

AEROSPIKE
NEXTGEN
NOW
SUMMIT '20

Aerospike + Persistent Memory = Voila!

PayPal



Athreya Gopalakrishna

SENIOR ENGINEERING MANAGER
DATABASE CLOUD ENGINEERING
PAYPAL



Sai Devabhaktuni

SENIOR DIRECTOR OF DATABASE
PLATFORM ENGINEERING
PAYPAL

Database Infrastructure & Storage Footprint

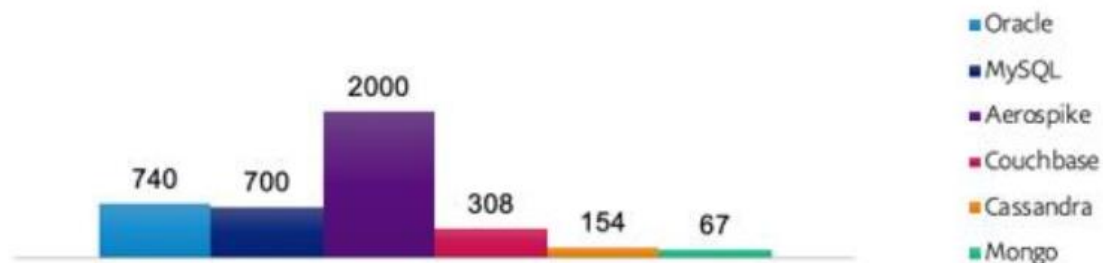
8M+
Execs/Sec

~4000
Database Instances

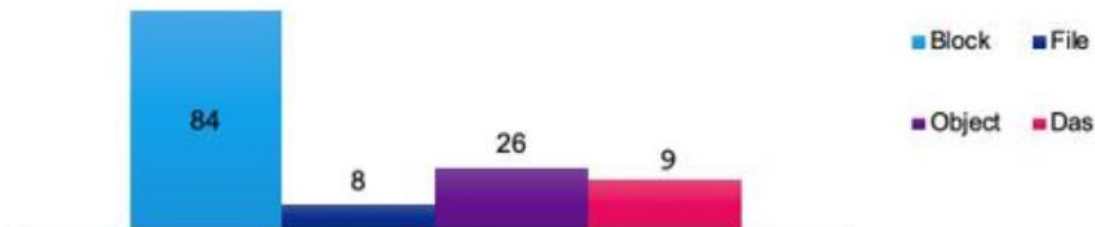
32% Y-o-Y
DB Storage Growth

~100 PB
Total DB Storage

Host Count by Database Type



Storage Footprint (PB) by Type (Utilization)



Executions/Sec range: 2.5M to 6.5M

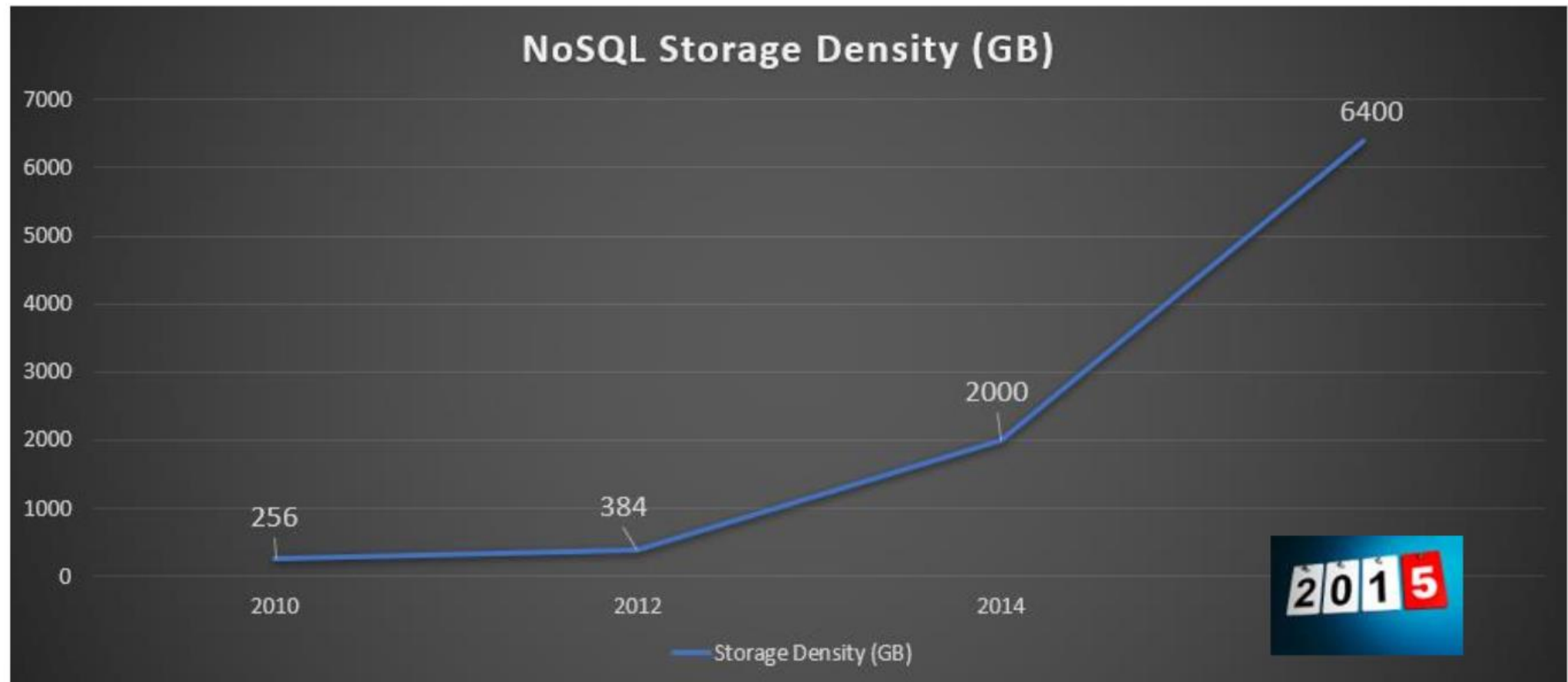
Aerospike TPS



Why Aerospike at PayPal?

- Industry leading in leveraging Hybrid Memory Architecture
- Simplicity – KV datastore
- Extreme performance – Pushing NVMe flash limits
- Higher storage density standards

Evolution 2010 - 2015





Transition from NoSQL Cache to Aerospike

Cost (\$)
↓
3x (\$)

(In-Memory NoSQL)
(50TB)

HP DL360P – 1U



RAM – 384GB/Server



Metric	Cost
Read Latency	~1-2ms
Write Latency	~1ms
Read Throughput	200K
Write Throughput	200K

Total Cost = \$12.5m

Space/Power
↓
8x (Datacenter)



18 Racks
of server 1024

Availability
↑
2-9's to 5-9's

- Inconsistent Performance
- Unstable
- Average Throughput – 100K TPS
- Low latency (~1ms)



Aerospike (Hybrid Memory NoSQL)
(50TB)

<AEROSPIKE>

Dell 630/730– 1U/2U



RAM – 512GB/Server



NVMe/PCIe SSD
(2x3.2TB = 6.4TB)

Metric	Cost
Read Latency (DR: SW)	200 microseconds
Write Latency (DR: SW)	150 microseconds
Peak Read Throughput	940K
Peak Write Throughput	500K
M Size	3M DR RAM 6.4 TB SSD
L Size	10 GB RAM 12.8 TB SSD

