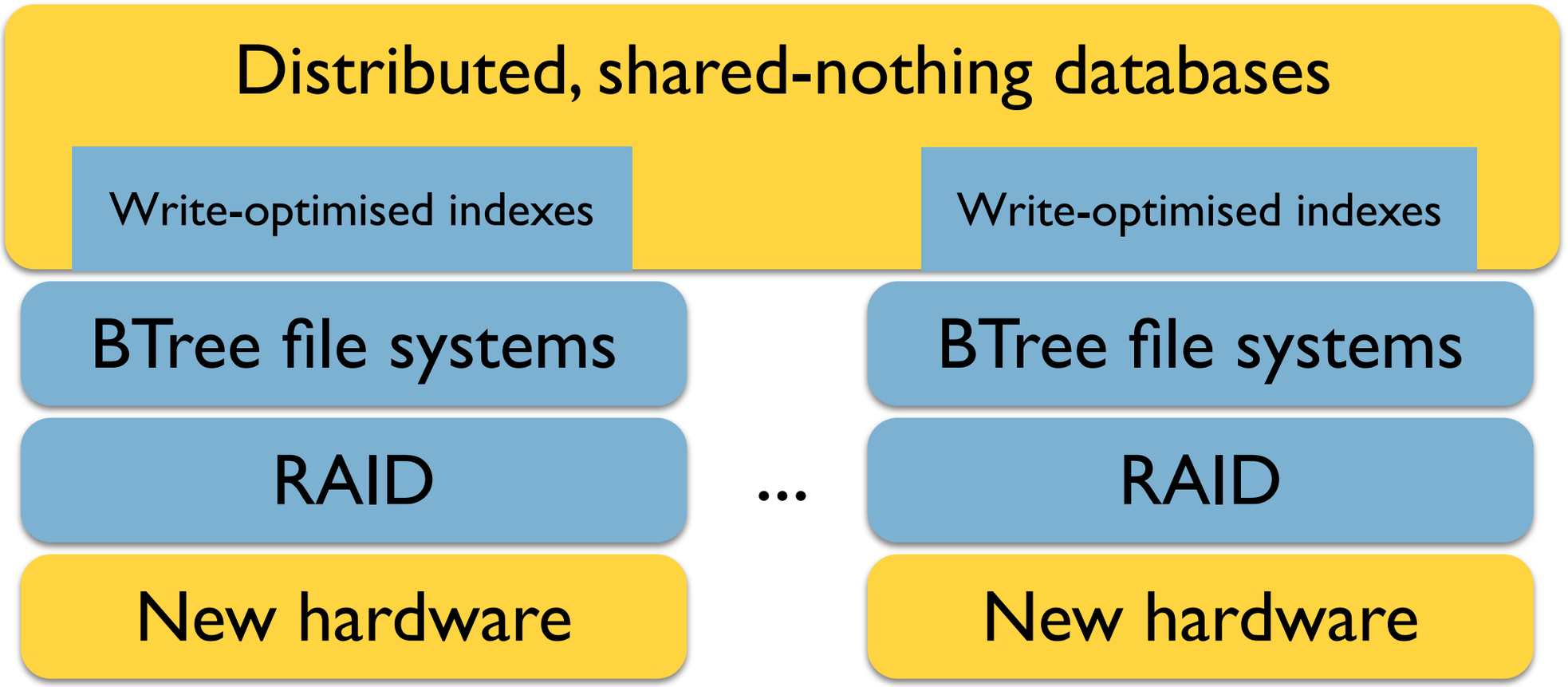
1990



Acunu

Two Revolutions

2010



Acunu

Bridging the Gap

2011

Distributed, shared-nothing databases

Castle

Castle

...

New hardware

New hardware

Acunu

Acunu

Open API

Management

Deployment

Monitoring

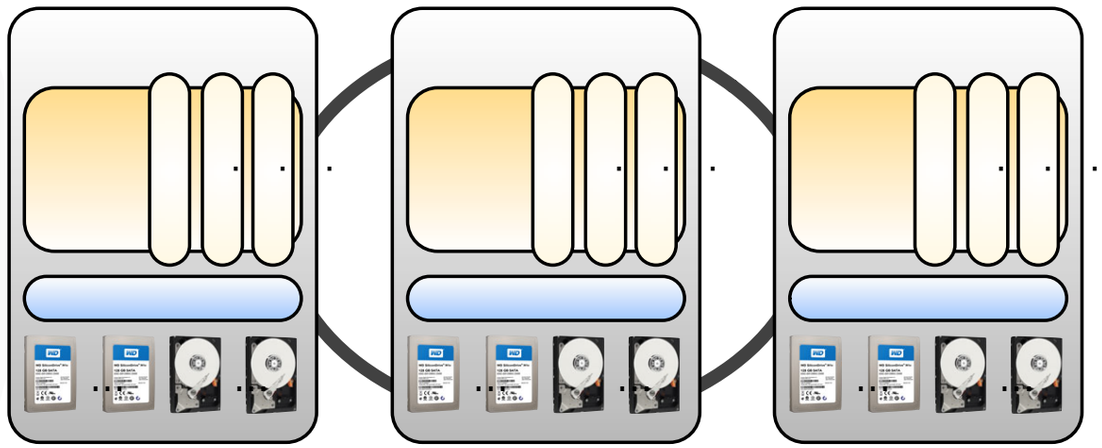
Cassandra

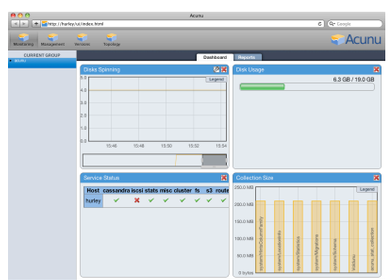
Memcached

...

Acunu Storage Core

Big Data Applications





Cross-Cluster Management UI

Acunu

1. Predictability

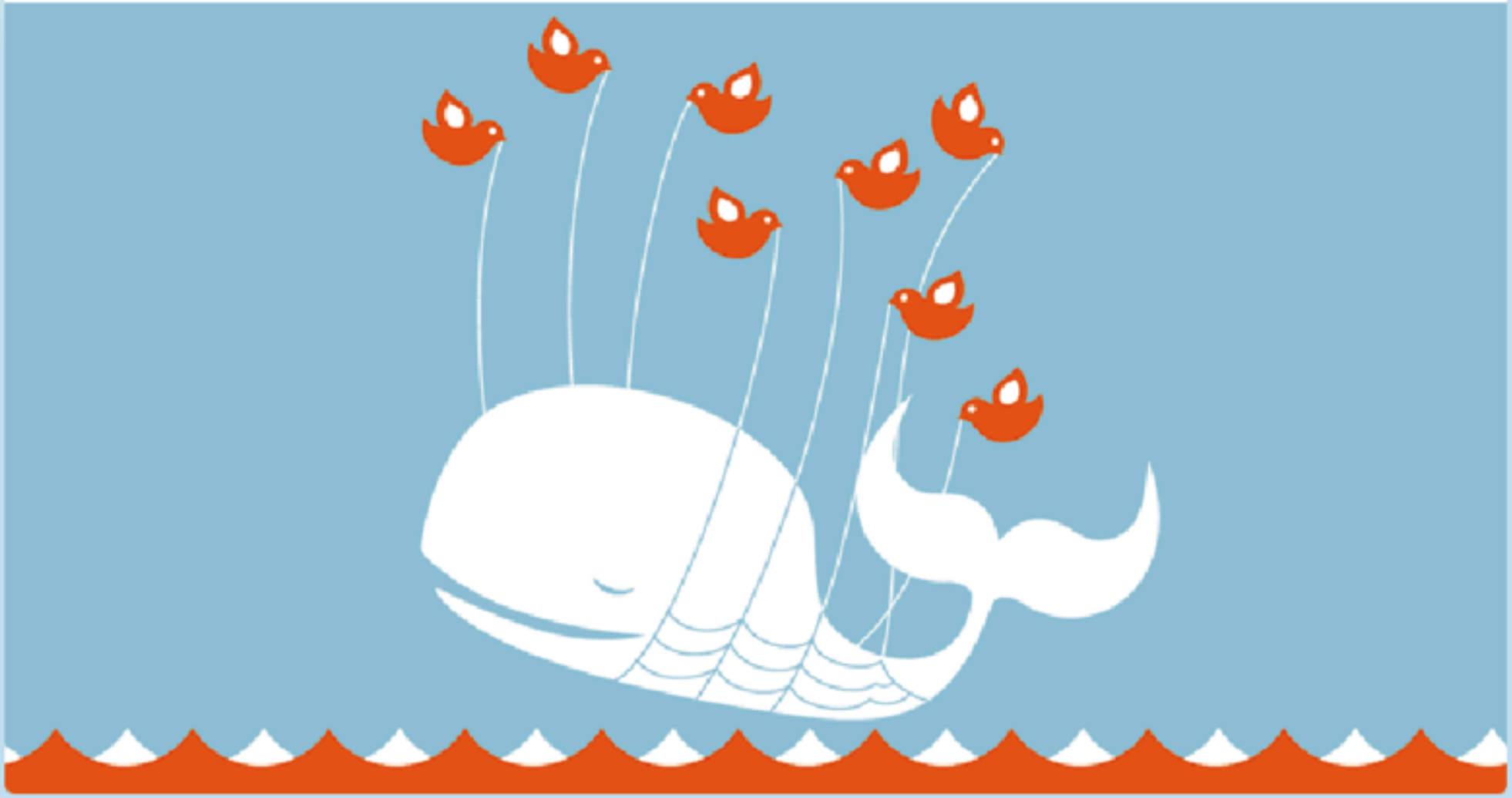
Acunu

twitter

Home ›

Twitter is over capacity.

