Using Distributed Fault Tolerant Memory in Virtualized Data Centers

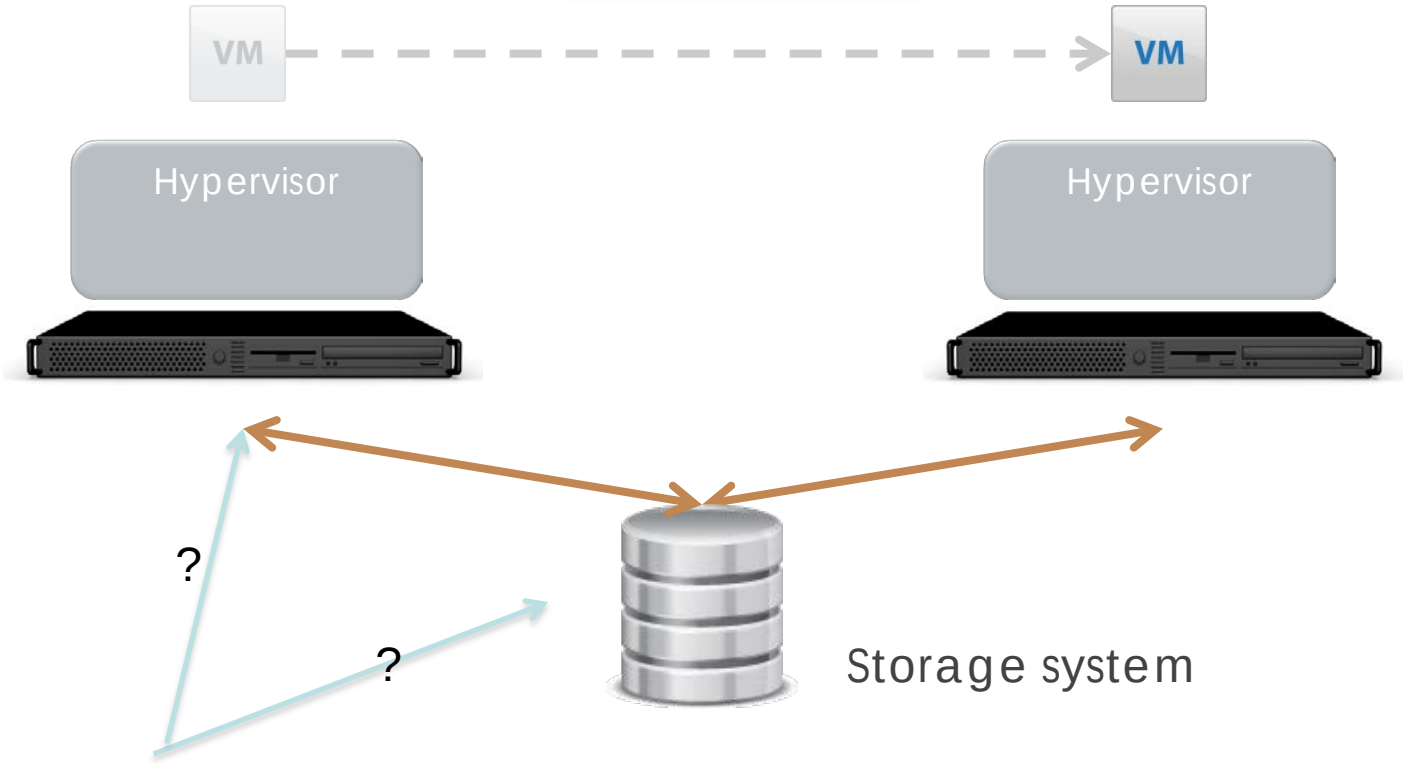
Woon Jung

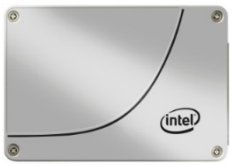
PernixData

Agenda

- Background
- Motivation for DFTM
- Survey of challenging problems with DFTM
- Q&A