Total Cost = \$3.5m



3 Racks
of servers = 120

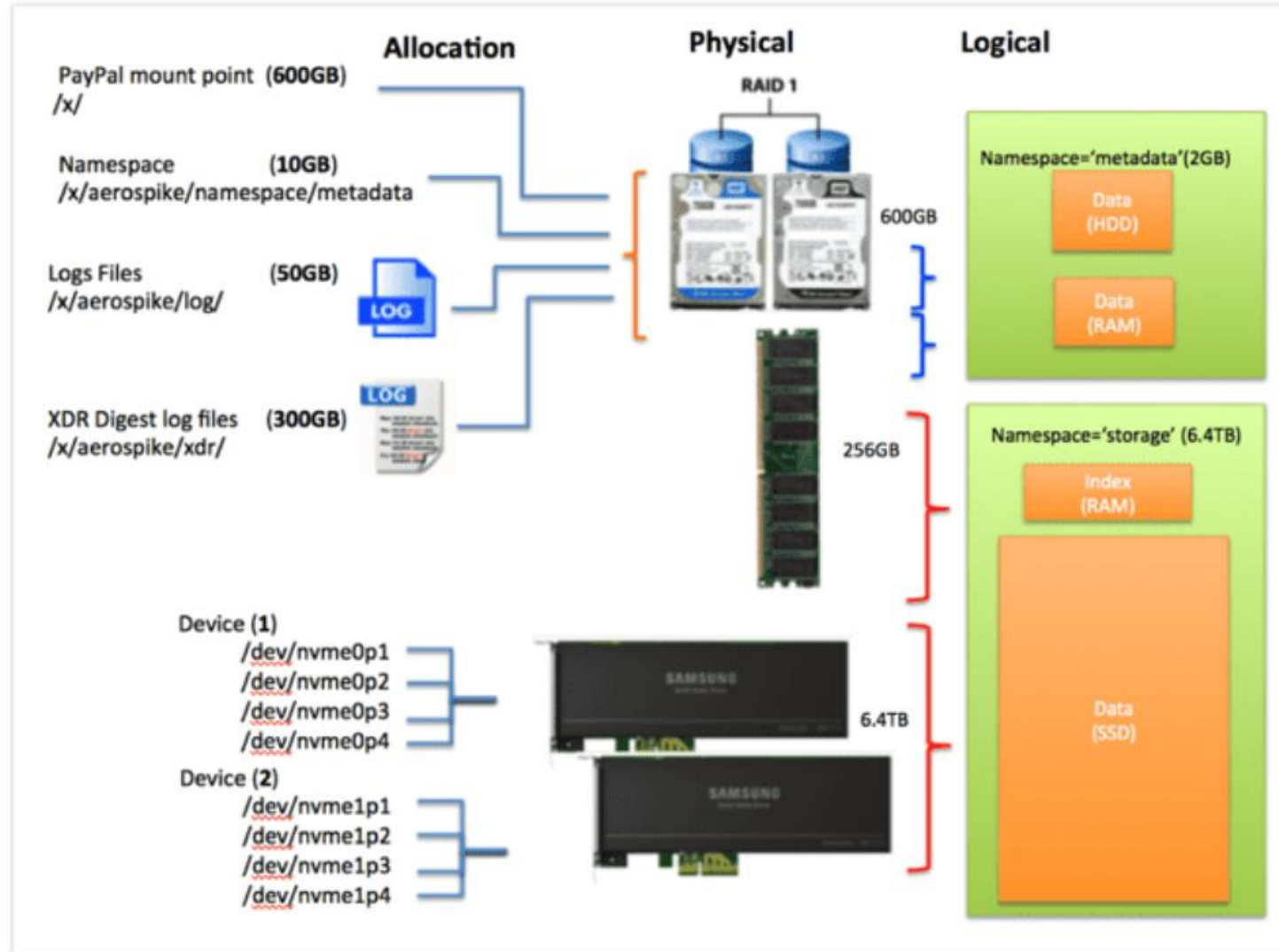
- Consistent Performance
- Stable
- High Throughput – 2M TPS
- Ultra low latency (~200us)

Storage Density Challenge – In Aerospike Hybrid Mode

Storage Density	Re-Index	Full Rebalance
0.5B Keys	~15 minutes	~1+ hour
1B Keys	~30 minutes	~2+ hours
2B Keys	~1 hours	~4+ hours
4B Keys	~2 hours	~8+ hours

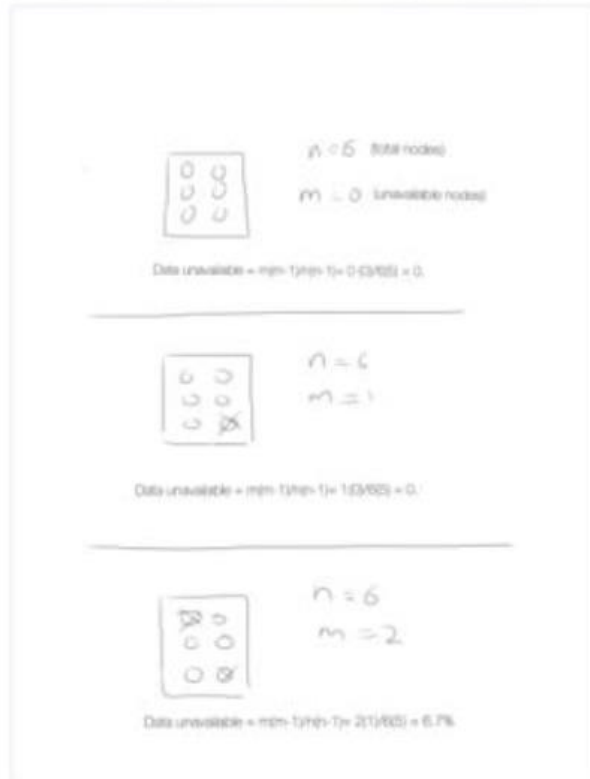


Aerospike 2015 SKU – 256GB RAM, 6.4TB SSD

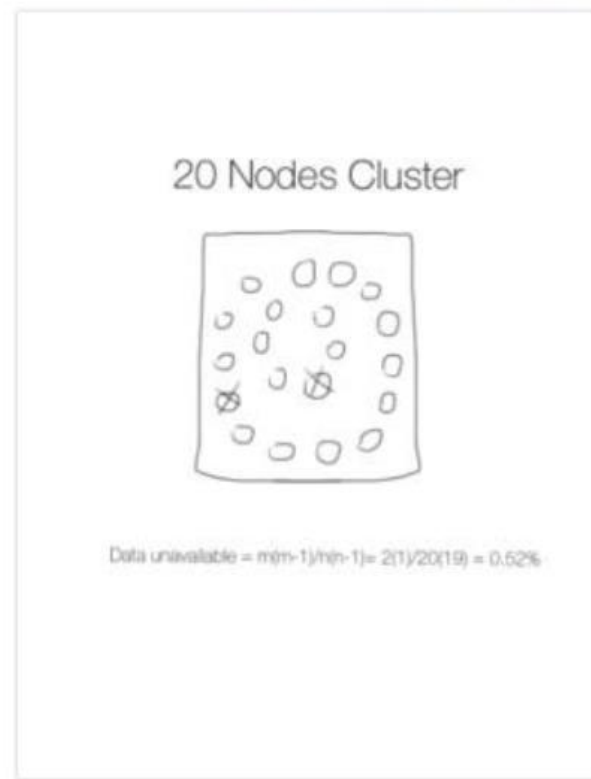


Cluster Designing Process

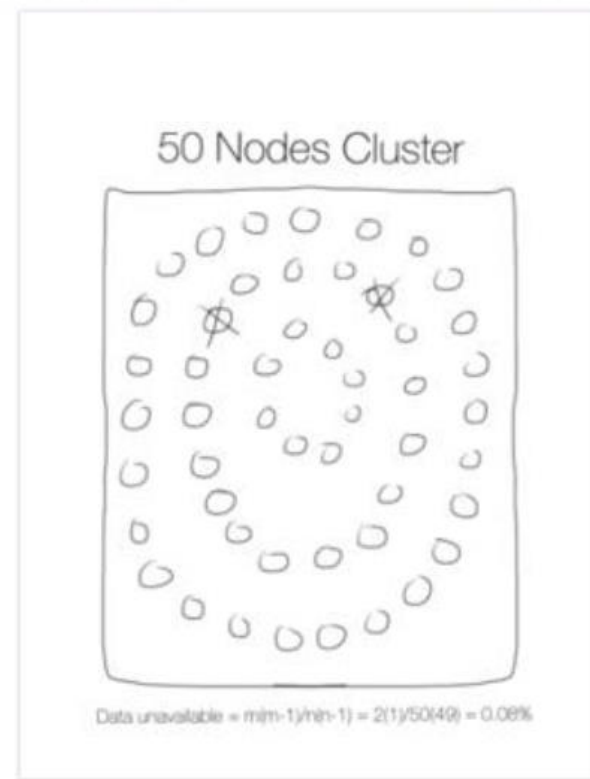
Example: Small Cluster (2 node failure)