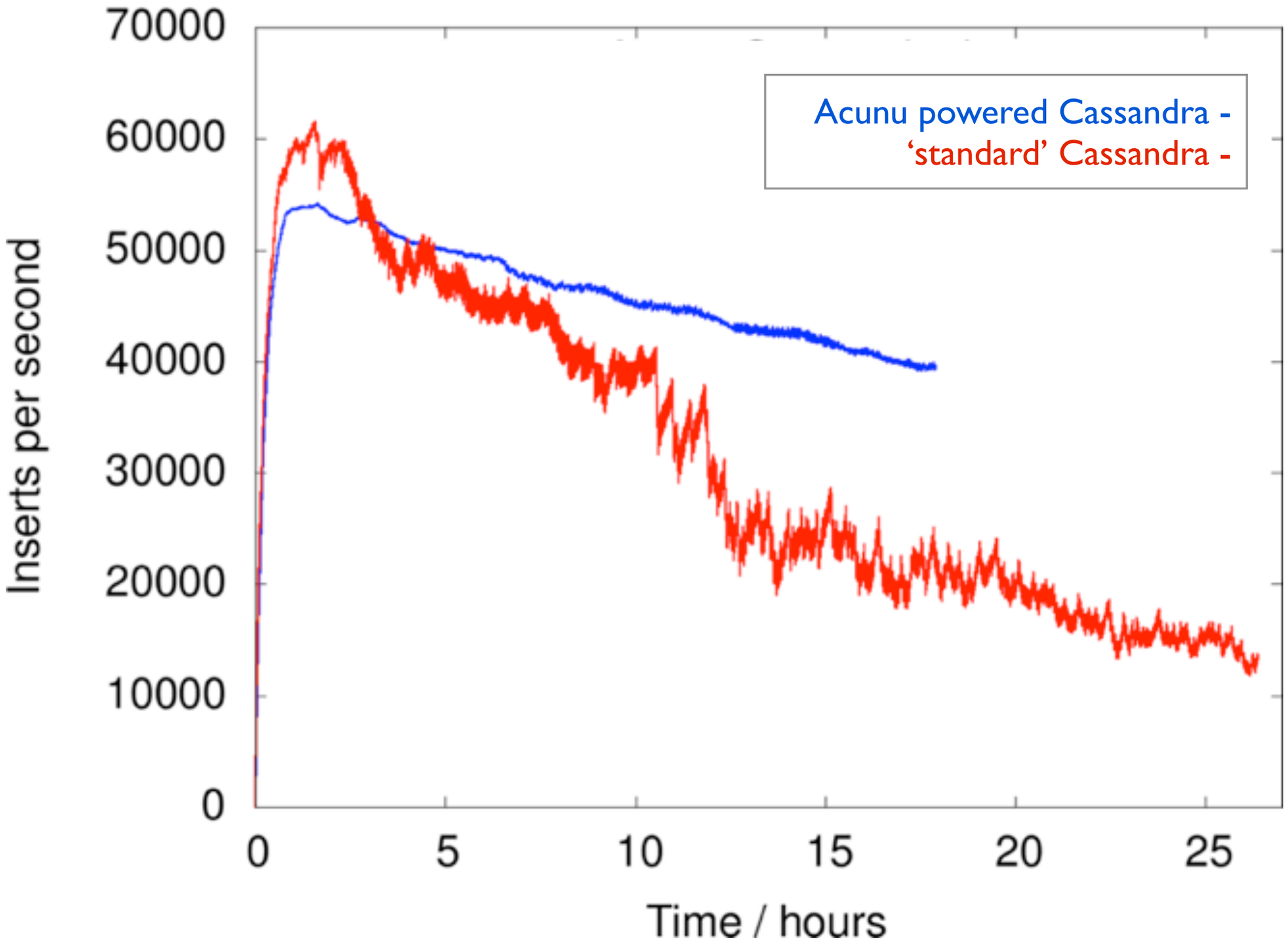
Please wait a moment and try again. For more information, check out Twitter Status »



