SDC
STORAGE DEVELOPER CONFERENCE
SNIA ■ SANTA CLARA, 2014

I/O Acceleration by Host Side Resources

Chethan Kumar
PernixData

Story So Far …

Virtualization has resulted in

- Longer I/O path
  - Through layers of storage abstraction
- Exponential growth in the load on the storage
- Substantial increase in repeated data access
- Revelation that storage cannot scale like compute, memory and network

Time for “Real Storage Virtualization”

- Move frequently accessed data close to consumers
- Send only new data to storage
- Best use of existing storage for data services (persistency, backup and disaster recovery)
- Utilize host resources for high speed I/O

Agenda

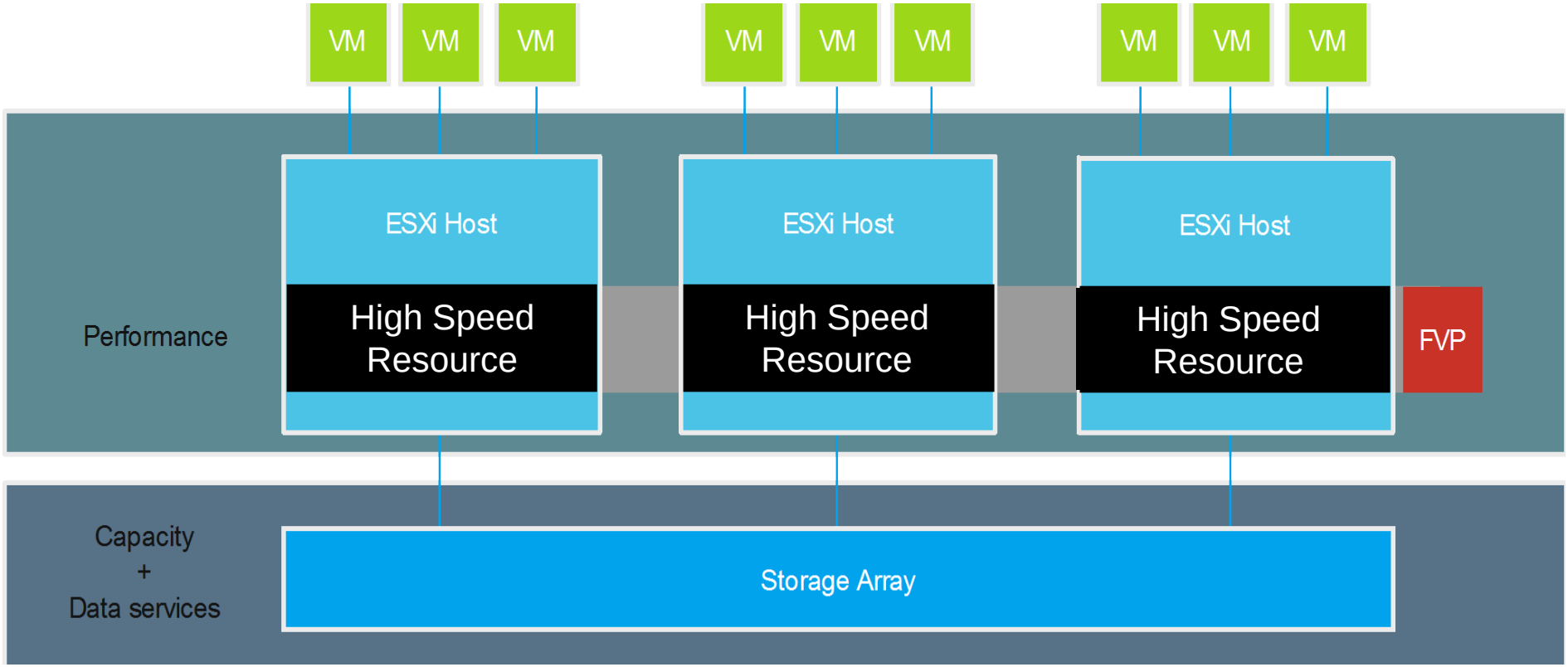
- Acceleration Tier – A Glance
- Read Acceleration
  - Static VM
  - Mobile VM
- Write Acceleration
  - Write Through
  - Write Back
- Tackling the increasing demands
- Putting it All Together
- Advanced Possibilities

Host Side High Speed Resources

- Serial ATA (SATA),. based Solid State Disks (SSDs)
- Serial Attached SCSI-2 (SAS) based SSDs
- PCIe based SSDs
- Flash on DIMMs
- RAM
- 10Gigabit Ethernet

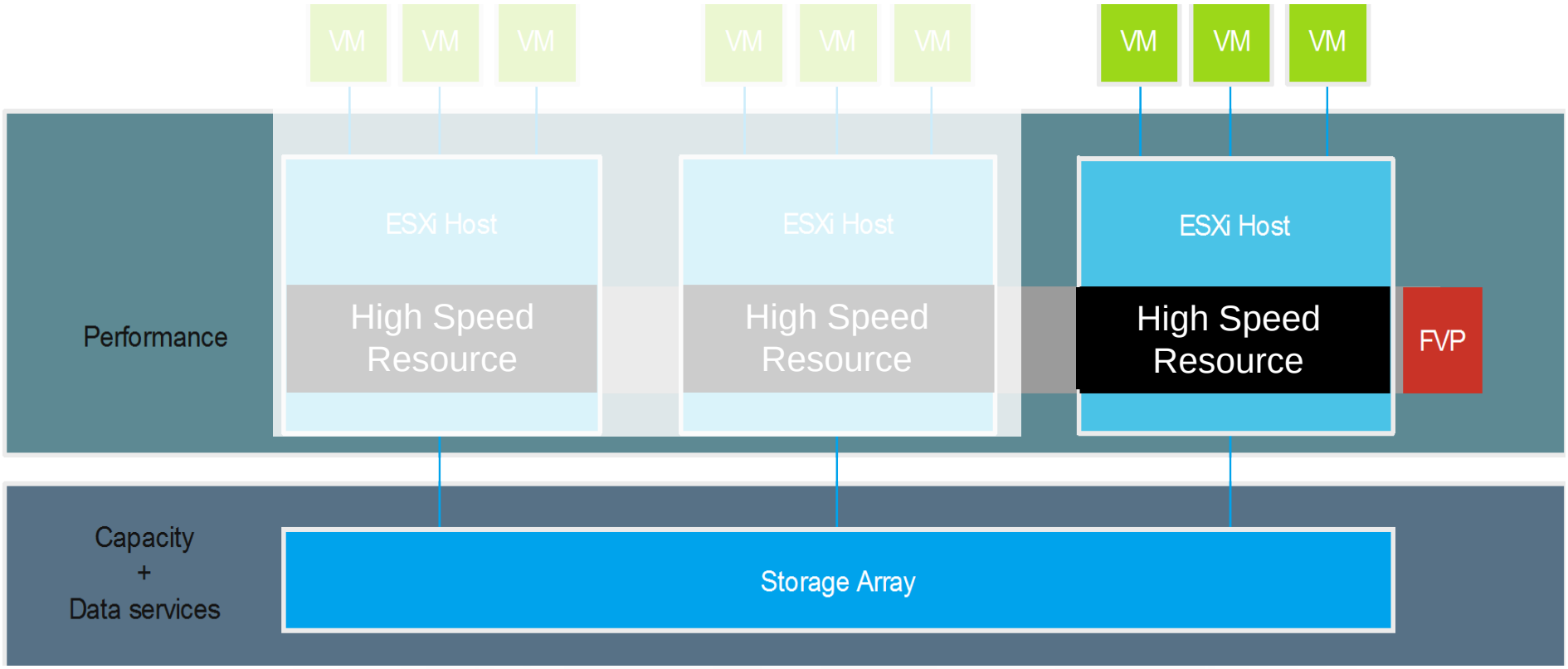
Acceleration Tier

Compute layer is ***now performance layer*** as well!