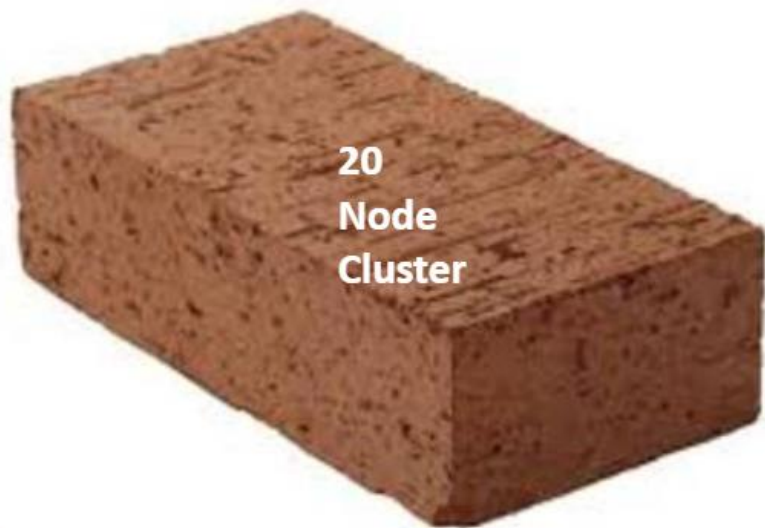
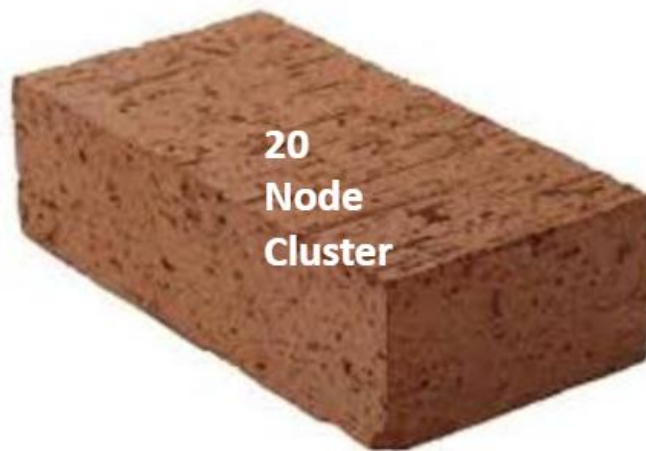
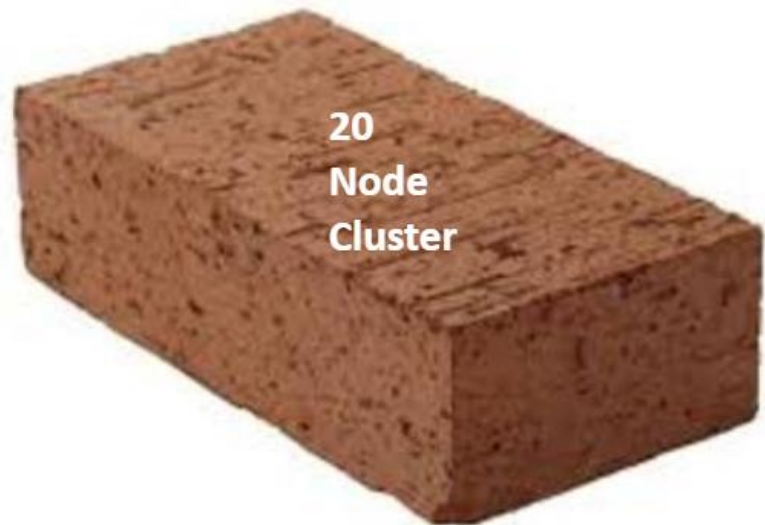


Example: Medium Cluster (2 node failure)



Example: Large Cluster (2 node failure)





Expansion Method



Tortilla



Personal Cheese Pizza



INTEL AND MICRON PRODUCE BREAKTHROUGH MEMORY TECHNOLOGY

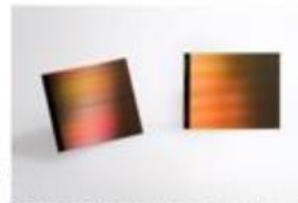
New Class of Memory Unleashes the Performance of PCs, Data Centers and More

NEWS HIGHLIGHTS

- Intel and Micron begin production on new class of non-volatile memory, creating the first new memory category in more than 25 years.
- New 3D XPoint™ technology brings non-volatile memory speeds up to 1,000 times faster¹ than NAND, the most popular non-volatile memory in the marketplace today.
- The companies invented unique material compounds and a cross point architecture for a memory technology that is 10 times denser than conventional memory².
- New technology makes new innovations possible in applications ranging from machine learning to real-time tracking of diseases and immersive 8K gaming.

SANTA CLARA, Calif., and BOISE, Idaho, July 28, 2015 – Intel Corporation and Micron Technology, Inc. today unveiled 3D XPoint™ technology, a non-volatile memory that has the potential to revolutionize any device, application or service that benefits from fast access to large sets of data. Now in production, 3D XPoint technology is a major breakthrough in memory process technology and the first new memory category since the introduction of NAND flash in 1989.

The explosion of connected devices and digital services is generating massive amounts of new data. To make this data useful, it must be stored and analyzed very quickly, creating challenges for service providers and system



3D Xpoint™ technology is up to 1000x faster than NAND and an individual die can store 128Gb of data



News In the Press



April 2019

TECHSPOT

TRENDINGFEATURESREVIEWS
THE BEST
DOWNLOADS
PRODUCT FINDER
FORUMS

By Eric Hamilton on April 2, 2019, 5:37 PM | 17 comments

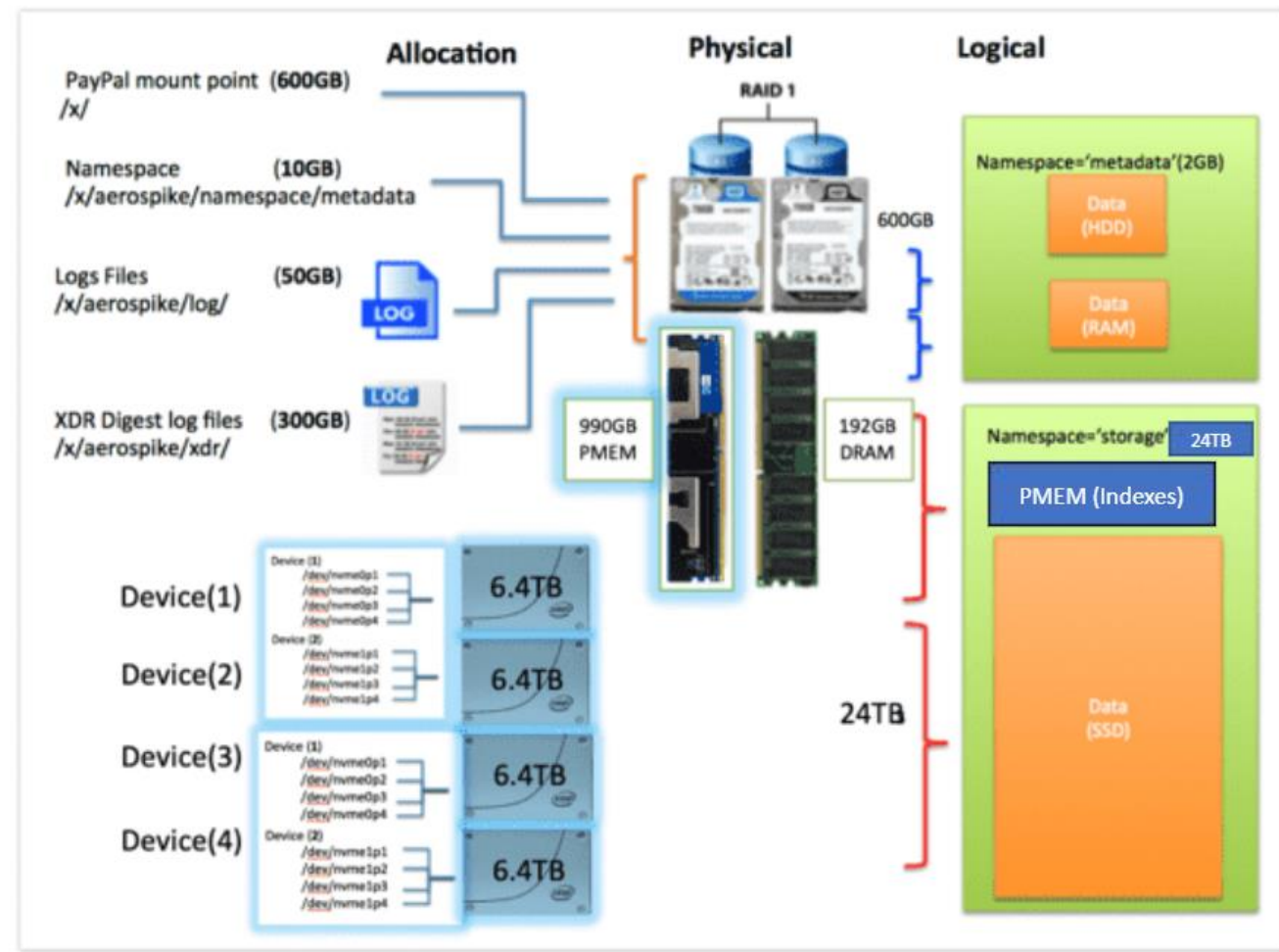


Why it matters: Intel's long awaited Optane DC Persistent Memory DIMMs have arrived and they could greatly transform the way servers and data centers handle data sets. The new DIMMs will see Intel attempting to bridge the price and performance gap between DRAM and NAND and the new class of memory opens up several unique use cases.