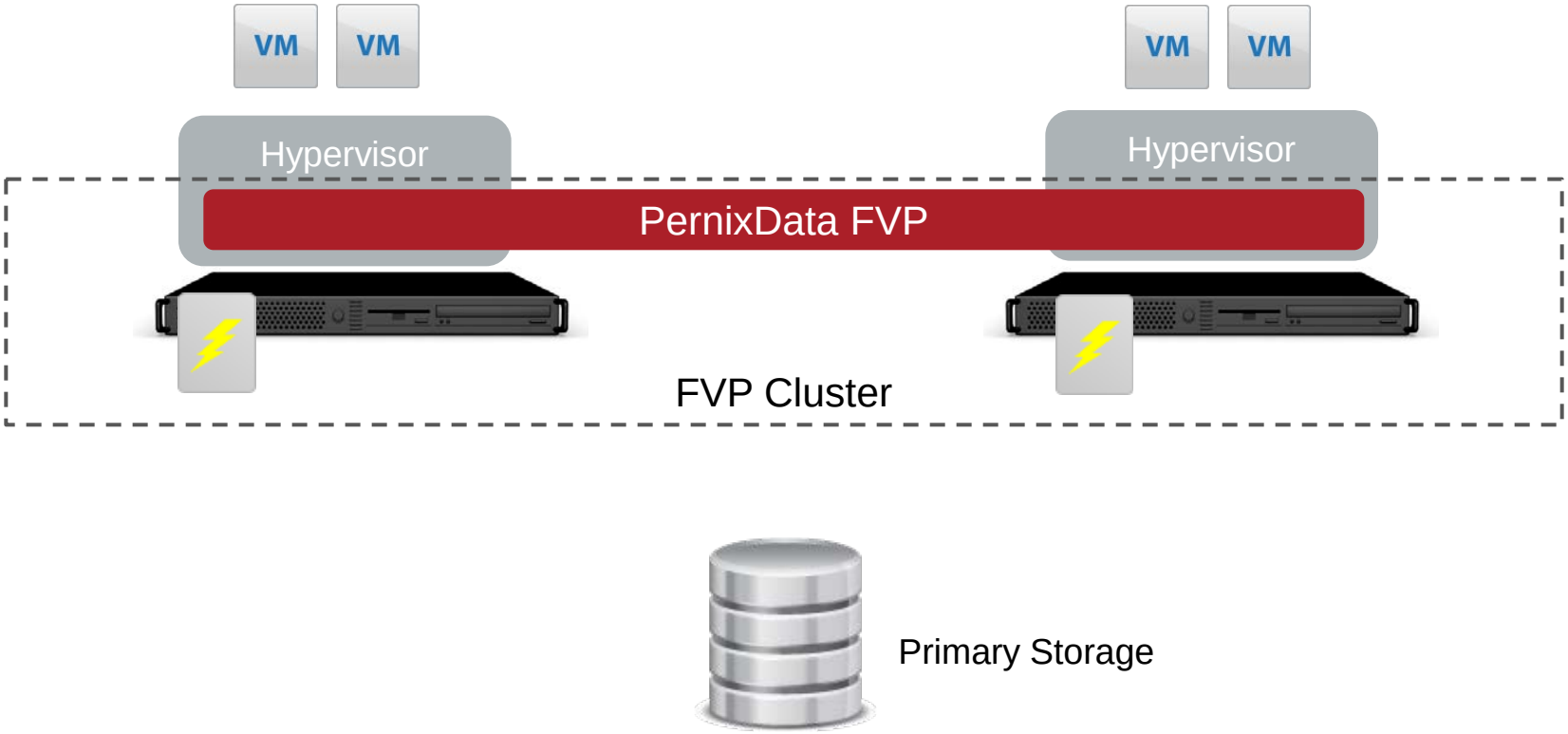
Background





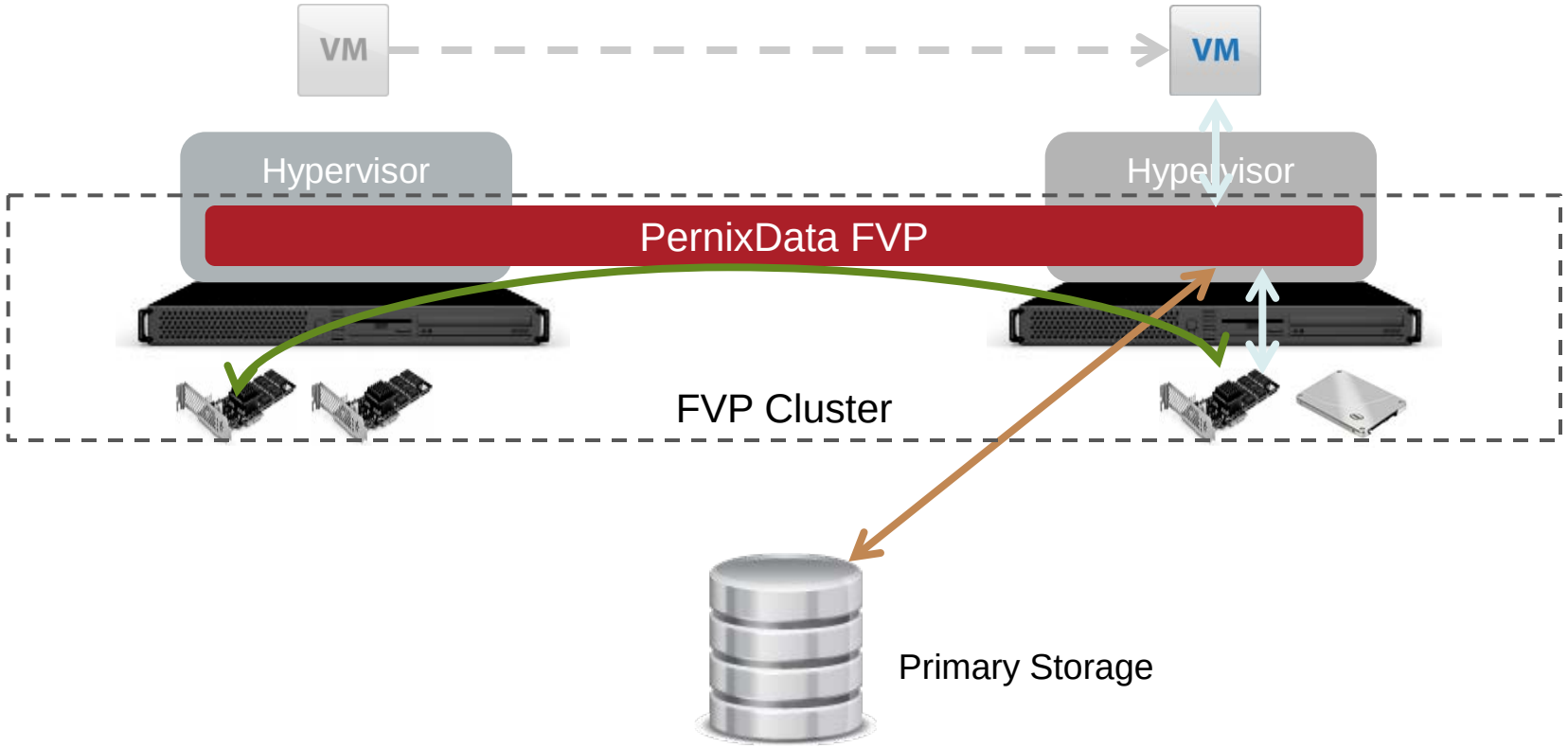