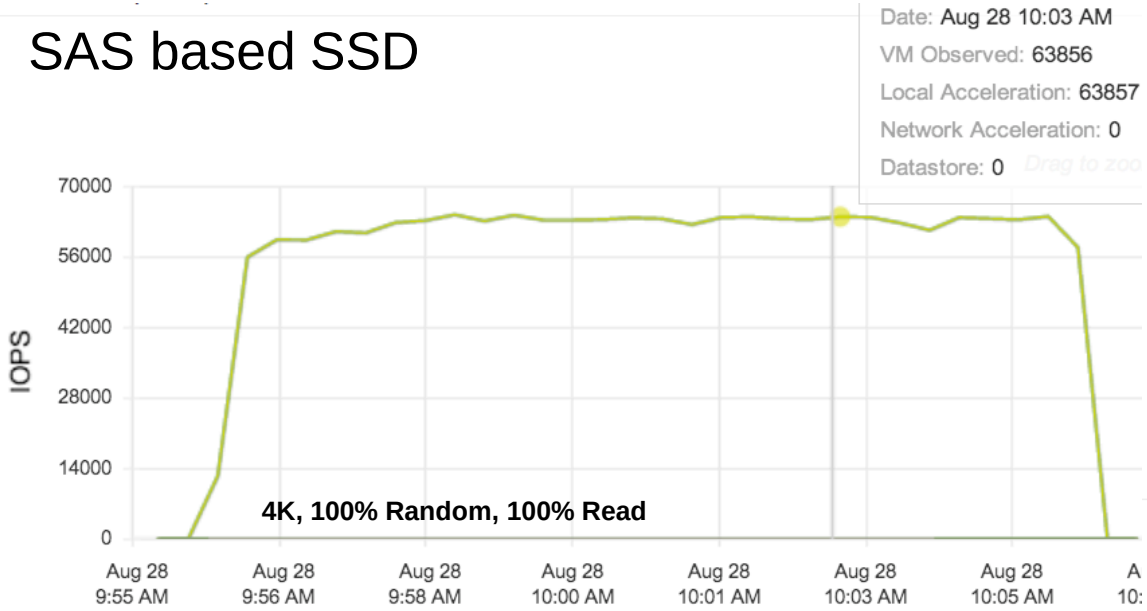
Accelerating Reads and Writes

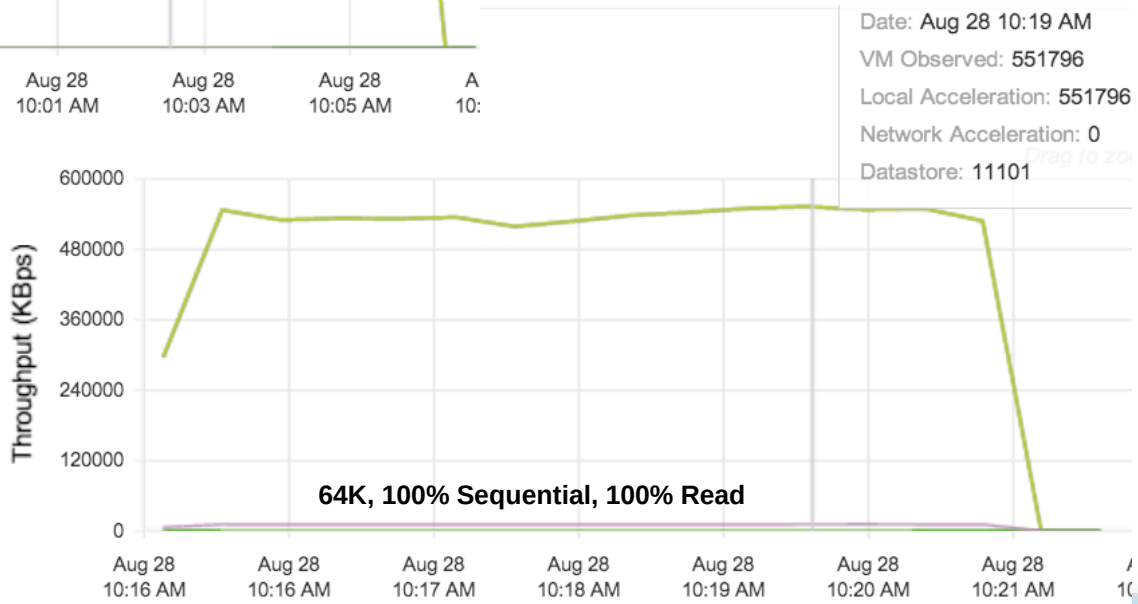
An Example



Accelerating Reads (Static VM)

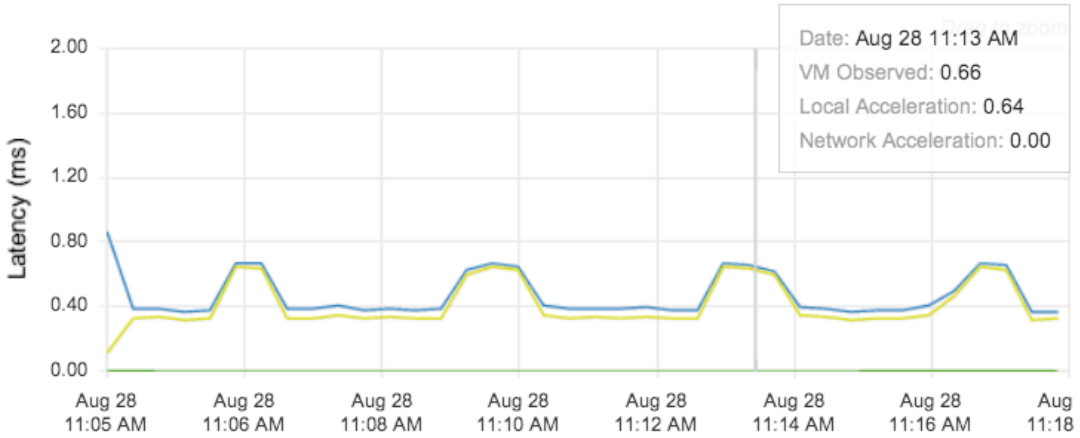
SAS based SSD



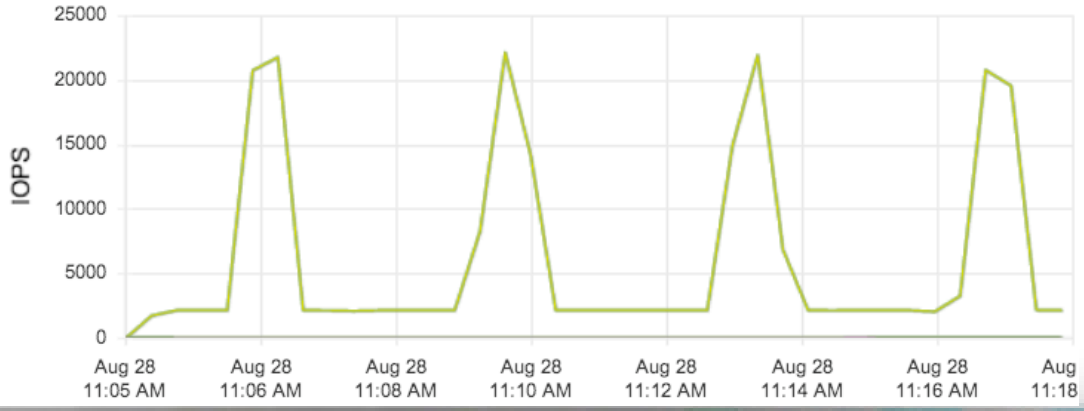


Accelerating Reads (Static VM)