English Deutsch Español Français Italiano

Small random inserts

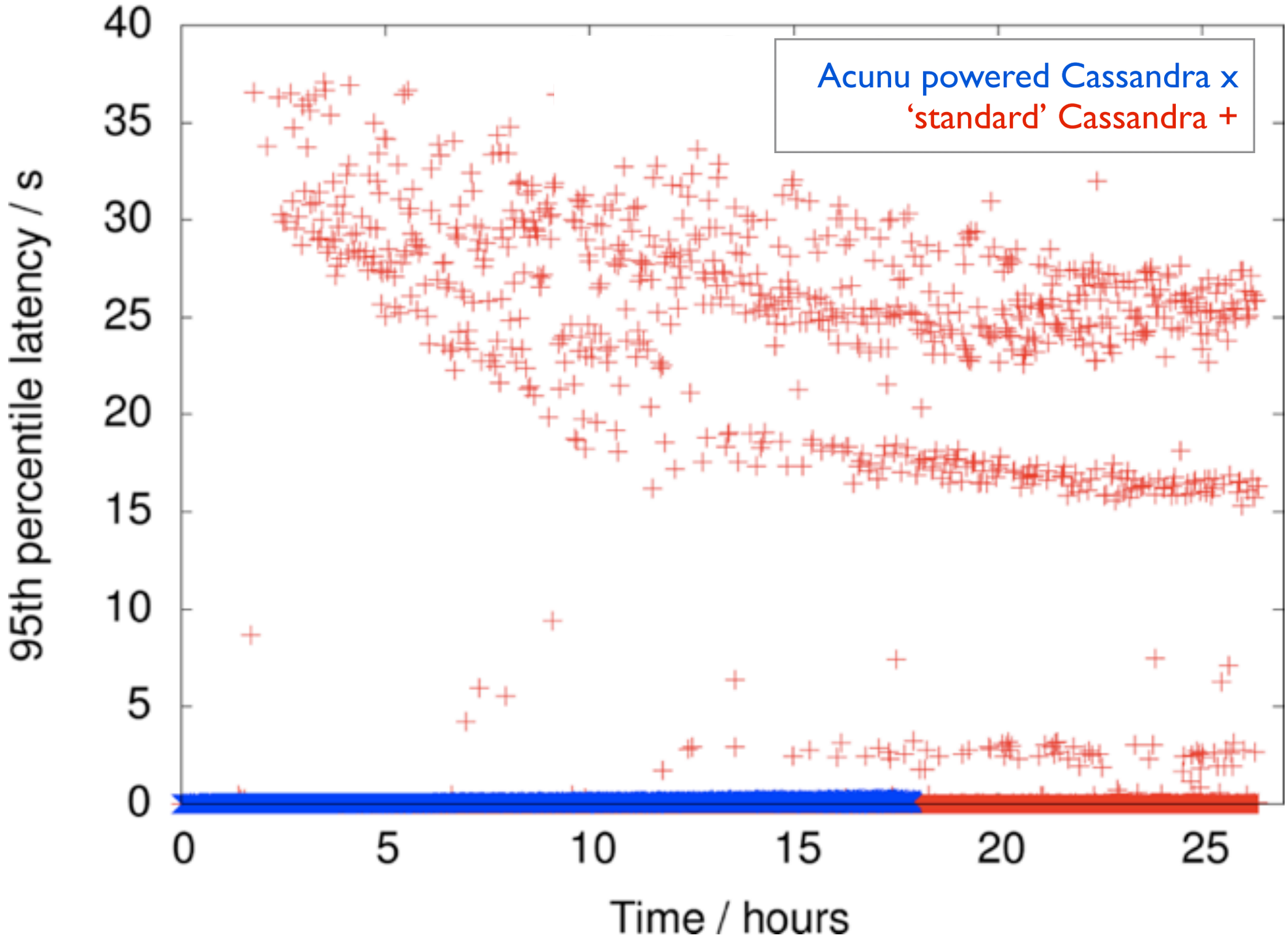
Inserting 3 billion rows



Acunu

Insert latency

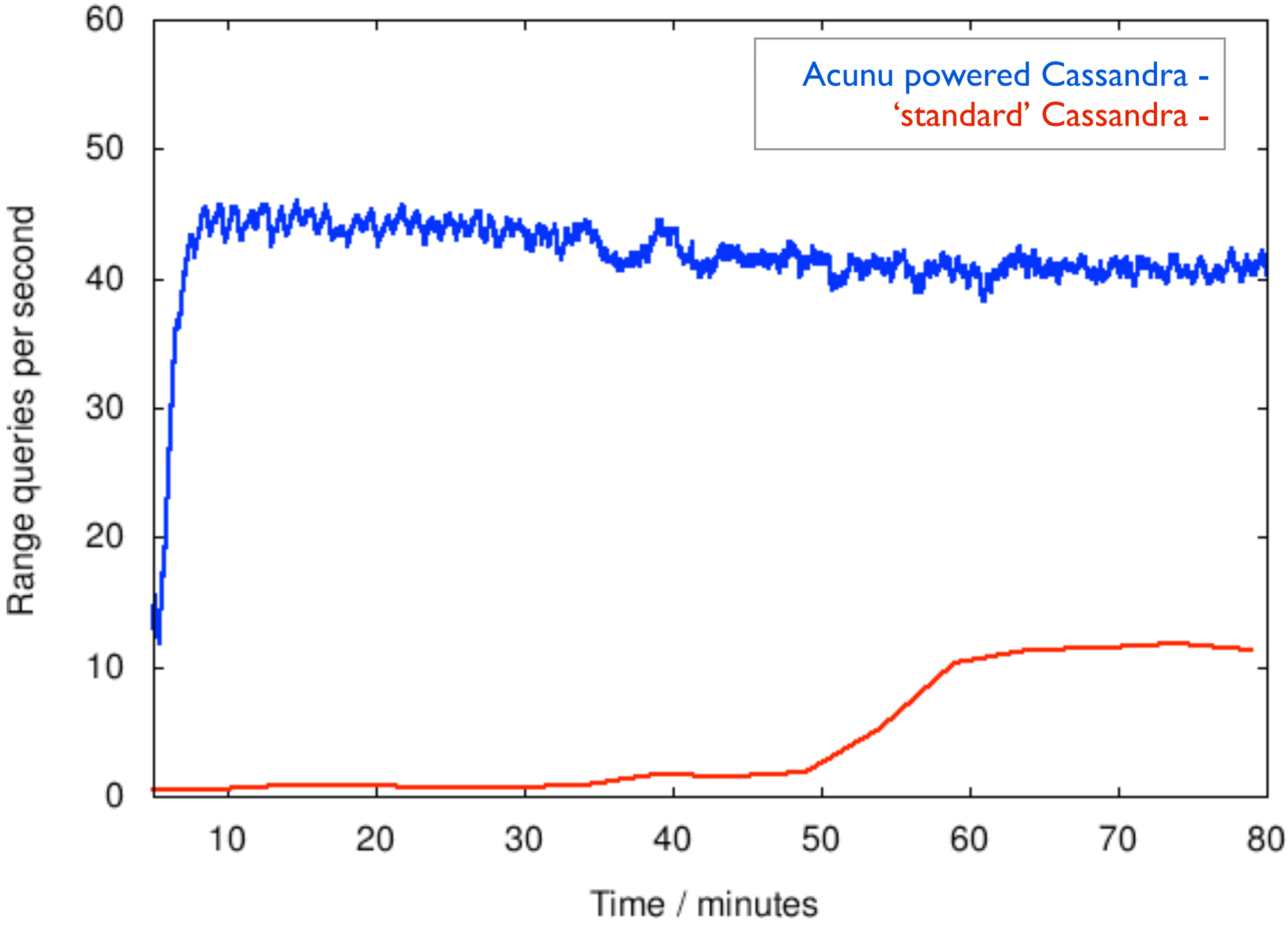
While inserting 3 billion rows



Acunu

Small random range queries

Performed immediately after inserts



Acunu

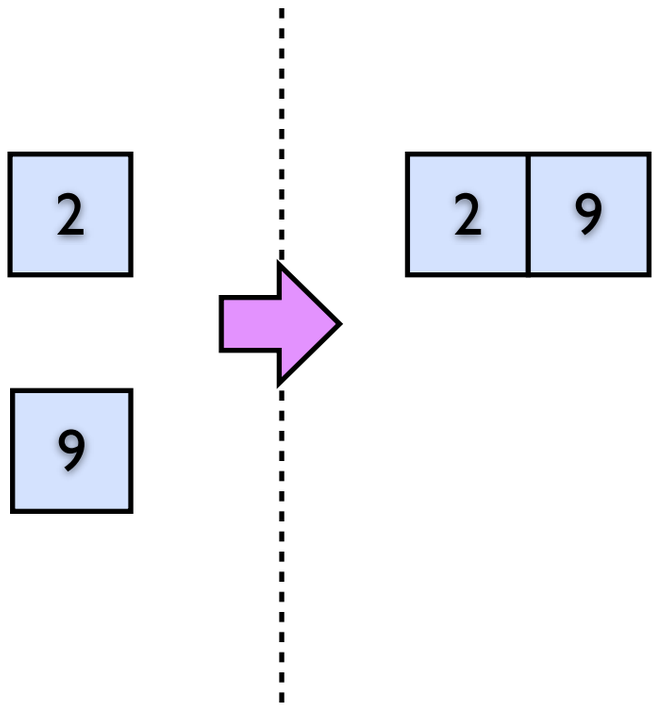
Performance summary

	Standard	**Acunu**	**Benefits**
inserts rate 95% latency	~32k/s ~32s	~45k/s ~0.3s	**>1.4x** **>100x**
gets rate 95% latency	~100/s ~2s	~350/s ~0.5s	**>3.5x** **>4x**
range queries 95% latency	~0.4/s ~15s	~40/s ~2s	**>100x** **>7.5x**

Acunu

Doubling Array

Inserts

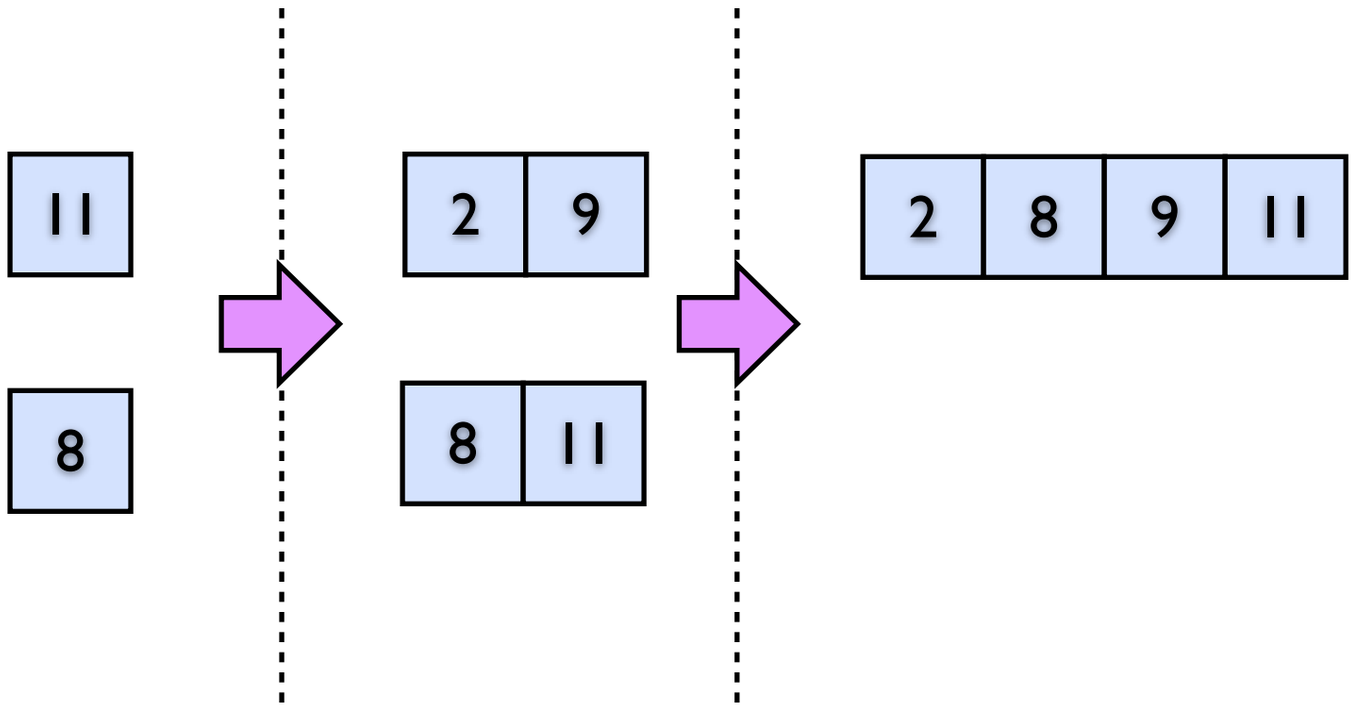


Buffer arrays in memory
until we have > B of them

Acunu

Doubling Array

Inserts



etc...