During Intel's [Data-Centric Innovation Day](#), the company announced the highly anticipated Optane DC Persistent Memory DIMMs. While the new [Cascade Lake Xeons](#) may be the star of the show, Intel's Optane is a key product in propelling the company forward to its data-centric future for which they've come up with a new slogan "Moving, Storing and Processing Data."

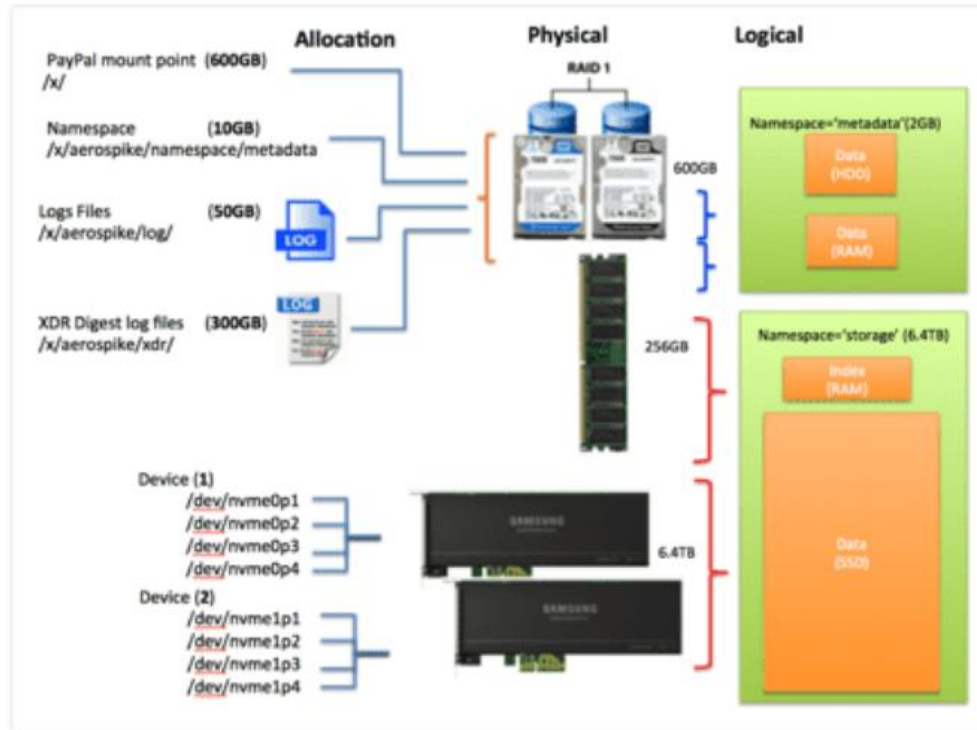
The new Optane DIMMs fit under the "store" designation as Intel is looking to leverage them to store more data affordably, while also potentially disrupting the memory market. Intel's Optane DIMMs will use 3D XPoint memory, a type of non-volatile memory that is something of an amalgamation of NAND and DRAM.