SAS based SSD

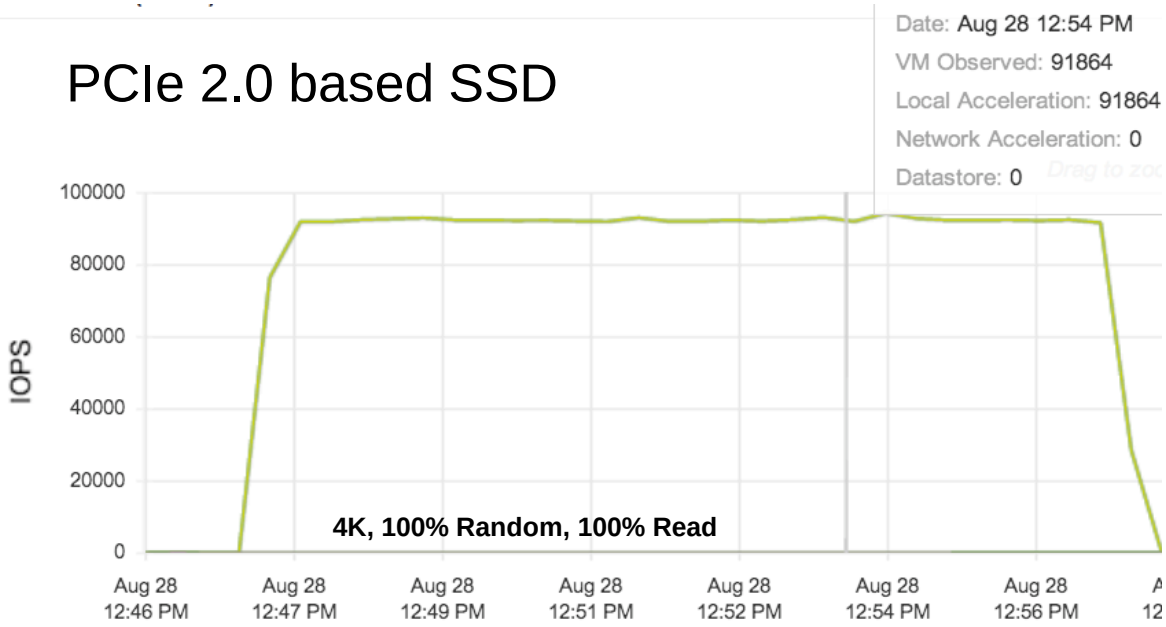


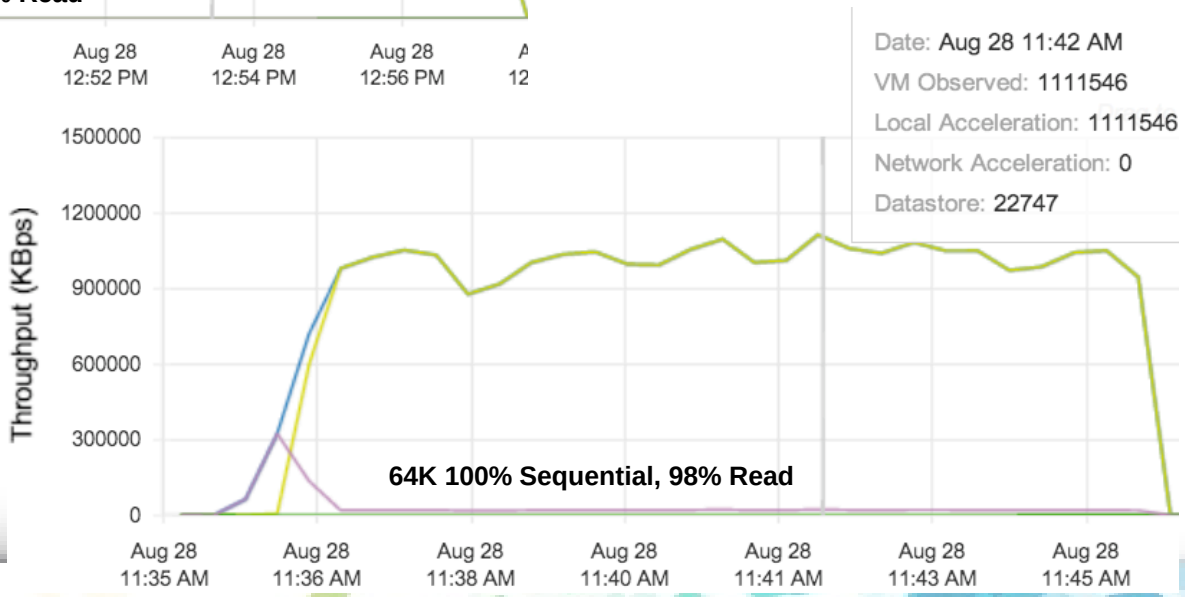
8K, Bursty I/O



Accelerating Reads (Static VM)

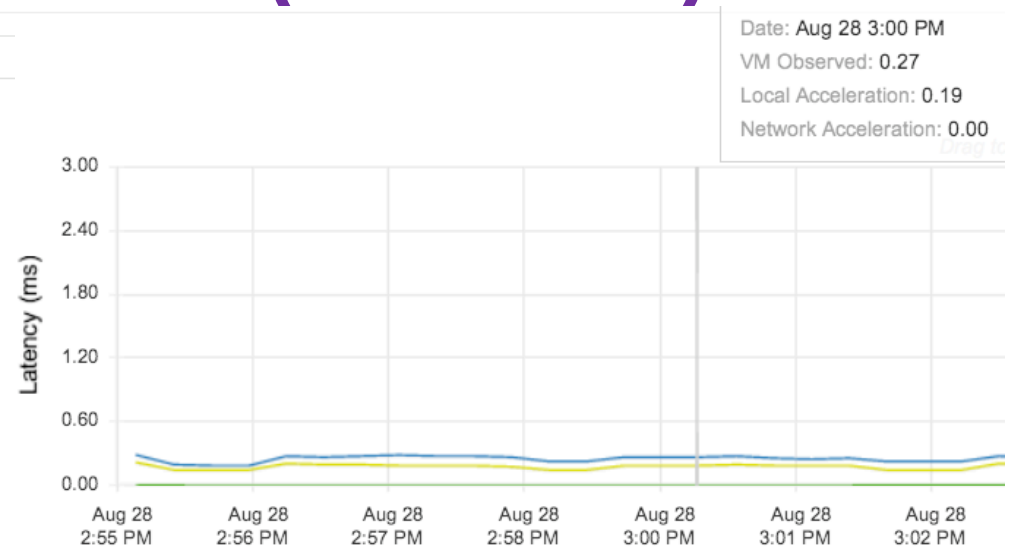
PCIe 2.0 based SSD



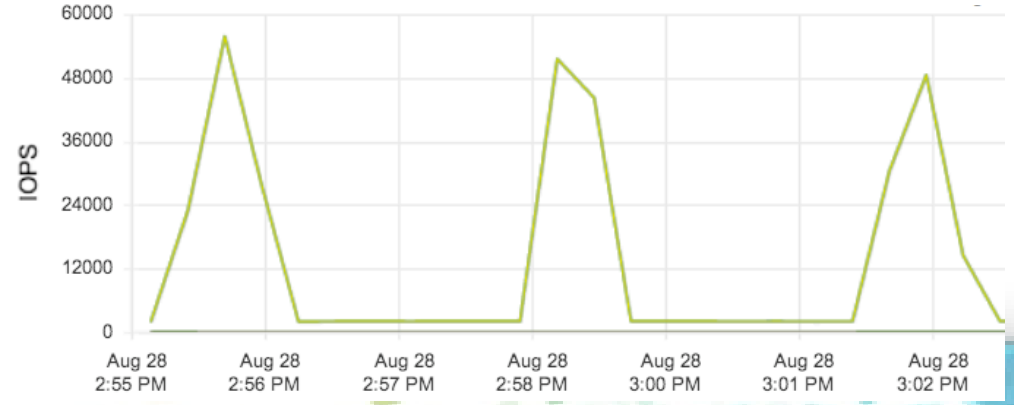


Accelerating Reads (Static VM)

PCIe 2.0 based SSD

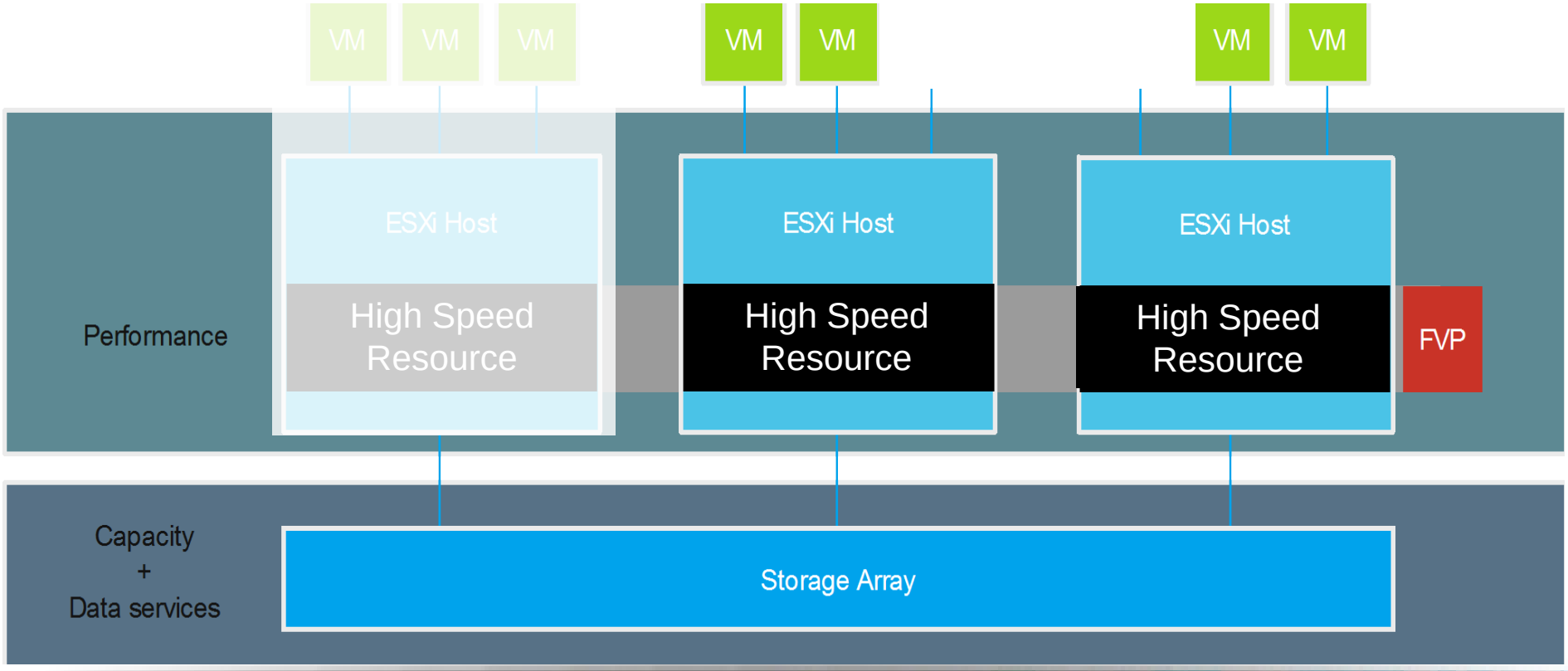


8K, Bursty I/O



Accelerating Reads (Mobile VM)

- VMs move across hosts in a vSphere cluster
  - What happens to VM's hot footprint?