Similar to log-structured merge trees (LSM), cache-oblivious lookahead array (COLA), ...

Acunu

Demo

https://acunu-videos.s3.amazonaws.com/dajs.html

Acunu

8KB @ 100MB/s, w/ 8ms seek
= **100 IOs/s**

100 / 5
= **20 updates/s**

~ log (2^30)/log 100
= **5 IOs/update**

	Update	**Range Query** (Size Z)
B-Tree	O($\log_B$ N) random IOs	O(Z/B) random IOs
Doubling Array	O((log N)/B) sequential IOs	O(Z/B) sequential IOs

~ log (2^30)/100
= **0.2 IOs/update**

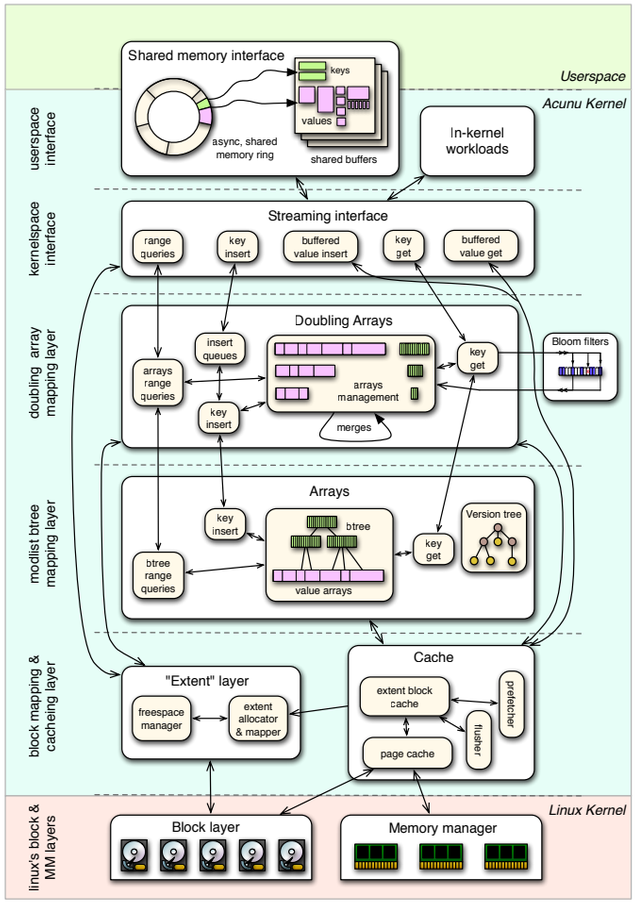
8KB @ 100MB/s
= **13k IOs/s**

13k / 0.2
= **65k updates/s**

B = "block size", say 8KB at 100 bytes/entry ~= 100 entries

Acunu

More



Castle

- Opensource (GPLv2, MIT for user libraries)
- http://bitbucket.org/acunu
- Loadable Kernel Module, targeting CentOS's 2.6.18
- http://www.acunu.com/blogs/andy-twigg/why-acunu-kernel/