New Aerospike PMEM SKU – May 2019





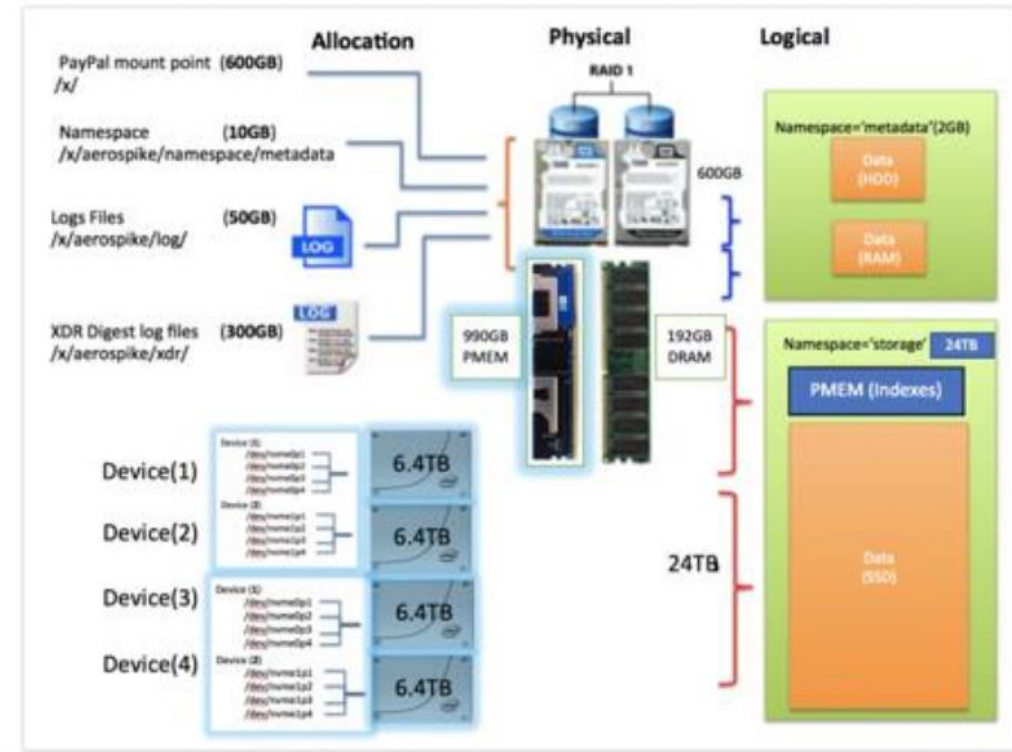
256GB RAM, 6.4TB NVMe



Aerospike SD



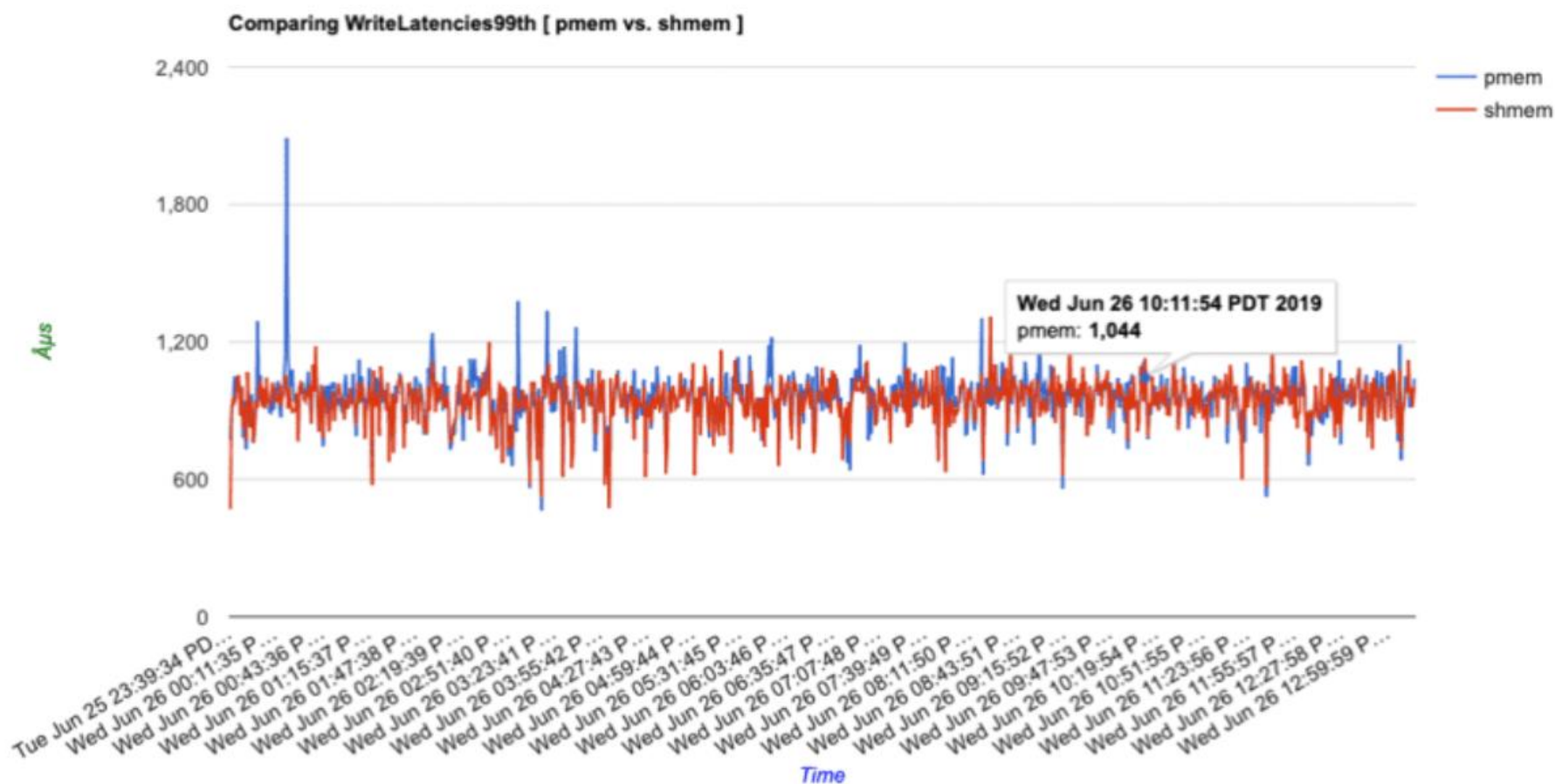
192GB RAM, 996GB PMEM, 24TB NVMe



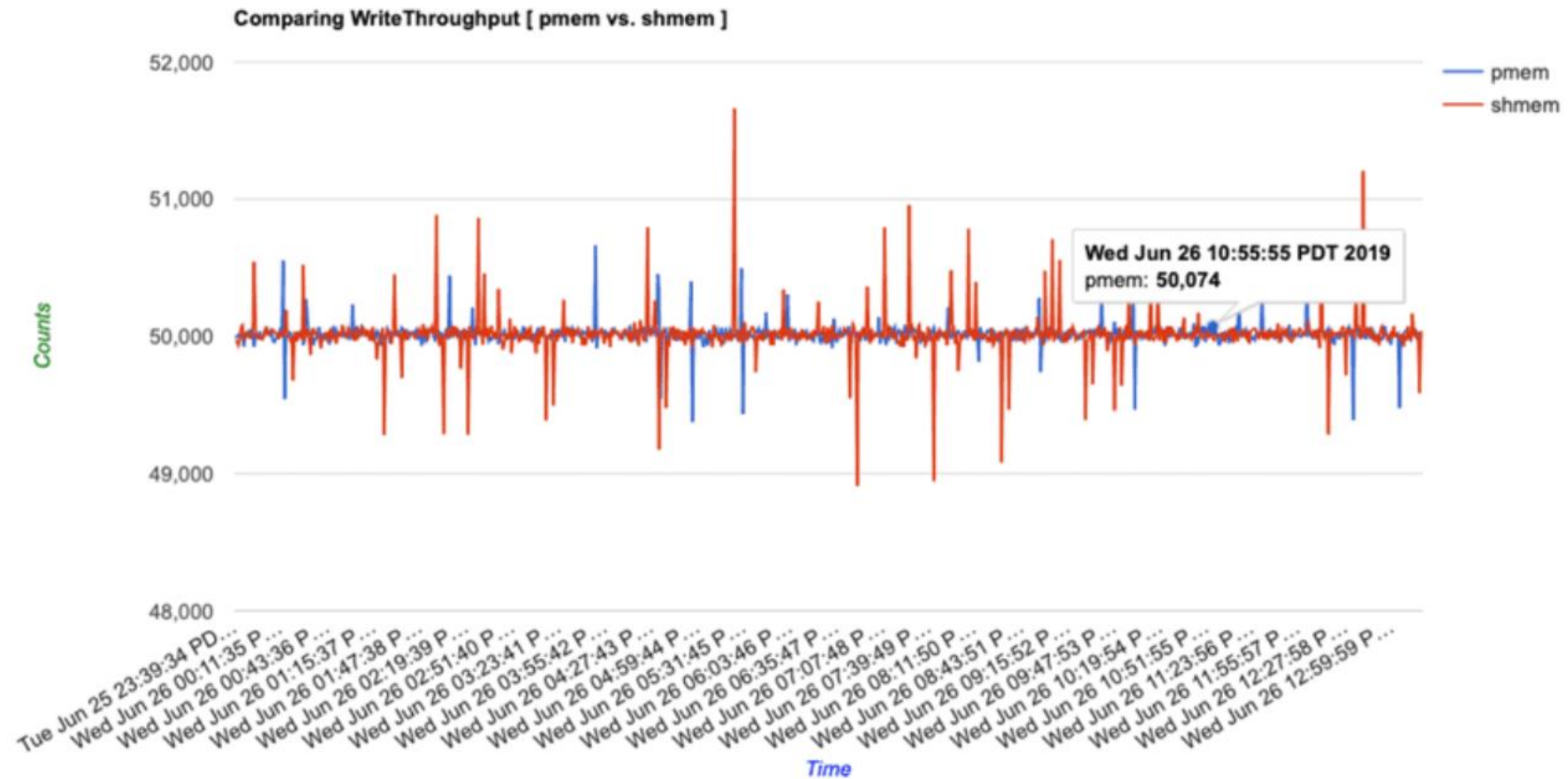
VS

Aerospike HD

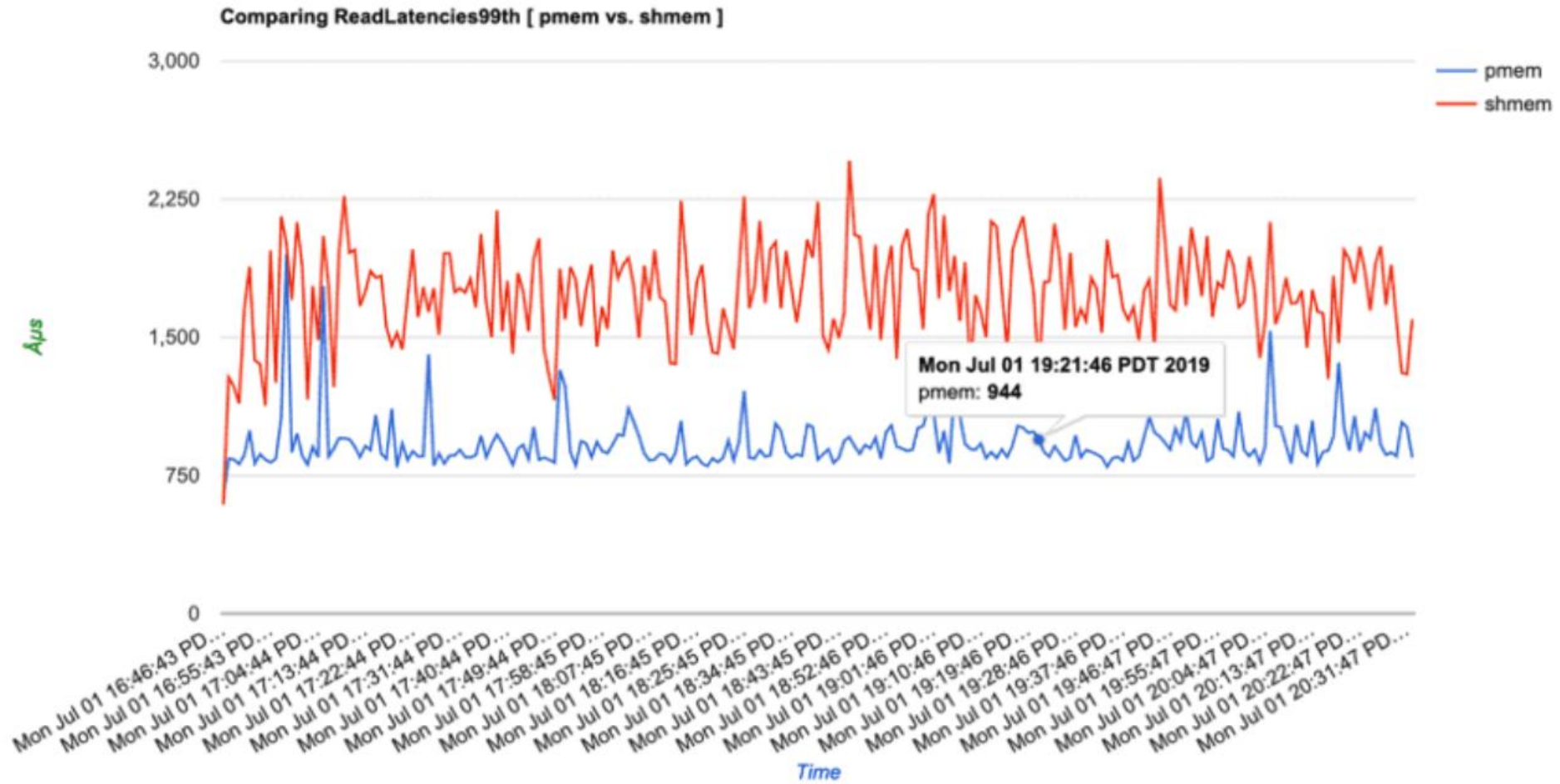