Accelerating Reads (Mobile VM)

Can VM's footprint be rebuilt after every migration?

Accelerating Reads (Mobile VM)

Can the footprint be migrated?

- **Proactive migration**
  - Overuse of resources
- **On-demand migration**
  - Migrate **enough** only when needed

Accelerating Reads (Mobile VM)

VM's footprint migrated on demand

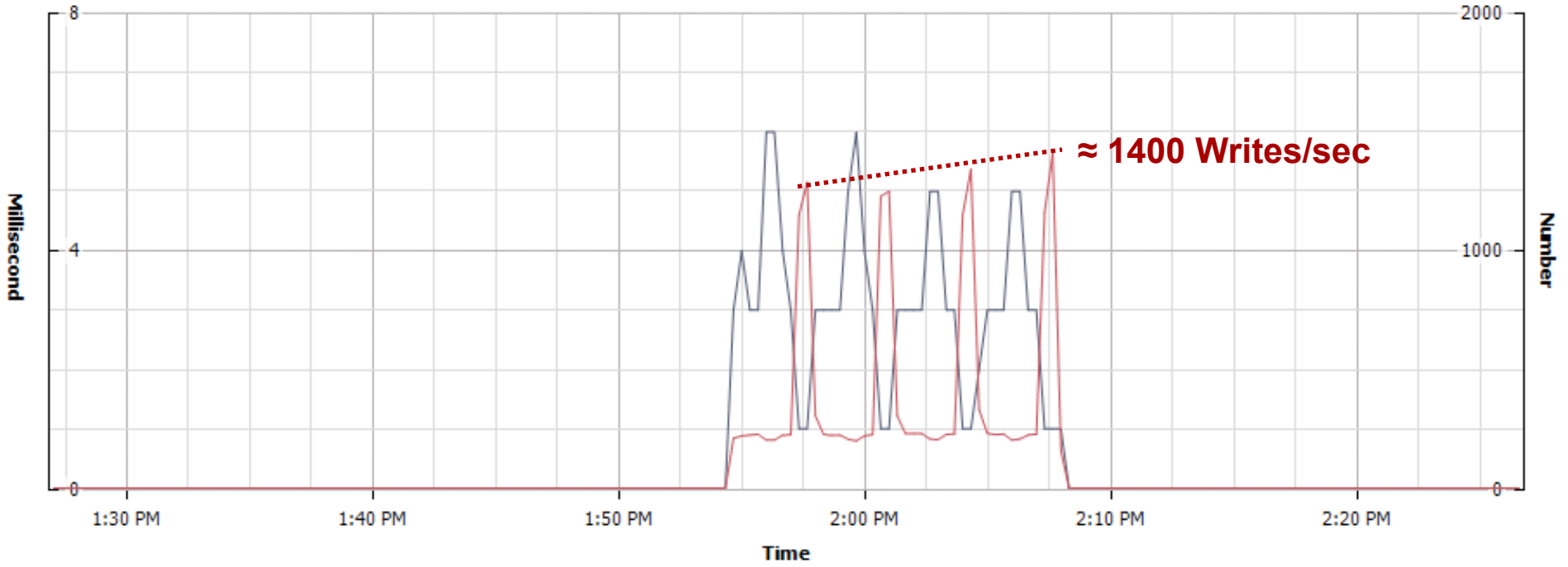
Accelerating Reads (Mobile VM)

On-demand migration – **15 minutes**

Name	Host	Local Accel	Network Accelera
Win2k8-fullclone	perf-5.pernixdat…	1.5 MB	67.61 GB
Win2k8-lc10	perf-3.pernixdat…	512 KB	0 GB
Win2k8-lc12	perf-3.pernixdat…	512 KB	0 GB
Win2k8-lc11	perf-3.pernixdat…	512 KB	0 GB

Name	Host	Local Accel	Network Accelera
Win2k8-fullclone	perf-5.pernixdat…	21.27 GB	67.61 GB
Win2k8-lc10	perf-3.pernixdat…	512 KB	0 GB
Win2k8-lc12	perf-3.pernixdat…	512 KB	0 GB
Win2k8-lc11	perf-3.pernixdat…	512 KB	0 GB

Accelerating Writes

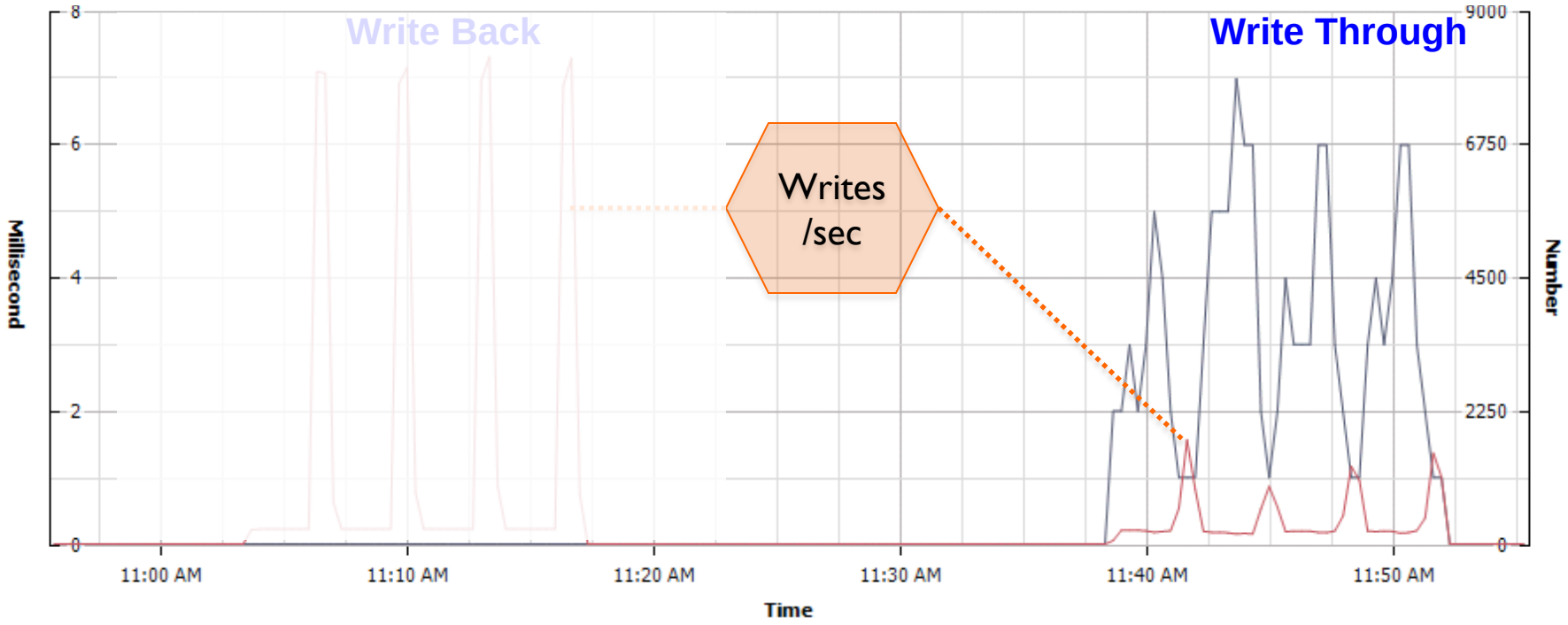


Performance Chart Legend

Key	Object	Measurement	Rollup	Units	Latest	Maximum	Minimum	Average
	scsi1:0	Write latency	Average	Millisecond	0	6	0	0.7
	scsi1:0	Average write requests per second	Average	Number	0	1414	0	97.1