Acunu

http://goo.gl/wXNDQ

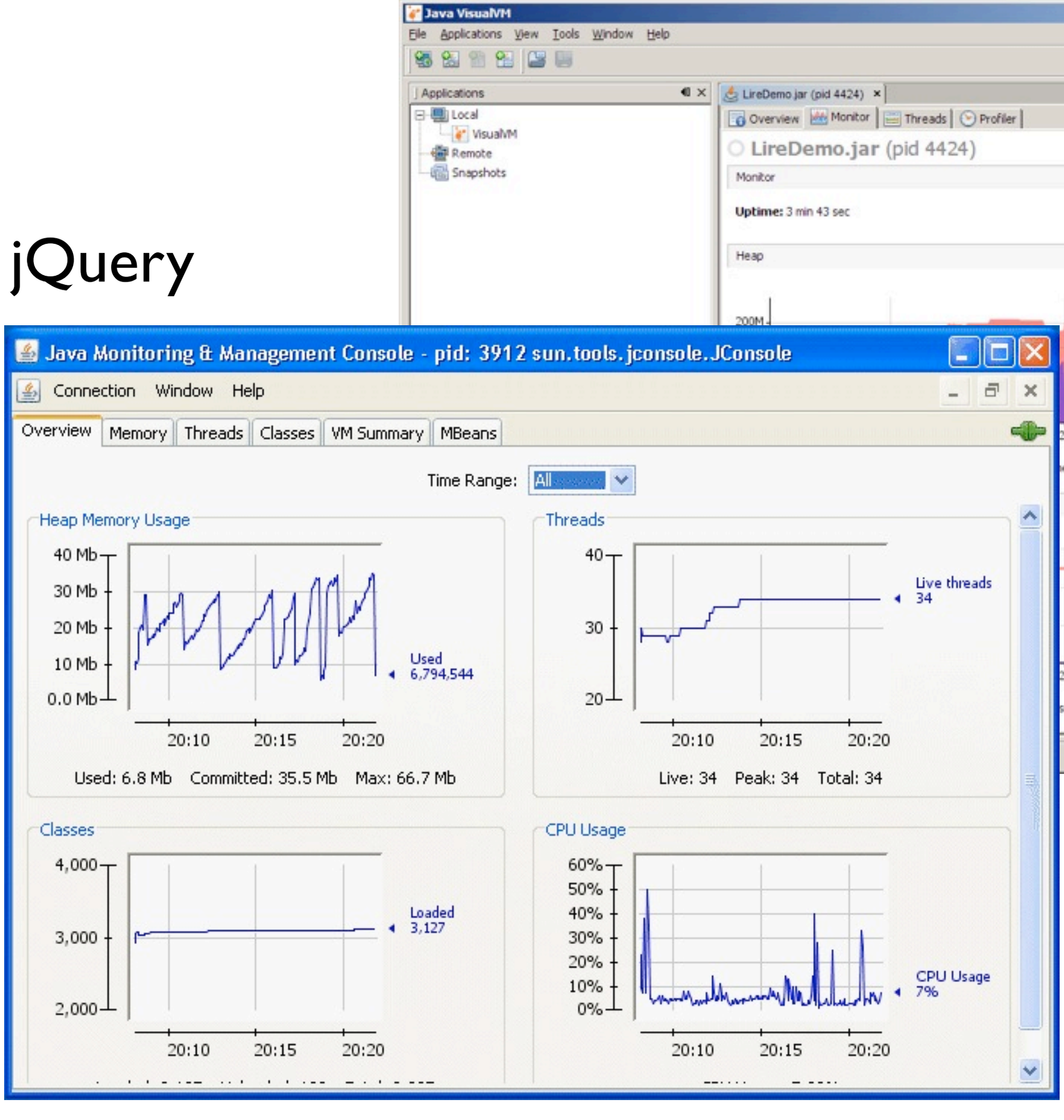
http://goo.gl/gzihe

Acunu

2. Monitoring

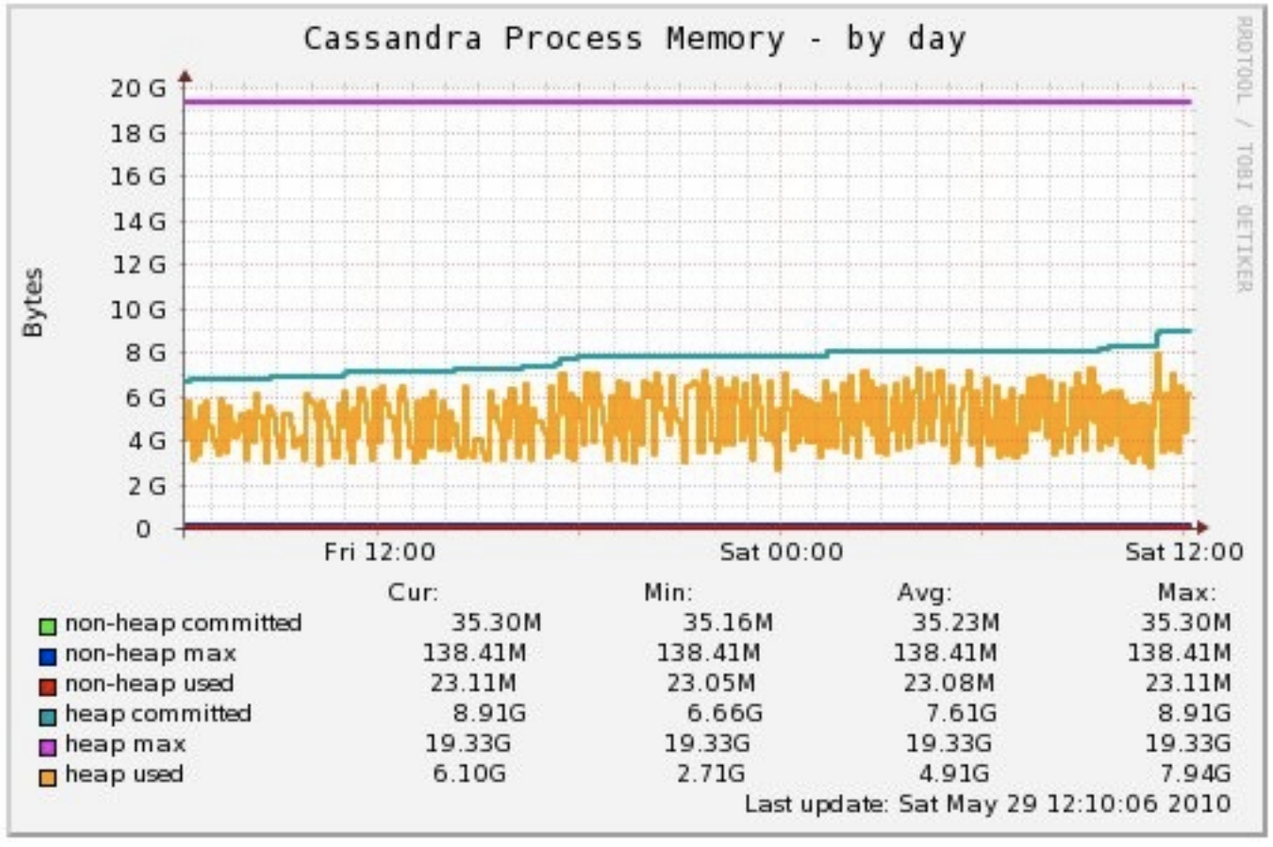
Acunu

jQuery

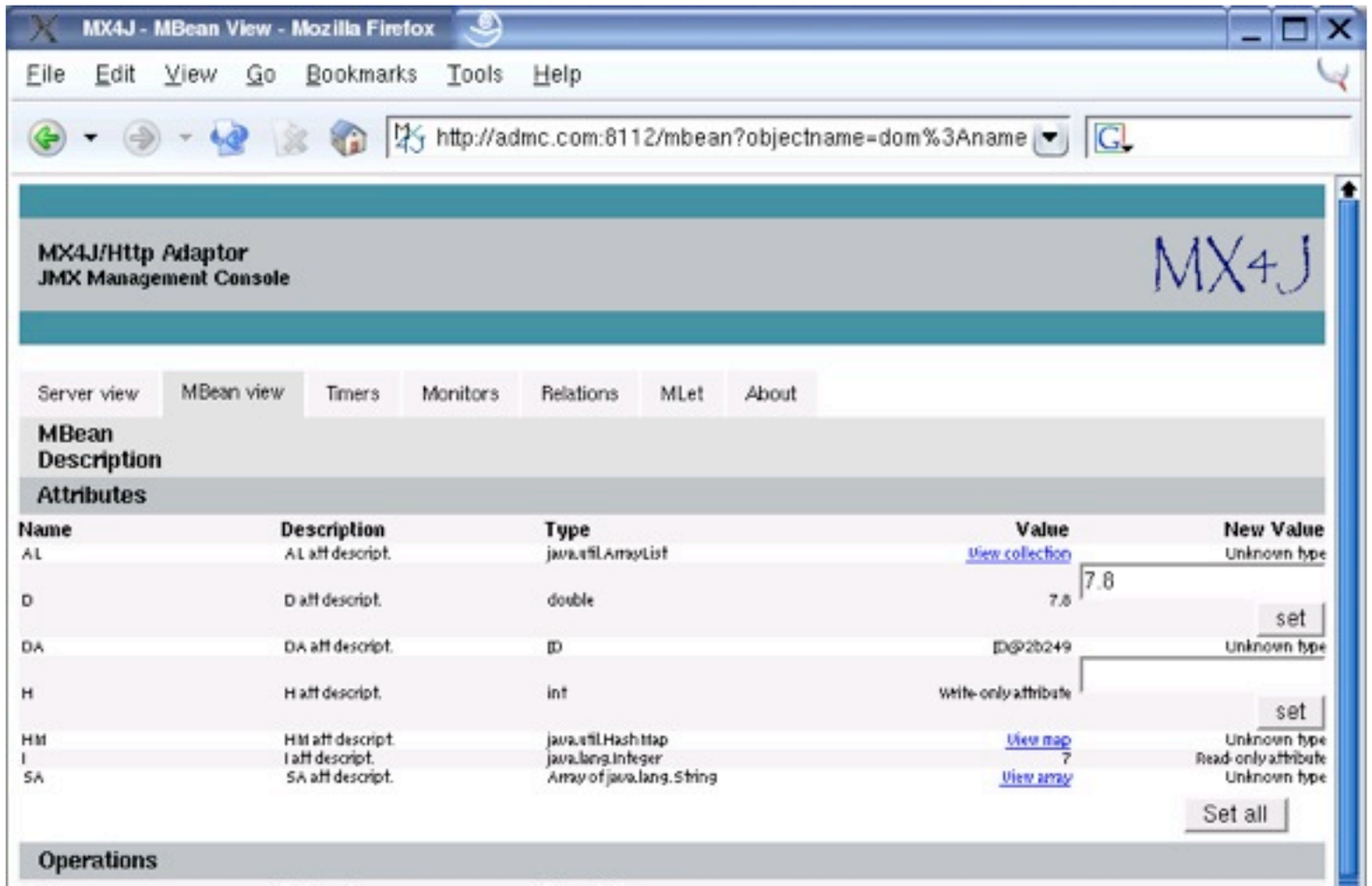


VisualVM

Acunu

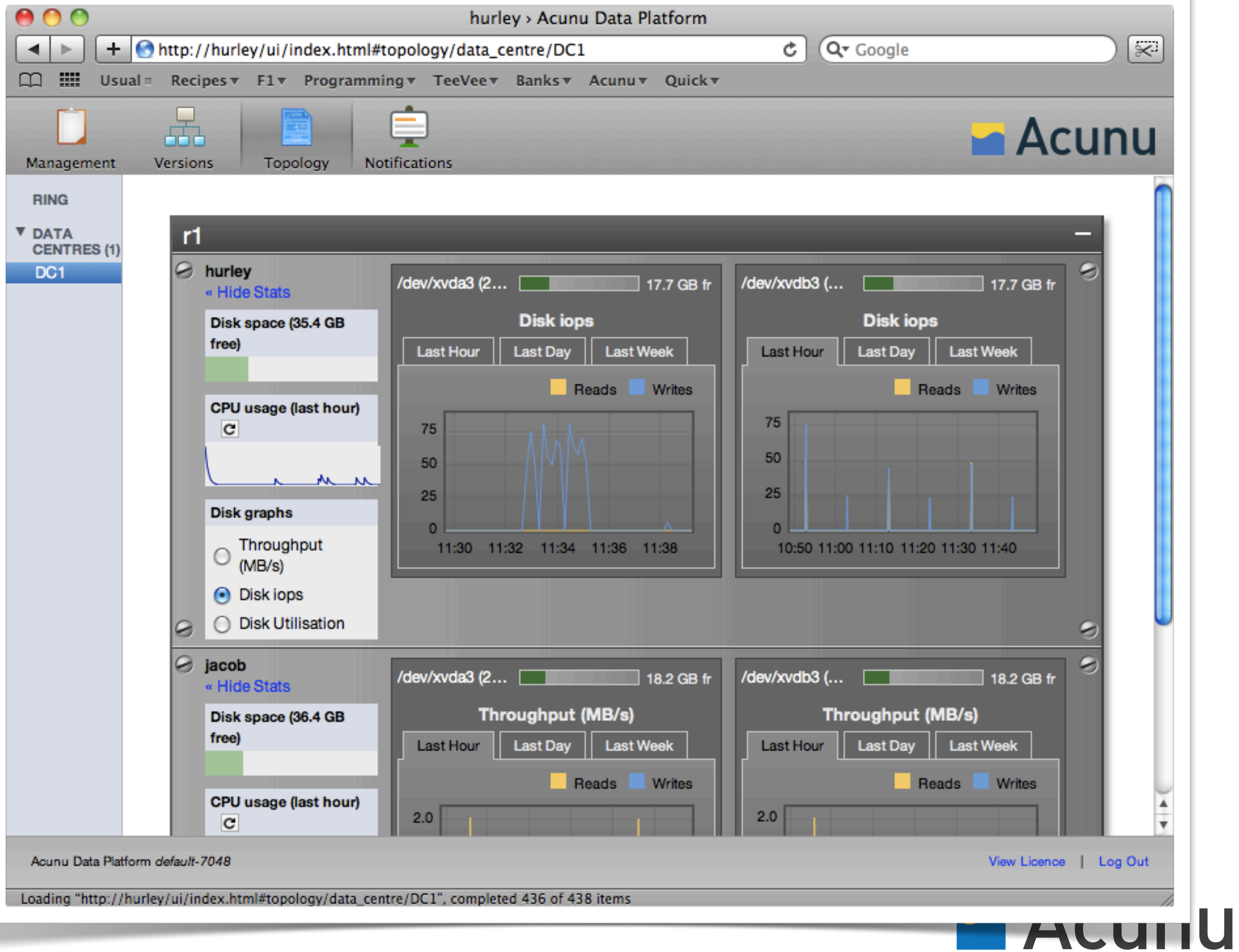


mx4j: Rest-JMX adapter



Munin, Nagios etc

Acunu


hurley › Acunu Data Platform
http://hurley/ui/index.html#topology/data_centre/DC1
Google
Usual
Recipes
F1
Programming
TeeVee
Banks
Acunu
Quick
Management
Versions
Topology
Notifications
Acunu
RING
DATA CENTRES (1)
DC1
r1
hurley
« Hide Stats
Disk space (35.4 GB free)
CPU usage (last hour)
Disk graphs
Throughput (MB/s)
Disk iops
Disk Utilisation
/dev/xvda3 (2...
17.7 GB fr
Disk iops
Last Hour
Last Day
Last Week
Reads
Writes
75
50
25
0
11:30
11:32
11:34
11:36
11:38
/dev/xvdb3 (...
17.7 GB fr
Disk iops
Last Hour
Last Day
Last Week
Reads
Writes
75
50
25
0
10:50
11:00
11:10
11:20
11:30
11:40
jacob
« Hide Stats
Disk space (36.4 GB free)
CPU usage (last hour)
/dev/xvda3 (2...
18.2 GB fr
Throughput (MB/s)
Last Hour
Last Day
Last Week
Reads
Writes
2.0
/dev/xvdb3 (...
18.2 GB fr
Throughput (MB/s)
Last Hour
Last Day
Last Week
Reads
Writes
2.0
Acunu Data Platform default-7048
View Licence
Log Out
Loading "http://hurley/ui/index.html#topology/data_centre/DC1", completed 436 of 438 items

ryan › Acunu Data Platform

ryan/ui/index.html#management/hosts

Management | Versions | Topology | View Licence | Log Out | Acunu

- HOSTS
- COLLECTIONS
- KEYSPACES
- DISKS
- BUCKETS

Hosts

Add | Change Partitioner

Name	IP Address	Cassandra Partitioner	Cassandra Token	CPU usage (last hour)	Cassandra Status	S3 Status
ryan	10.2.129.17	RandomPartitioner	60610040111976956729740652633953450669		✓	✗

ryan › Acunu Data Platform

ryan/ui/index.html#management/column_family/200f9319-fc23-2a92-54b2-75f0dcaad797

Management | Versions | Topology | View Licence | Log Out | Acunu

- HOSTS
  - ryan
- COLLECTIONS
- KEYSPACES
  - hello
    - test
- DISKS
- BUCKETS
- REPORTS

General

Name	test
Comment	
Type	Standard
Column Comparator	BytesType
Subcolumn Comparator	
Collection UUID	44d07349-f498-5f21-6faa-afe7a2f9c9d4

Edit | Clone | Destroy | Take Snapshot

Ops

Last Hour | Last Day | Last Week

Gets | Puts | Big Gets | Big Puts | Range Queries Keys

1.00 / 0.75 / 0.50 / 0.25 / 0.00

08:40 6/9 | 08:50 6/9 | 09:00 6/9 | 09:10 6/9 | 09:20 6/9 | 09:30 6/9

Throughput (MB/s)

Last Hour | Last Day | Last Week

Reads | Writes

1.00 / 0.75 / 0.50 / 0.25 / 0.00

08:40 6/9 | 08:50 6/9 | 09:00 6/9 | 09:10 6/9 | 09:20 6/9 | 09:30 6/9

Latency

Last Hour | Last Day | Last Week