Write latency 99th percentile



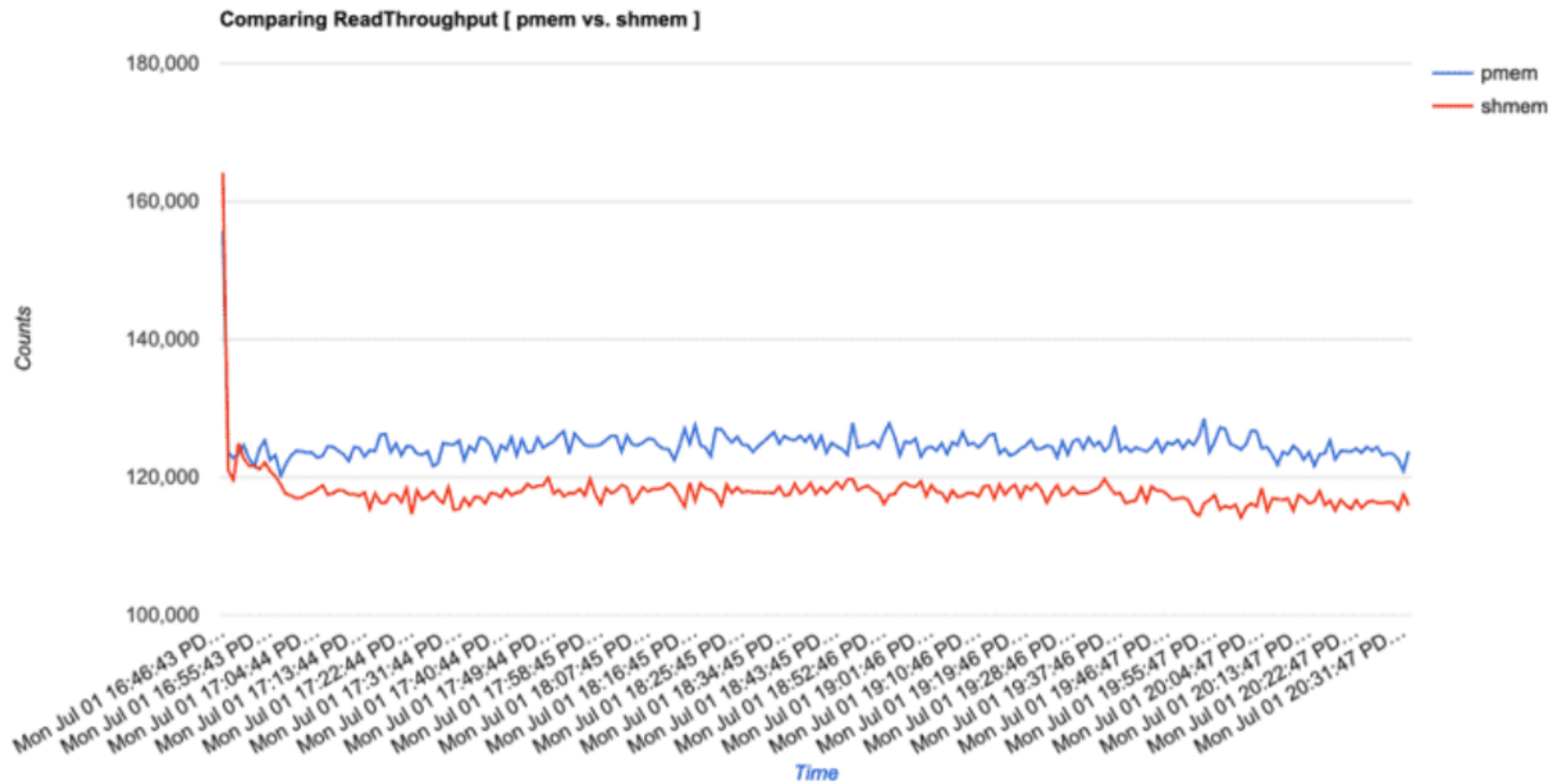
Write throughput



Read Latency 99th Percentile



Read Throughput





Large
Cheese
Pizza



Rationalizing AerospikeHD Usage – 4 Months

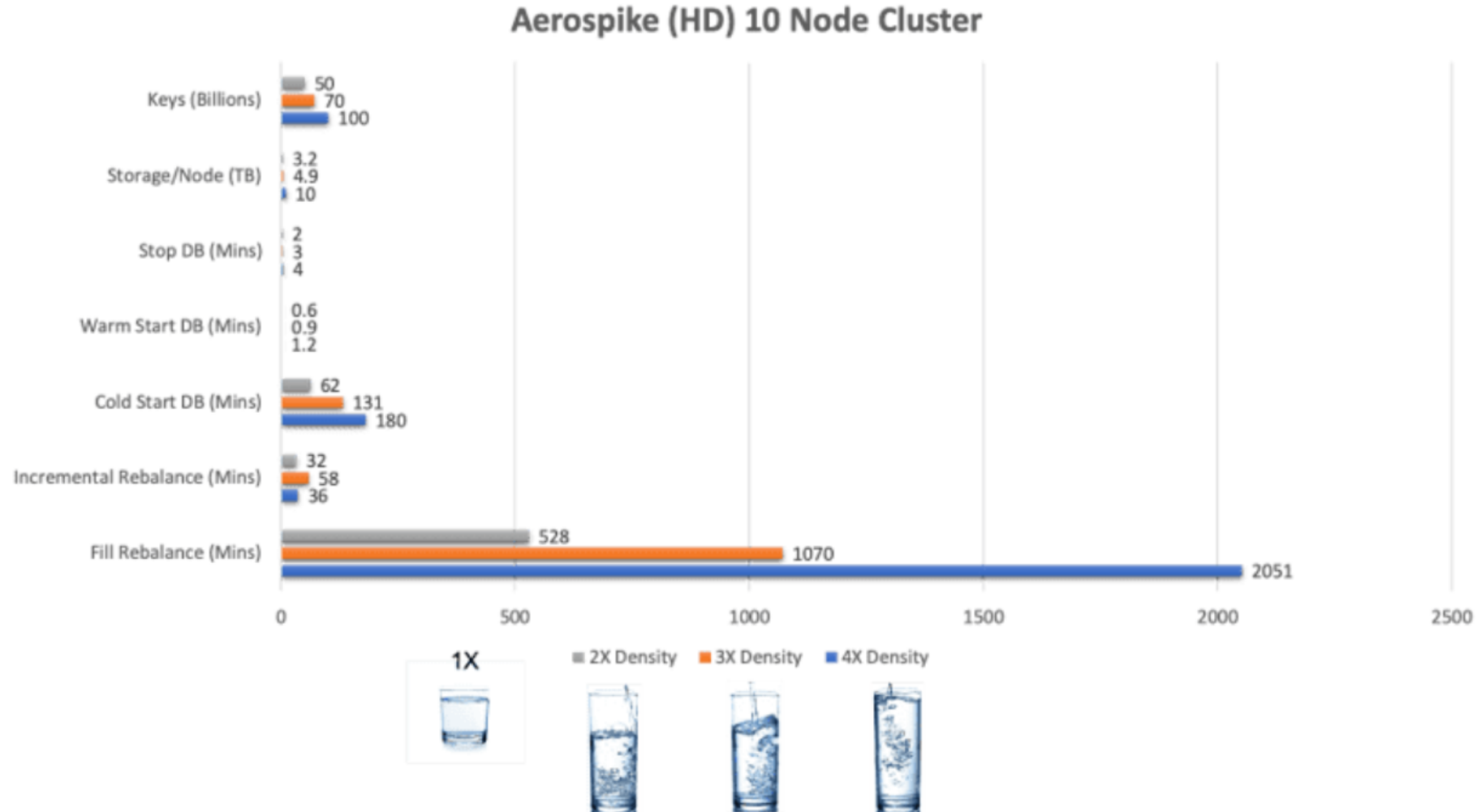
Per Cluster:

1x = 25B Keys

2x = 50B Keys

3x = 70B Keys

4x = 100B Keys



Comparison Evaluation

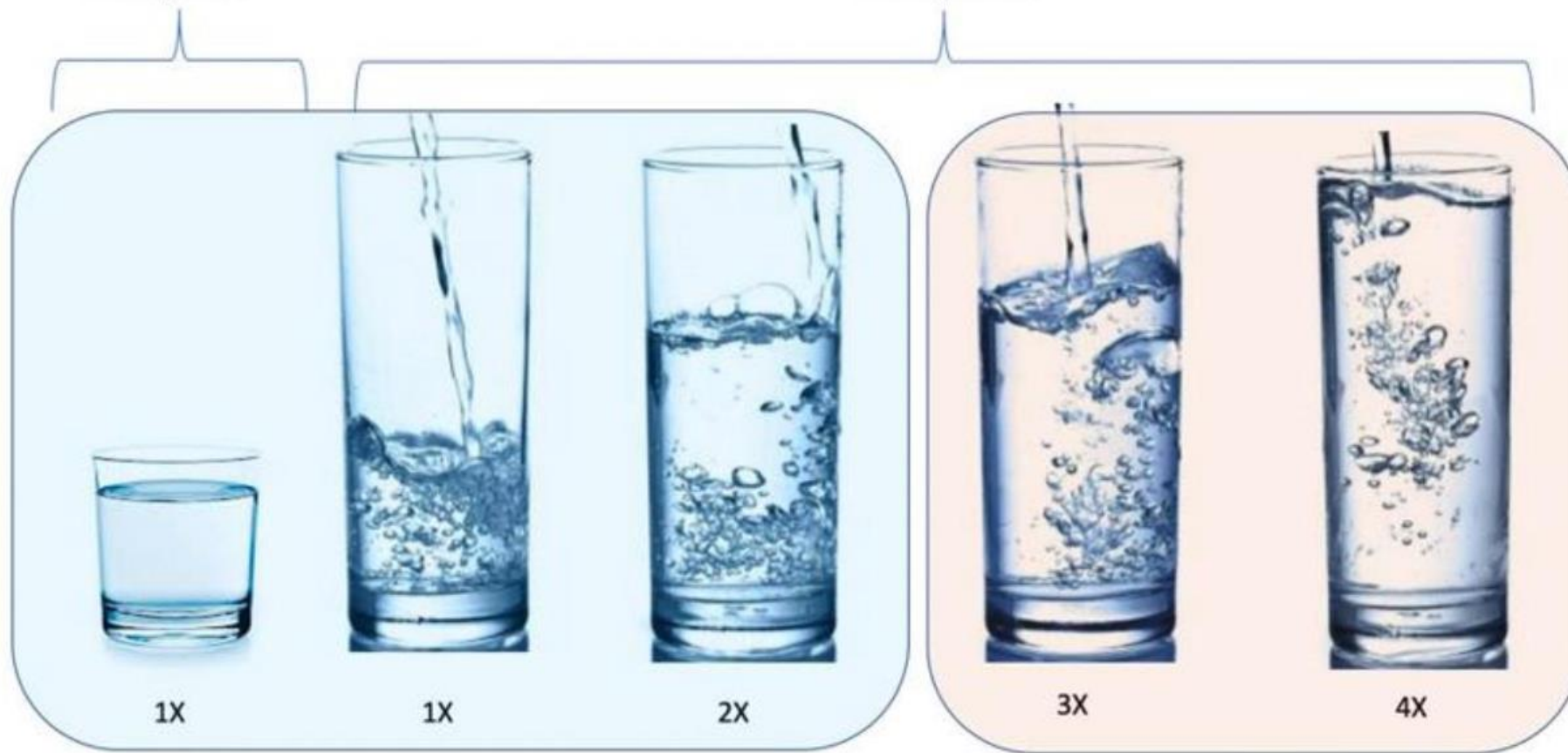
	AerospikeSD	AerospikeHD	Improvement
Max Keys/Node	~2 Billion	~10 Billion	5x
Max Usable Storage/Node	3.2 TB	12.8 TB	4x
Nodes/Cluster	20	10	2x
Usable Keys/Cluster	21B	21B	2x
~Cost/ClusterGroup	~\$1.3M	~\$0.9M	30+% drop in price
Reindexing Time	59 minutes	4 minutes	12x
Shutdown Time	1 minute	4 minutes	Loss of 3 minutes
Reboot+Reindex Time	1 hour	8 minutes	8x
Replication factor	2	3	NA
Rack/Power/Space	20U	10U	2x
Clusters/Rack	2	4	2x
Failure tolerance	1 node	2 nodes	Yes
Rolling software upgrade	10+ hours	~5 hours	Yes
Rolling OS upgrade	10+ hours	~6 hours	Yes
Cluster Creation	11 minutes	11 minutes	No