Accelerating Writes

PCIe 2.0 based SSD

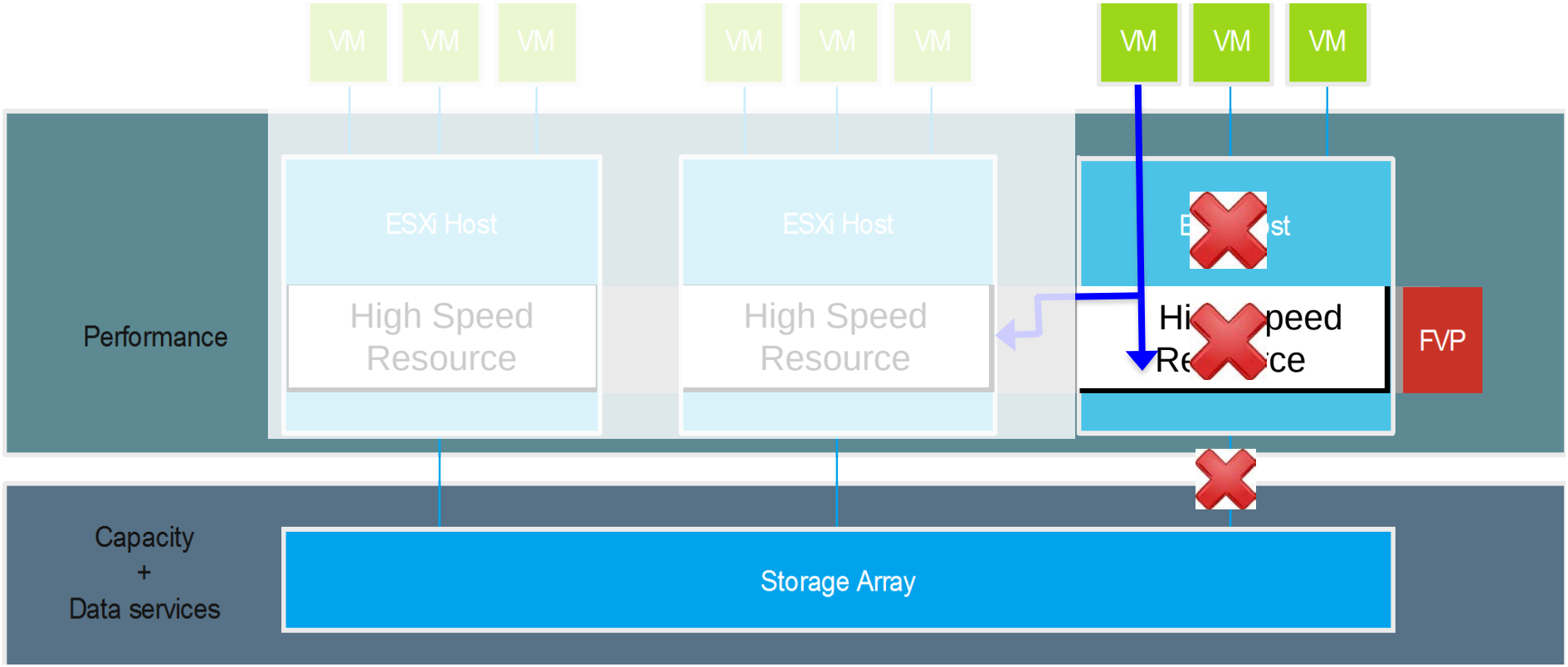


Performance Chart Legend

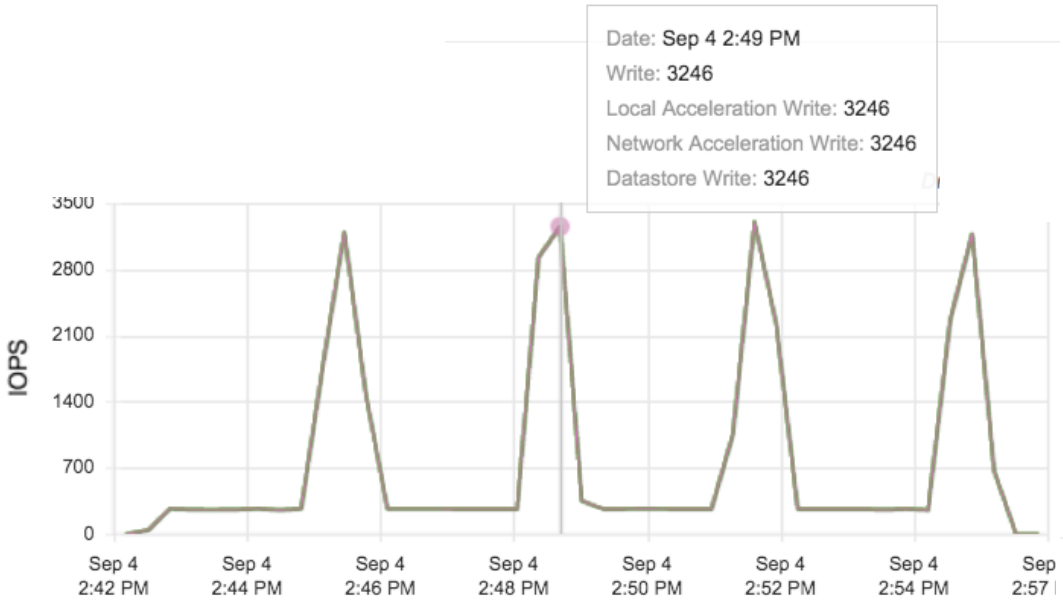
Key	Object	Measurement	Rollup	Units	Latest	Maximum	Minimum	Average
	scsi1:0	Write latency	Average	Millisecond	0	7	0	0.739
	scsi1:0	Average write requests per second	Average	Number	0	8244	0	511.872