95th Percentile Read Latency | 95th Percentile Write Latency | Read Latency | Write Latency

1.00 / 0.75 / 0.50 / 0.25 / 0.00

08:40 6/9 | 08:50 6/9 | 09:00 6/9 | 09:10 6/9 | 09:20 6/9 | 09:30 6/9

3. Operations

Acunu

```
-bash-3.2$ nodetool
...
Available commands:
  ring                    - Print informations on the token ring
  join                    - Join the ring
  info                    - Print node informations (uptime, load, ...)
  cfstats                 - Print statistics on column families
  version                 - Print cassandra version
  tpstats                 - Print usage statistics of thread pools
  drain                   - Drain the node (stop accepting writes and flush all column families)
  decommission            - Decommission the node
  compactionstats         - Print statistics on compactions
  disablegossip           - Disable gossip (effectively marking the node dead)
  enablegossip            - Reenable gossip
  disablethrift           - Disable thrift server
  enablethrift            - Reenable thrift server
  netstats [host]         - Print network information on provided host (connecting node by defau
  move <new token>        - Move node on the token ring to a new token
  removetoken status|force|<token> - Show status of current token removal, force completion of
  setcompactionthroughput <value_in_mb> - Set the MB/s throughput cap for compaction in the sys
  snapshot [keyspaces...] -t [snapshotName] - Take a snapshot of the specified keyspaces using
  clearsnapshot [keyspaces...] -t [snapshotName] - Remove snapshots for the specified keyspaces
  flush [keyspace] [cfnames] - Flush one or more column family
  repair [keyspace] [cfnames] - Repair one or more column family
  cleanup [keyspace] [cfnames] - Run cleanup on one or more column family
  compact [keyspace] [cfnames] - Force a (major) compaction on one or more column family
  scrub [keyspace] [cfnames] - Scrub (rebuild sstables for) one or more column family
  invalidatekeycache [keyspace] [cfnames] - Invalidate the key cache of one or more column fami
  invalidaterowcache [keyspace] [cfnames] - Invalidate the key cache of one or more column fami
  getcompactionthreshold <keyspace> <cfname> - Print min and max compaction thresholds for a gi
  cfhistograms <keyspace> <cfname> - Print statistic histograms for a given column family
  setcachecapacity <keyspace> <cfname> <keycachecapacity> <rowcachecapacity> - Set the key and
  setcompactionthreshold <keyspace> <cfname> <minthreshold> <maxthreshold> - Set the min and ma
```