Usage Strategy

AerospikeSD

AerospikeHD

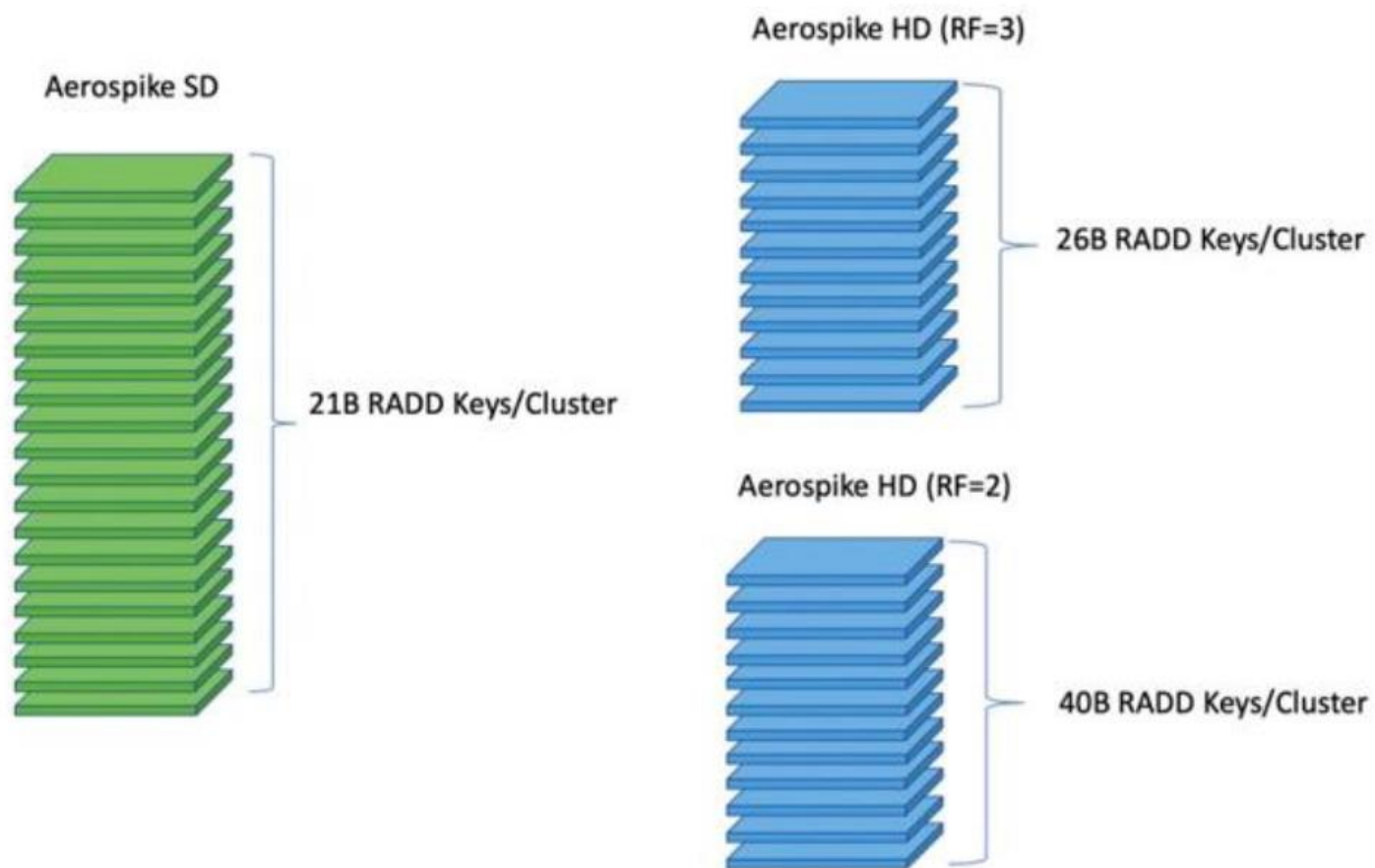


Per Node:
1x = 2B Keys
2x = 4B Keys
3x = 6B Keys
4x = 8B Keys

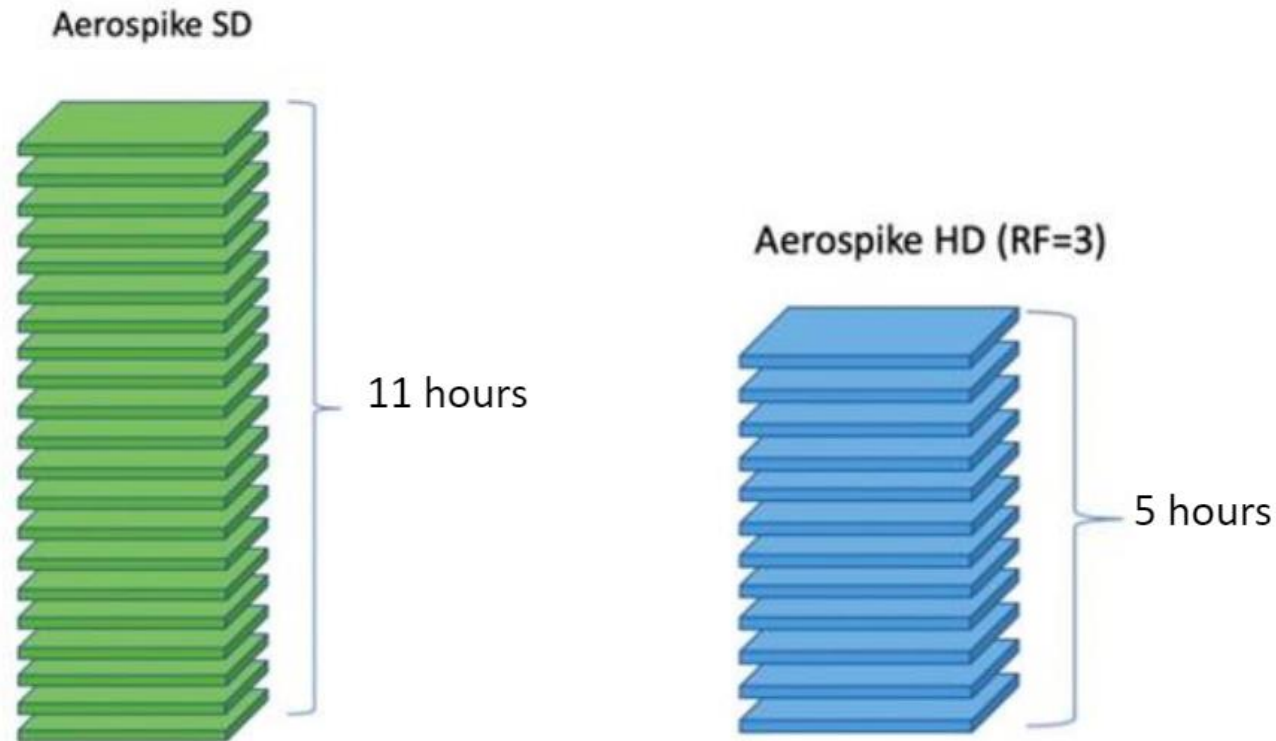
Online Analytics

Offline Analytics

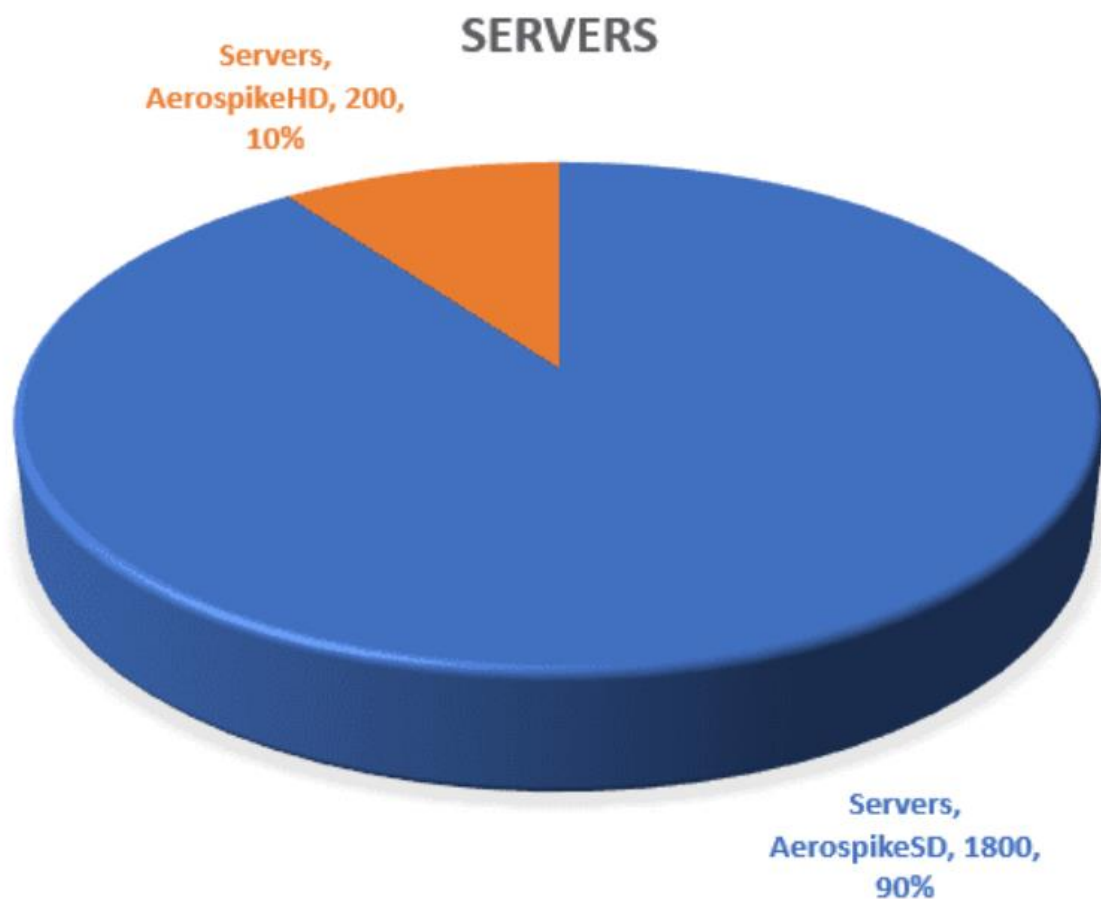
Transitioning from AerospikeSD to AerospikeHD



OS Patching on AerospikeSD vs AerospikeHD

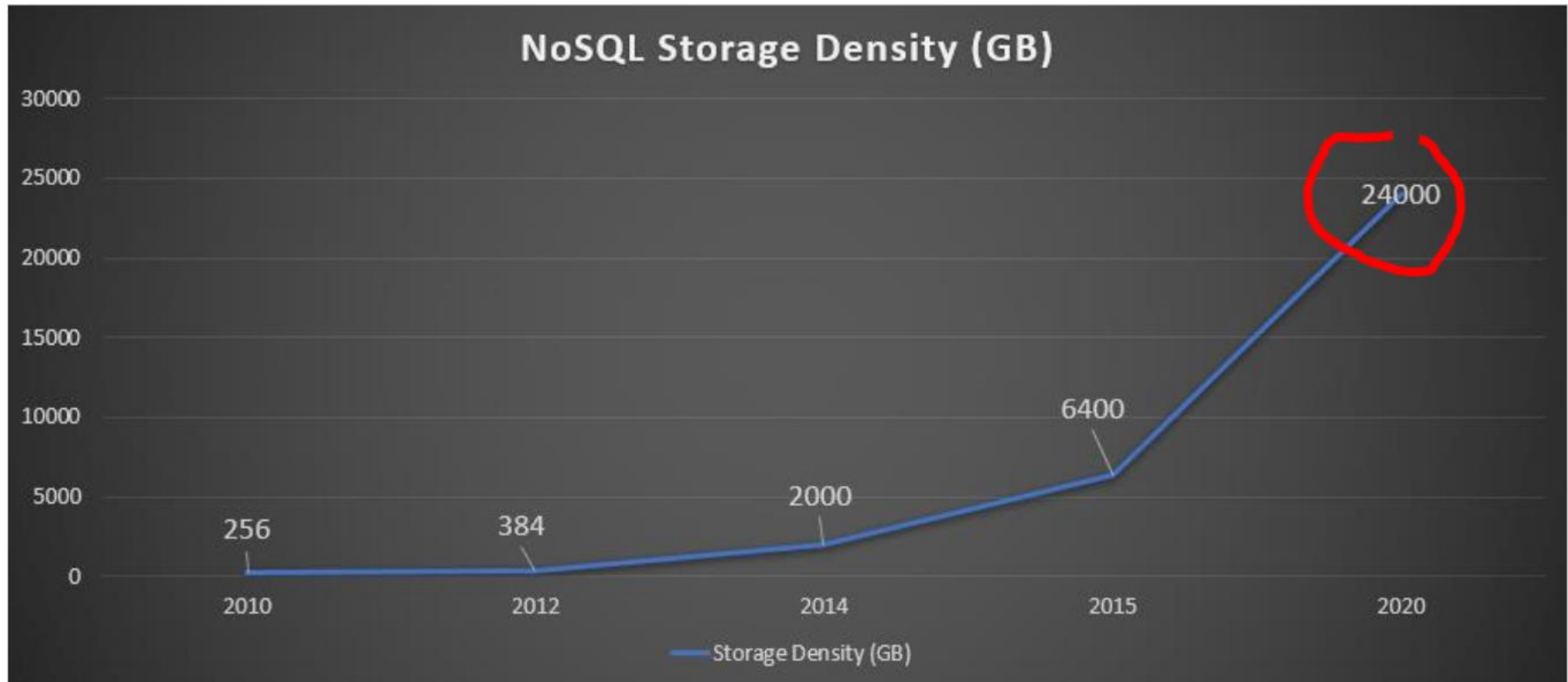


Aerospike Inventory - 2020



- 2000 Servers
- 60 Racks
- 106 Clusters
- 200 AerospikeHD
- 1800 AerospikeSD

2020





Aerospike + Persistent Memory = **Voila !**



Thank You !
(Stay safe)

Q & A