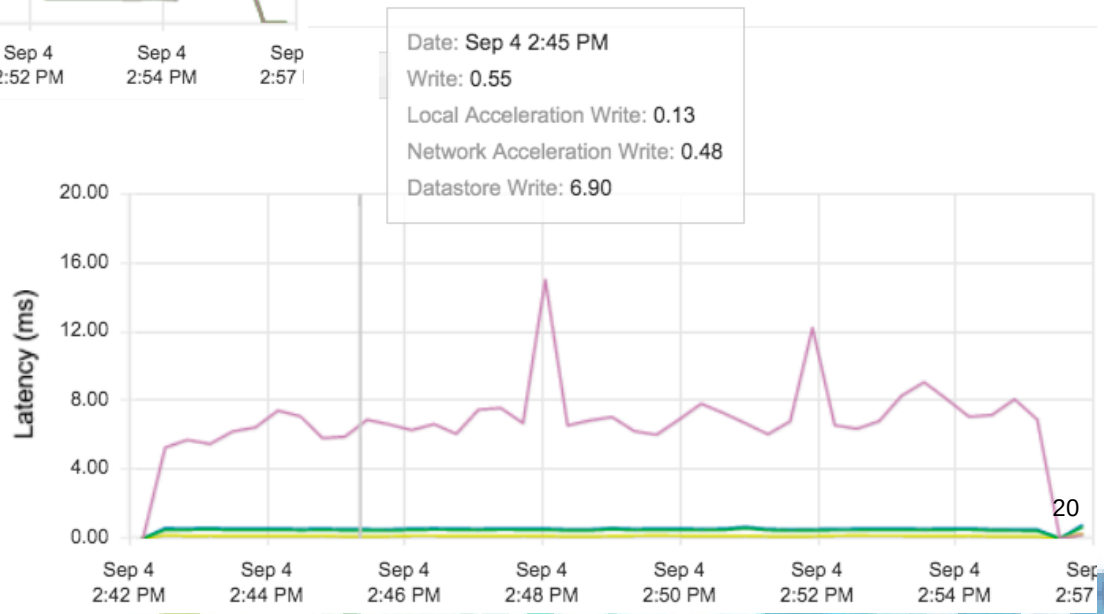
Fault Tolerant Write-Back

Writes to peer hosts over **Ethernet links** for fault tolerance



Fault Tolerant Write-Back



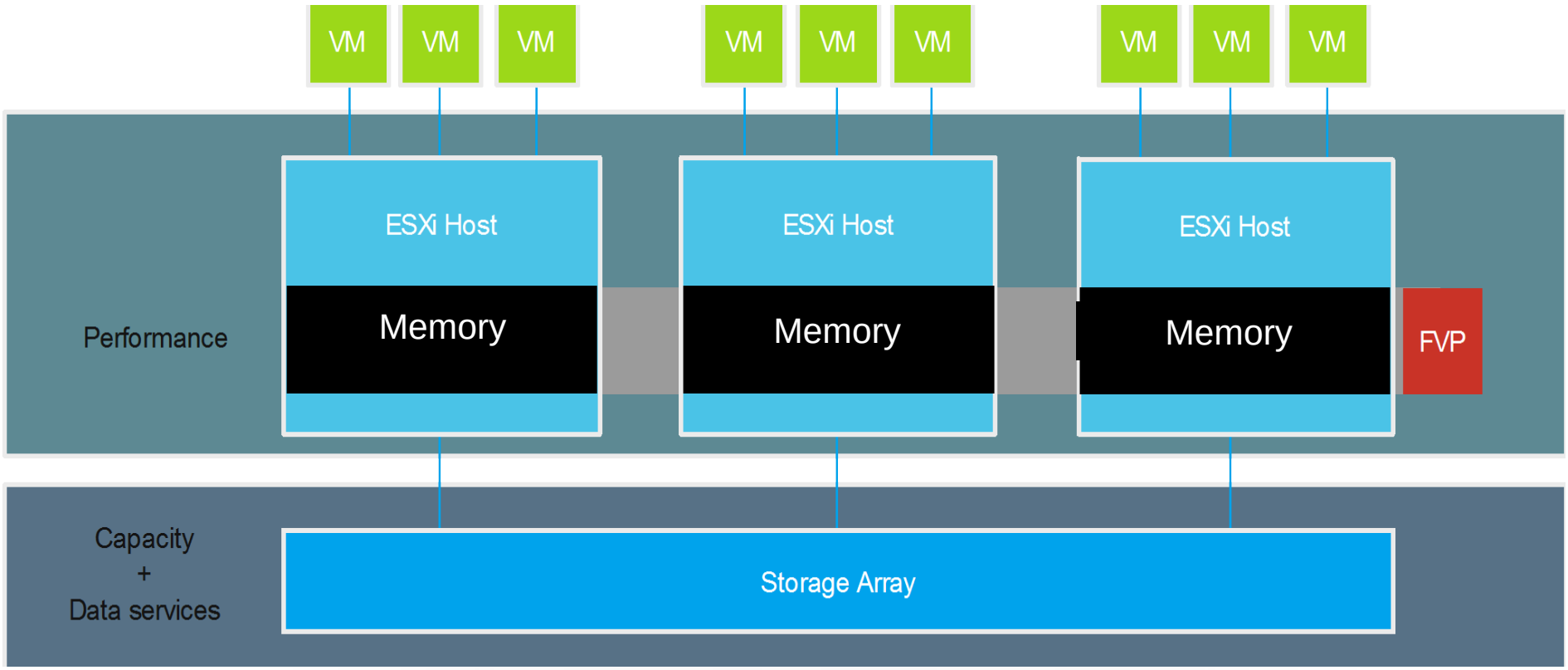


Tackling the Increasing Demands

- Ever increasing demand for I/Os
- Resources local to a host can satisfy host needs
  - Avoid local problems from becoming global
- Scale-out acceleration tier
- "Dividing and Conquering the Problem"