Desktop

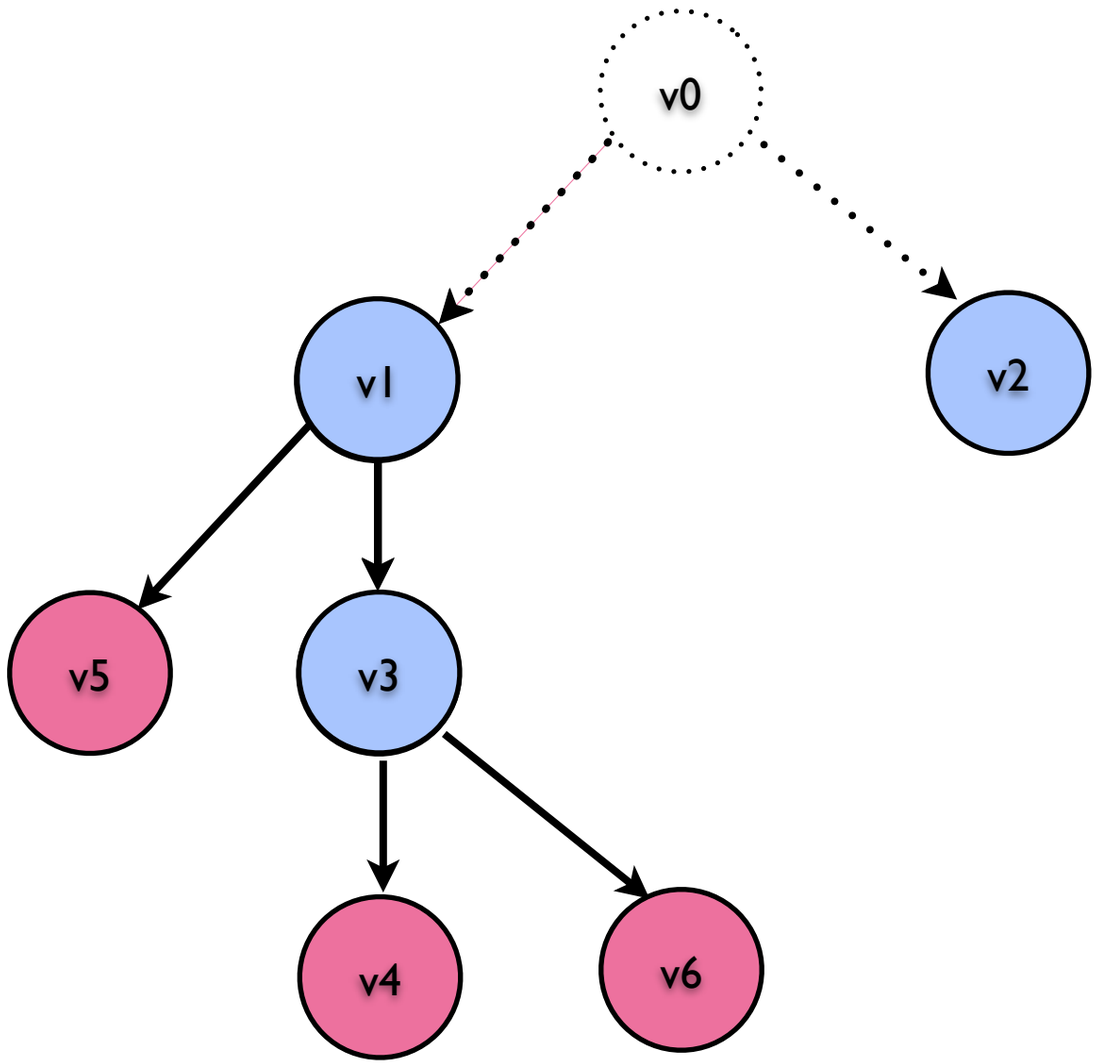
DEVICES
Paul's MacBook
iDisk
Time Machine Backups
SHARED
nexus
e521
viiv-64
PLACES
Desktop
Paul
Documents
Applications
SEARCH FOR

Leopard shots

1 item, 111.18 GB available

SNAPSHOTS*

* And clones!