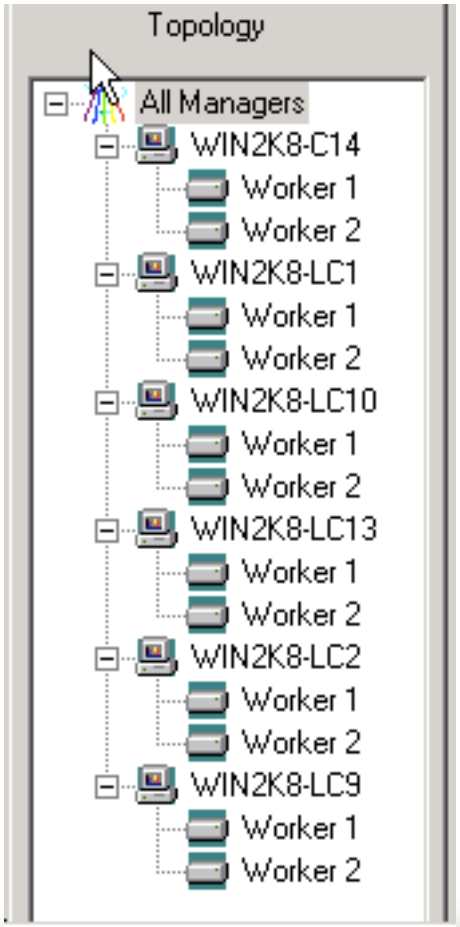
Scaling I/O Performance

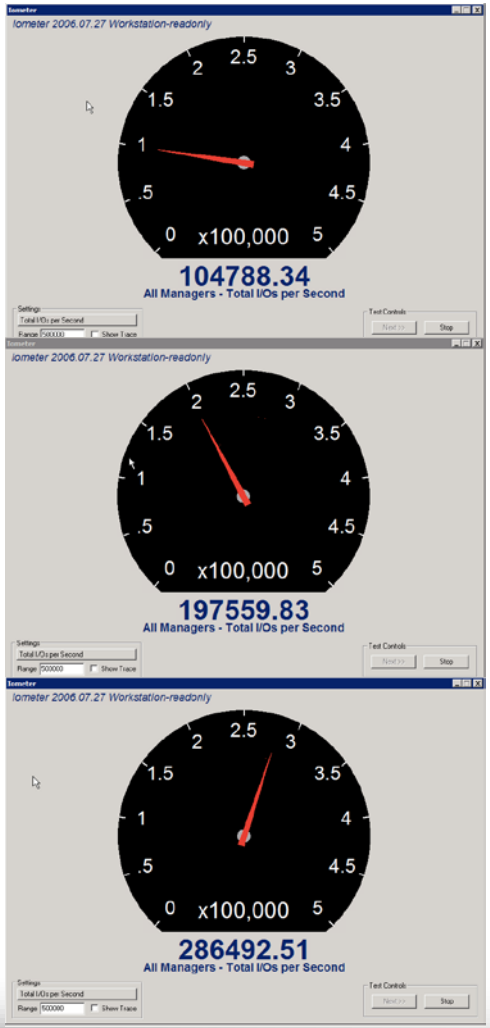
At **Cluster** level



RAM based I/O Acceleration

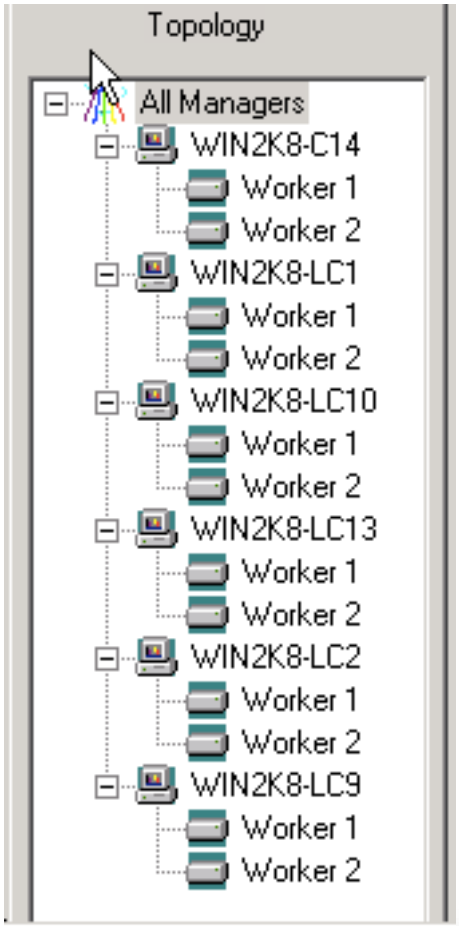
4KB random reads

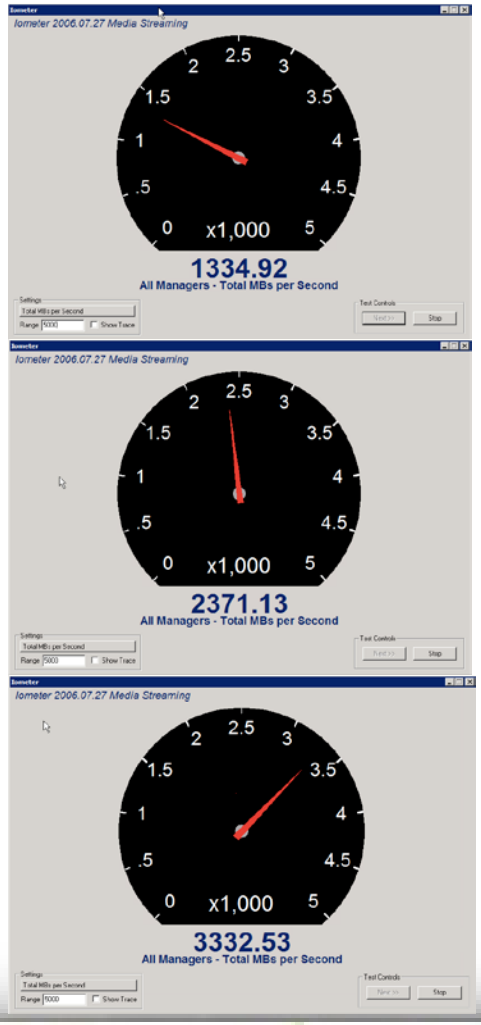


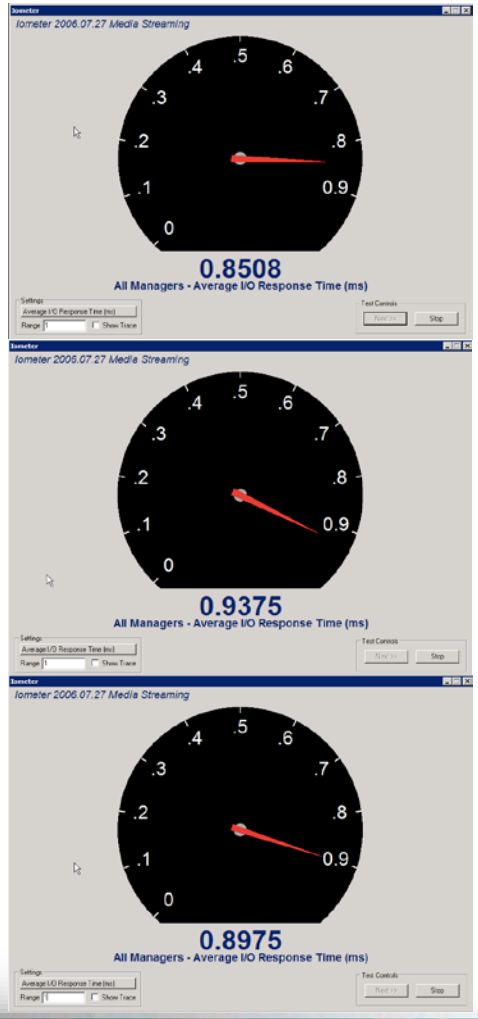


RAM based I/O Acceleration

64KB sequential reads



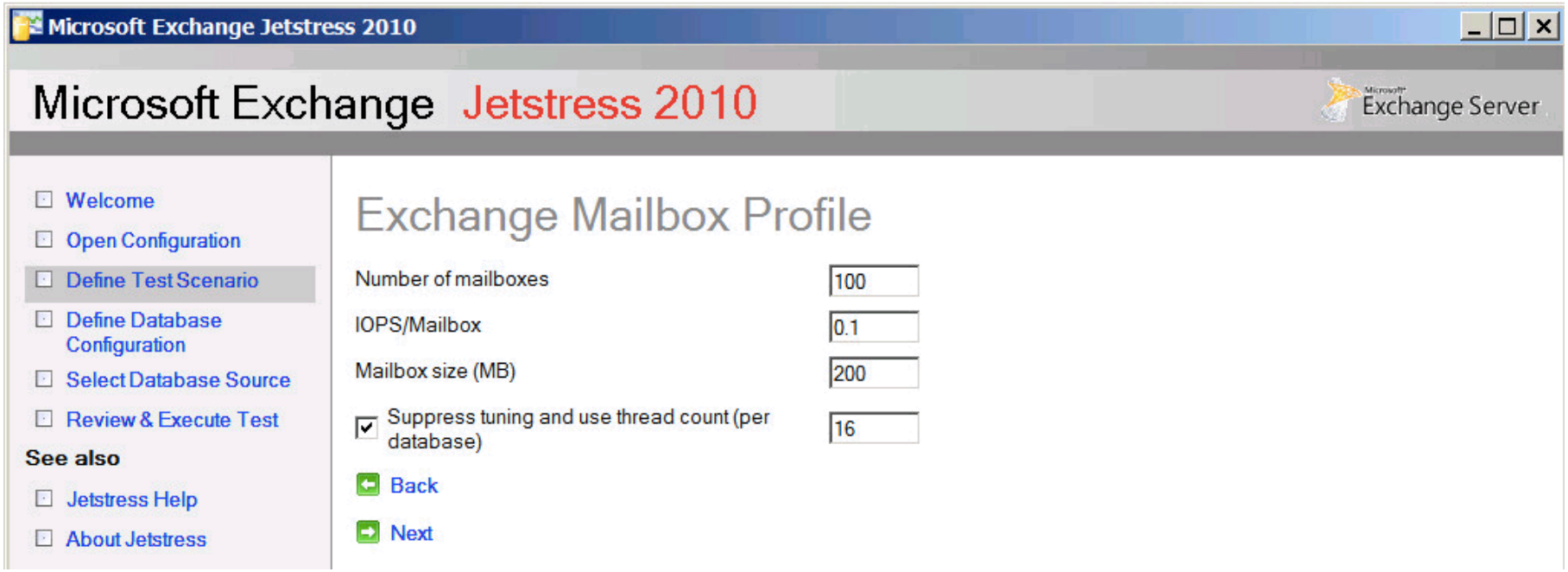




Putting it all in action…

Accelerating Enterprise Applications

MS Exchange Server 2010 (JetStress)



Accelerating Enterprise Applications

MS Exchange Server 2010 (JetStress)

Database Sizing and Throughput

Achieved Transactional I/O per Second	7686.156
Target Transactional I/O per Second	10
Initial Database Size (bytes)	22288793600
Final Database Size (bytes)	23396089856
Database Files (Count)	4

Transactional I/O Performance

MSExchange Database ==> Instances	I/O Database Reads Average Latency (msec)	I/O Database Writes Average Latency (msec)	I/O Database Reads/sec	I/O Database Writes/sec	I/O Database Reads Average Bytes	I/O Database Writes Average Bytes	I/O Log Reads Average Latency (msec)	I/O Log Writes Average Latency (msec)	I/O Log Reads/sec	I/O Log Writes/sec	I/O Log Reads Average Bytes	I/O Log Writes Average Bytes
Instance1716.1	6.453	12.124	902.939	998.893	33471.838	37214.747	0.000	3.489	0.000	227.129	0.000	13771.786
Instance1716.2	6.526	12.463	906.913	999.628	33436.336	37211.162	0.000	3.548	0.000	225.593	0.000	13854.196
Instance1716.3	6.514	12.732	913.405	1013.013	33409.578	37193.471	0.000	3.585	0.000	228.543	0.000	13679.762
Instance1716.4	6.605	12.241	929.657	1021.708	33445.276	37251.802	0.000	3.445	0.000	230.942	0.000	13705.067

Accelerating Enterprise Applications

MS Exchange Server 2010 (JetStress)

Database Sizing and Throughput

Achieved Transactional I/O per Second	15048.471
Target Transactional I/O per Second	100
Initial Database Size (bytes)	21038891008
Final Database Size (bytes)	23077322752
Database Files (Count)	4

Transactional I/O Performance

MSExchange Database ==> Instances	I/O Database Reads Average Latency (msec)	I/O Database Writes Average Latency (msec)	I/O Database Reads/sec	I/O Database Writes/sec	I/O Database Reads Average Bytes	I/O Database Writes Average Bytes	I/O Log Reads Average Latency (msec)	I/O Log Writes Average Latency (msec)	I/O Log Reads/sec	I/O Log Writes/sec	I/O Log Reads Average Bytes	I/O Log Writes Average Bytes
Instance2256.1	2.655	4.109	1760.810	1947.351	33280.746	36634.577	0.000	1.223	0.000	542.872	0.000	11095.013
Instance2256.2	2.613	4.301	1785.773	1991.121	33338.699	36581.943	0.000	1.140	0.000	548.401	0.000	11202.158
Instance2256.3	2.631	4.288	1789.353	1990.257	33392.681	36595.541	0.000	1.192	0.000	545.866	0.000	11107.949
Instance2256.4	2.563	4.167	1792.127	1991.679	33306.768	36579.590	0.000	1.131	0.000	548.614	0.000	11113.892

Real Life Scenarios

DVDStore[1] Workload on MS SQL Server 2008 Database

Application Metrics	Non-Accelerated Database	Accelerated Database	% Improvement
Orders per Sec	74.6	170.5	+129%
Order Completion time (ms)	231.5	93.4	+60%

http://www.pernixdata.com/resource/accelerating-virtualized-databases key=144c291ed857a627307bf3ebcf3a7c3f

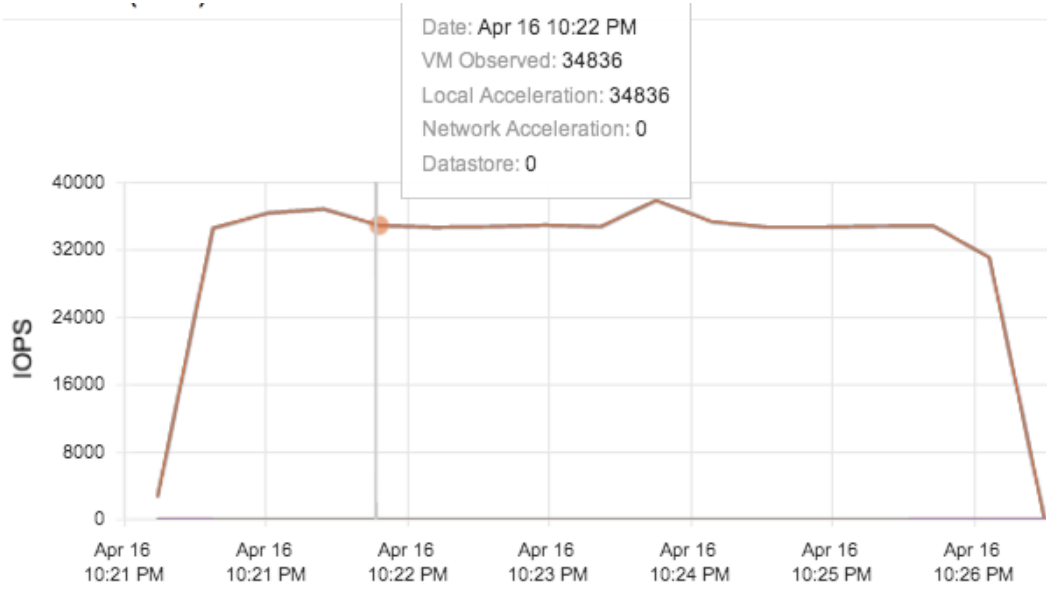
[1] *http://linux.dell.com/dvdstore/*

Advanced Possibilities

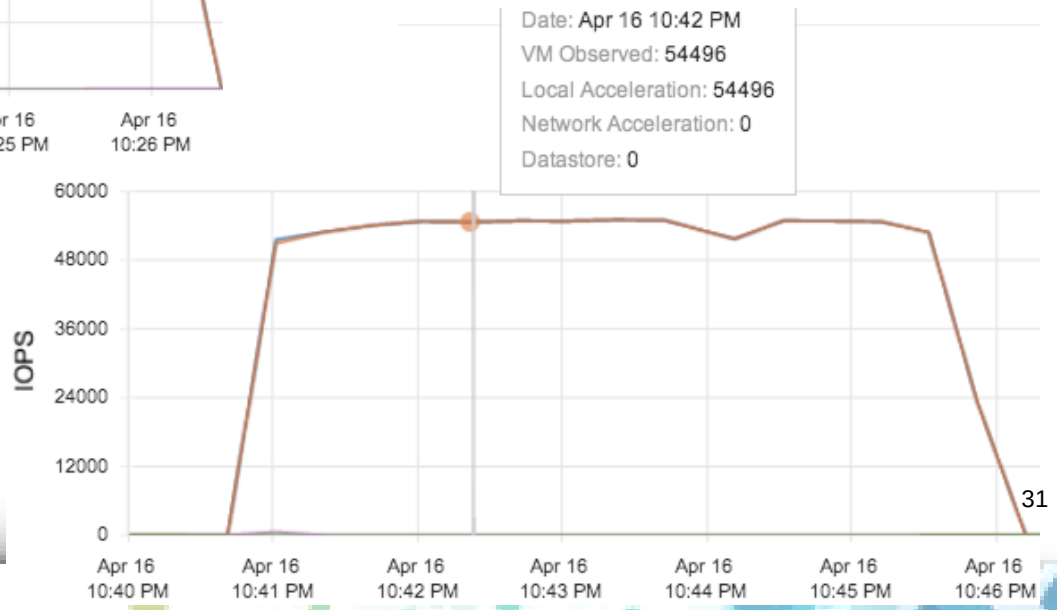
- Flash Faster Than Flash
- Fault Tolerant Writes

Flash Faster than Flash

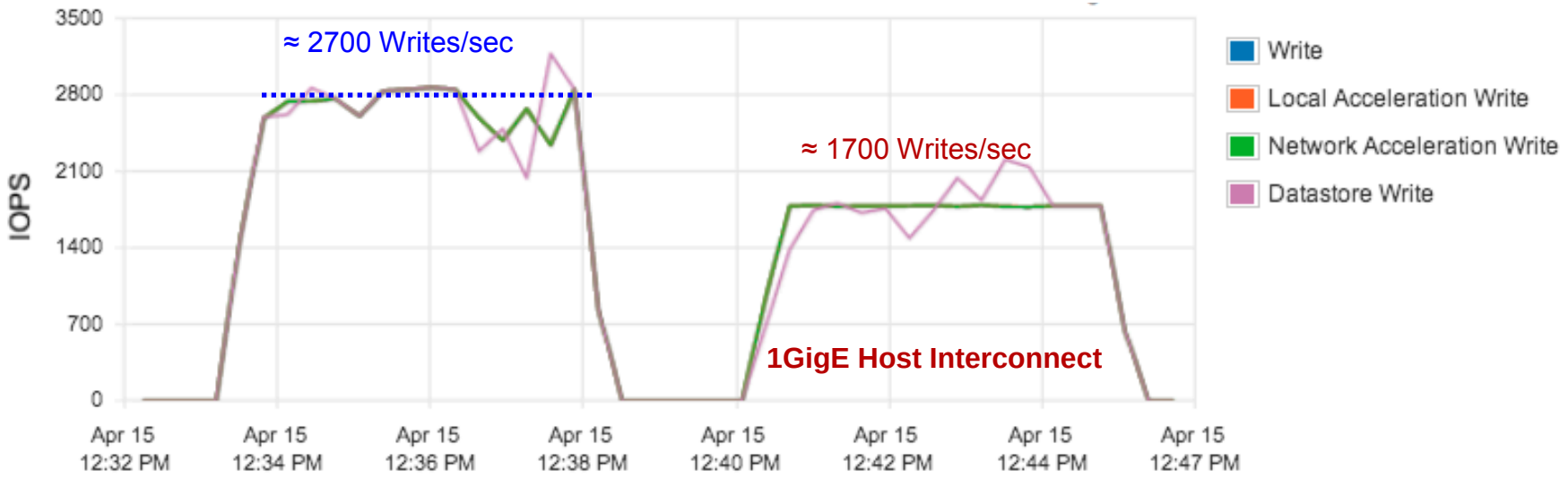
Can acceleration tier improve SSD characteristics?



4K, 100% Random, 100% Read



Fault Tolerant Writes



8K, 100% Random, 100% Write

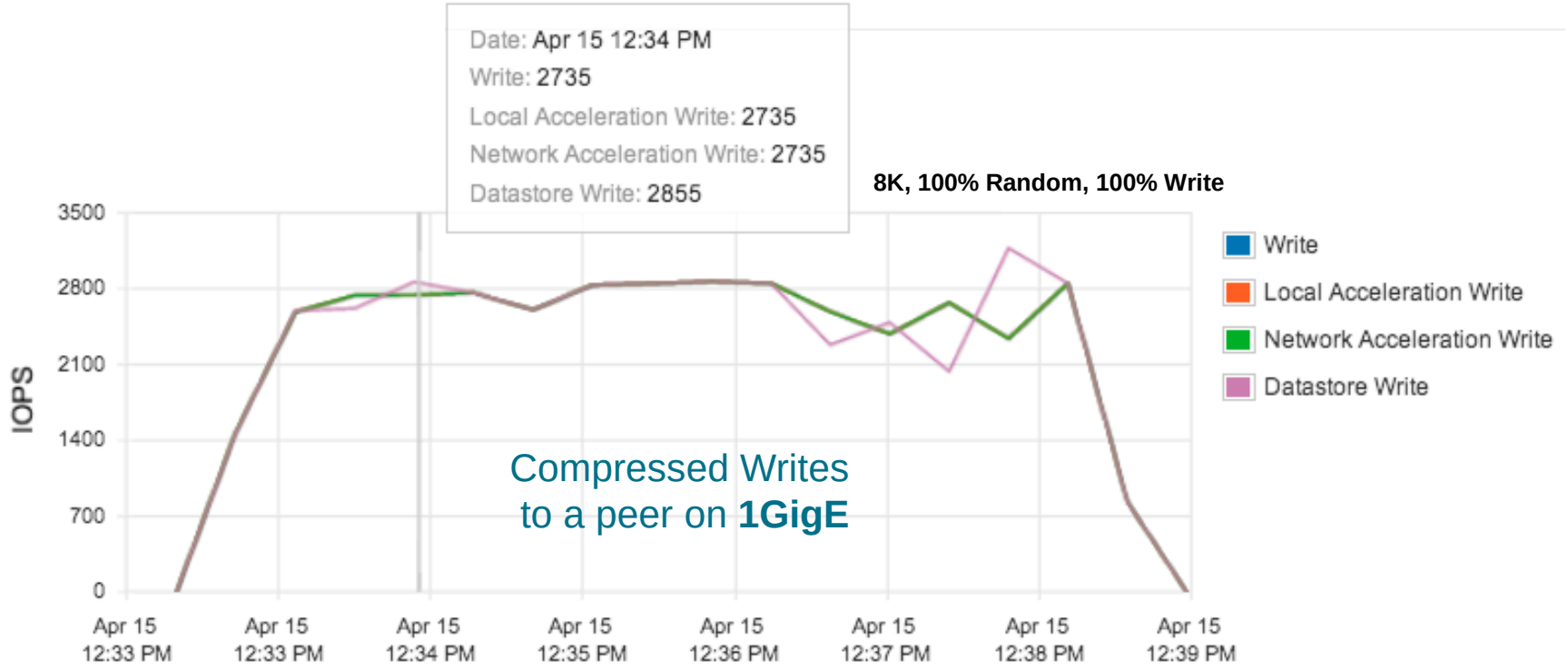
Fault Tolerant Writes

Compatible speed? No issues

8K, 100% Random, 100% Write

Fault Tolerant Writes

Speed mismatch? No issues



Summary

- Compute layer + Intelligent software ➔ high performant storage infrastructure
  - Previously impossible!!
- Host side high speed resources ➔ Insane I/O acceleration
- Results are mind-blowing in spite of clustering challenges
- Many more to come …

BACKUP

Storage Relief

In a 7 Month Period:

Historical (Since 2013-12-19 11:11:03 AM)

IOs Saved from Datastore	**8,308,458,857**
Datastore Bandwidth Saved	**317.99 TB**
Writes Accelerated	**10,280,658,943**

- **8 Billion Reads** didn't reach primary storage
- **318 Terabytes** of storage bandwidth not used
- **10 Billion Writes** saw significantly low latency