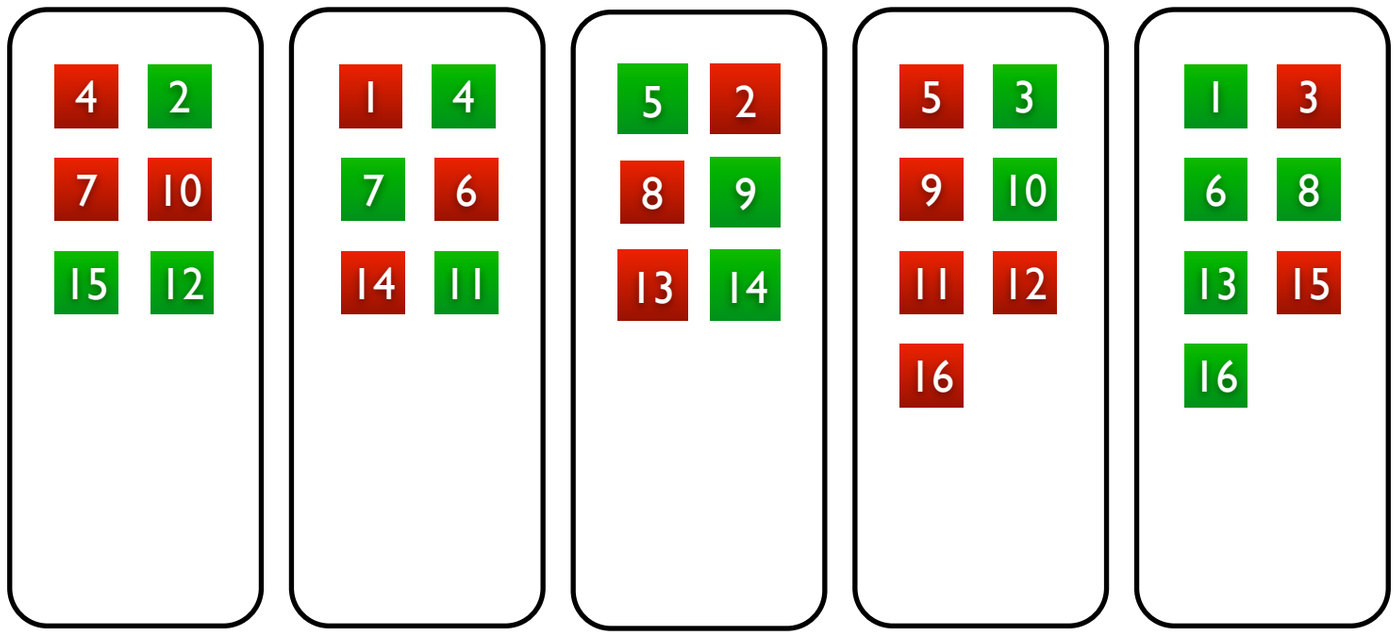


Acunu

Rebuild

Disk Layout: RDA

random duplicate allocation

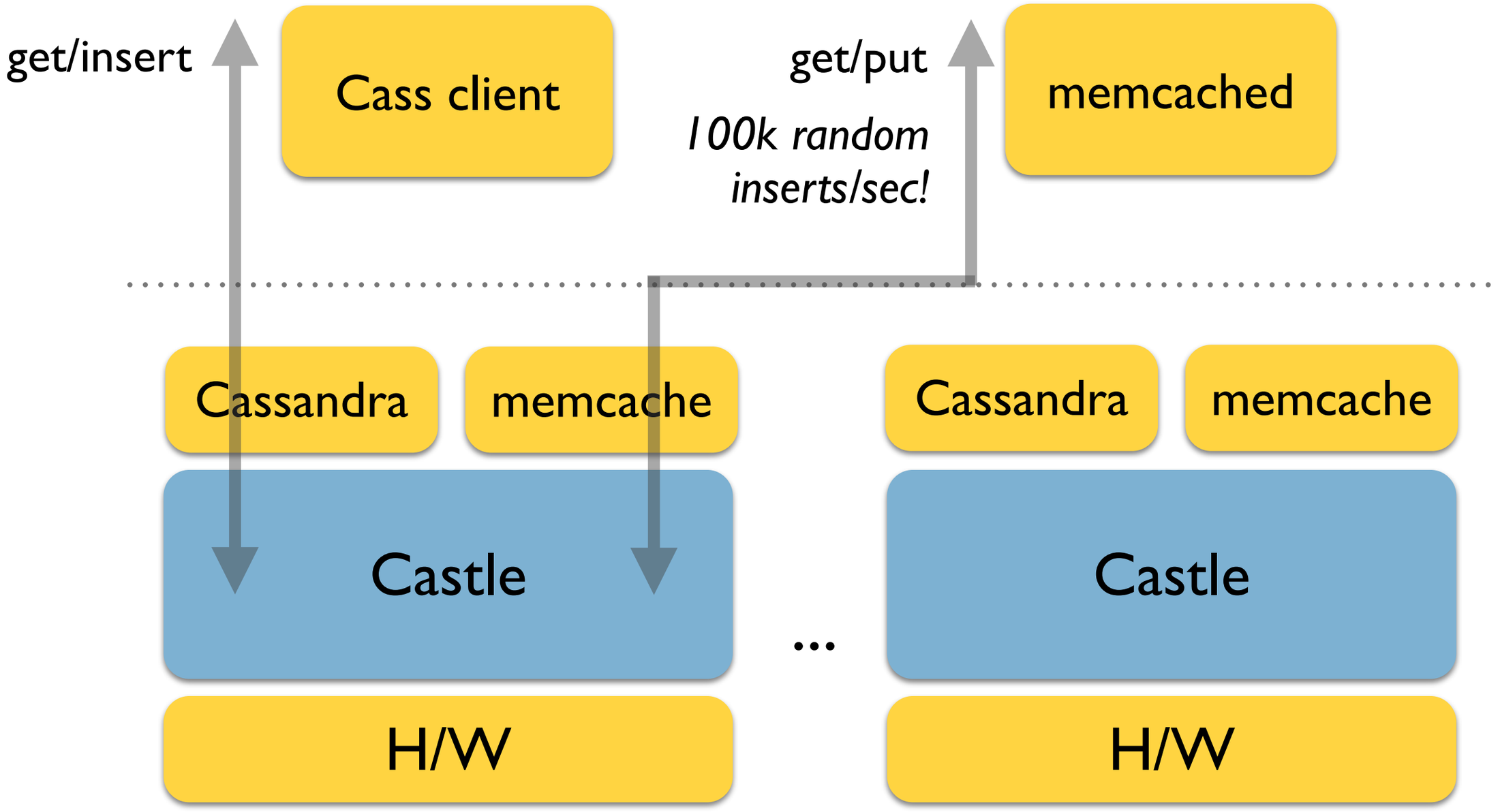


Acunu

Future

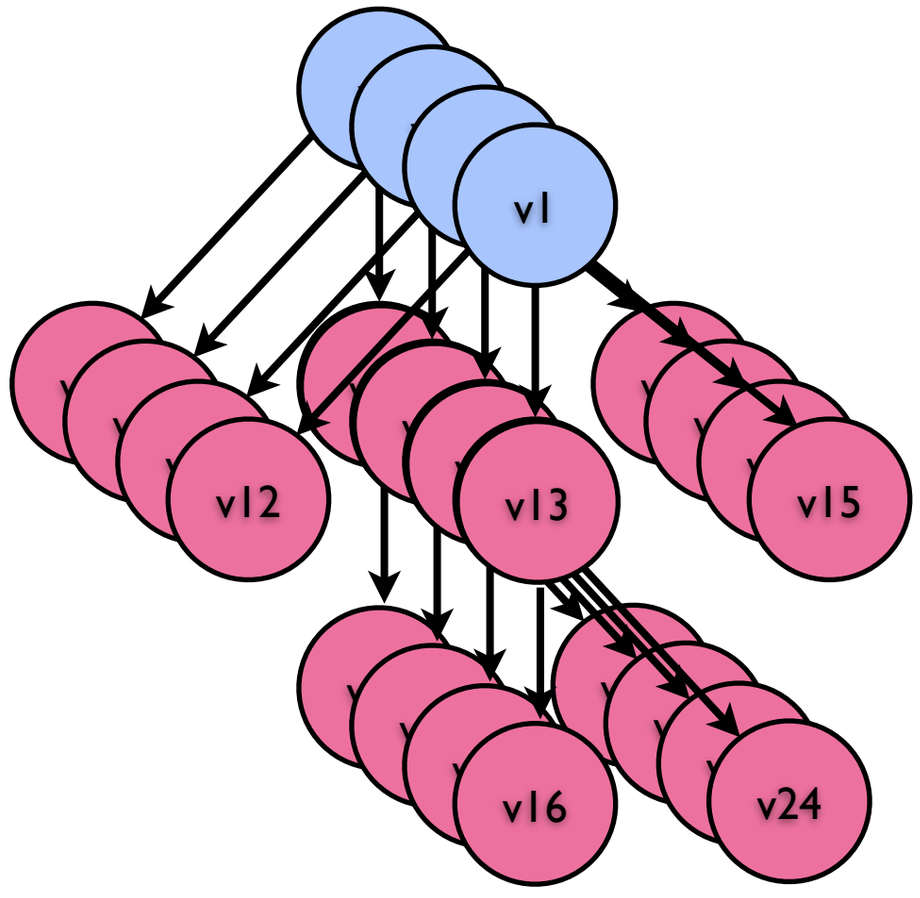
Acunu

Memcache + Cassandra

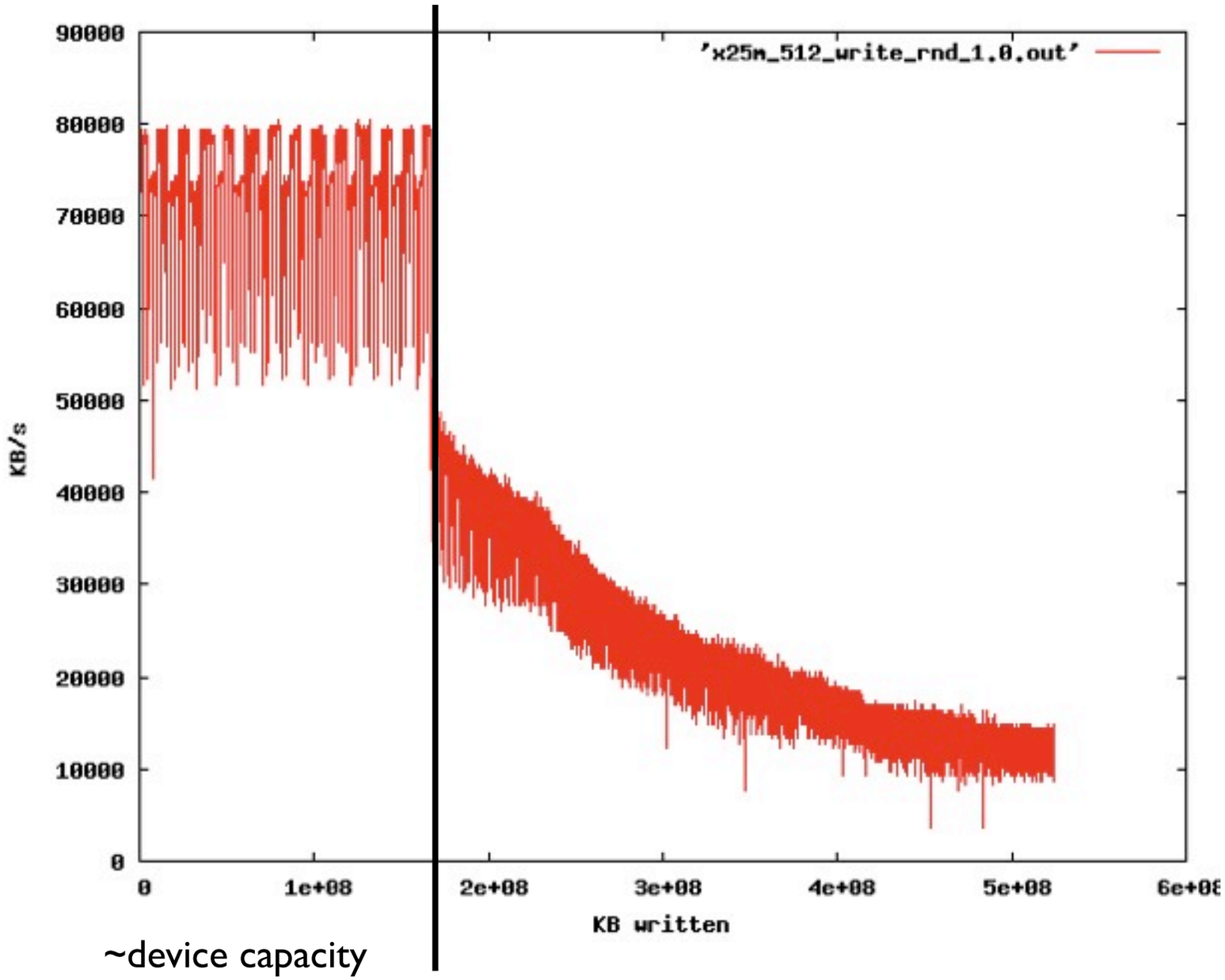


Acunu


SanDisk
SSD
SATA 5000 2.5"
32GB
SERIAL ATA
WARNING!
Contains parts susceptible to damage by Electrostatic Discharge (ESD). Product warranty will be void if label or cover is removed.


v1
v12
v13
v15
v16
v24

Acunu



Beware the “write cliff”...

- Castle: Predictable Performance for Big Data
- Monitoring: distributed, multi-master tools, give you aggregated and summarised view of your cluster
- Snapshots & Clones: addressing real problems with new workloads
- RDA: lightening fast rebuilds for massive disks

Acunu

Questions?

Tom Wilkie
@tom_wilkie
tom@acunu.com

http://bitbucket.org/acunu
http://github.com/acunu

http://www.acunu.com/download
http://www.acunu.com/insights

Acunu


BIG DATA?
No Problem.
Acunu