It started with FVP…

Hypervisor-based non-disruptive data acceleration

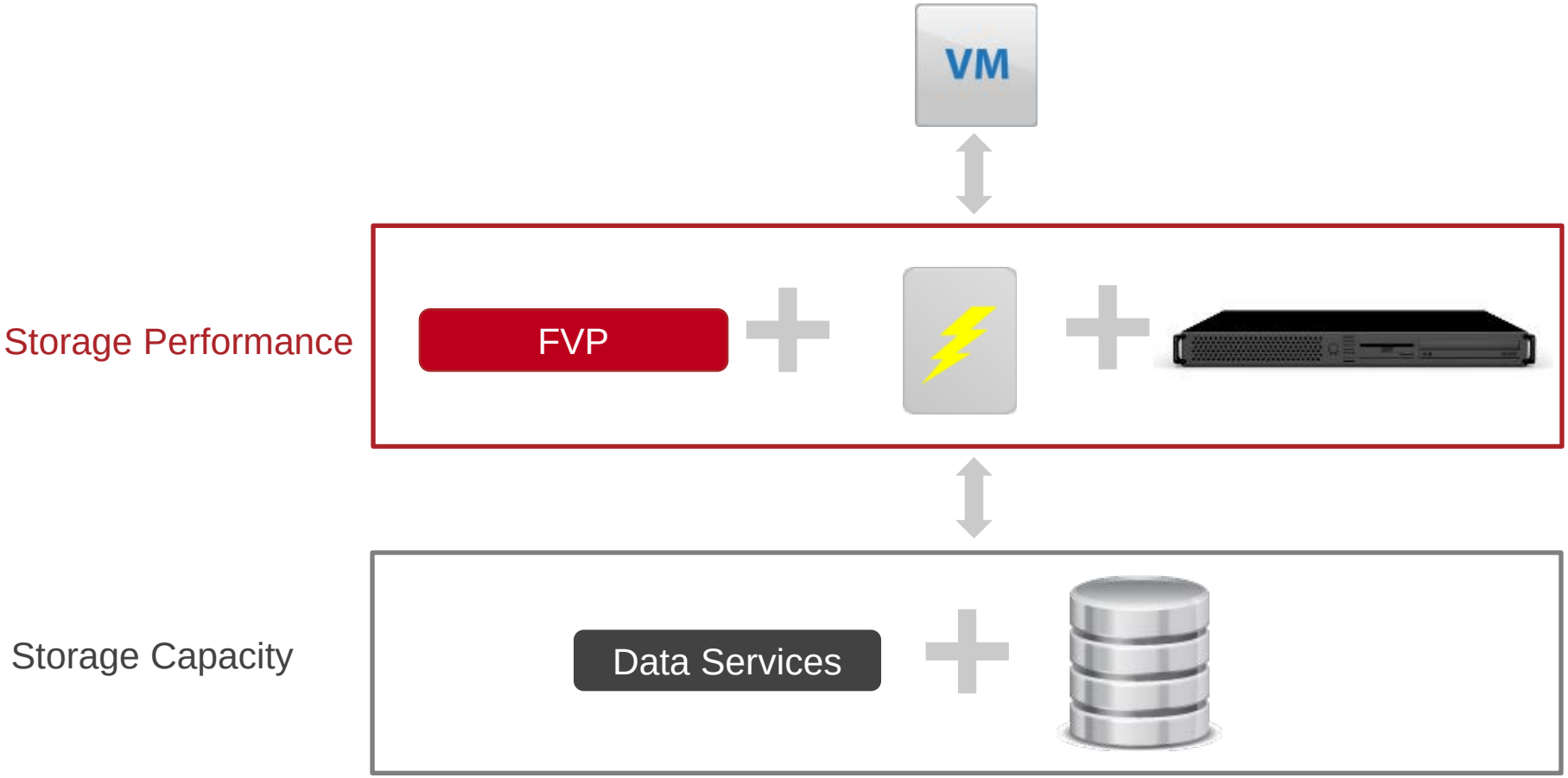


Clustered Data Tier

Server flash aggregation into cluster-wide pool



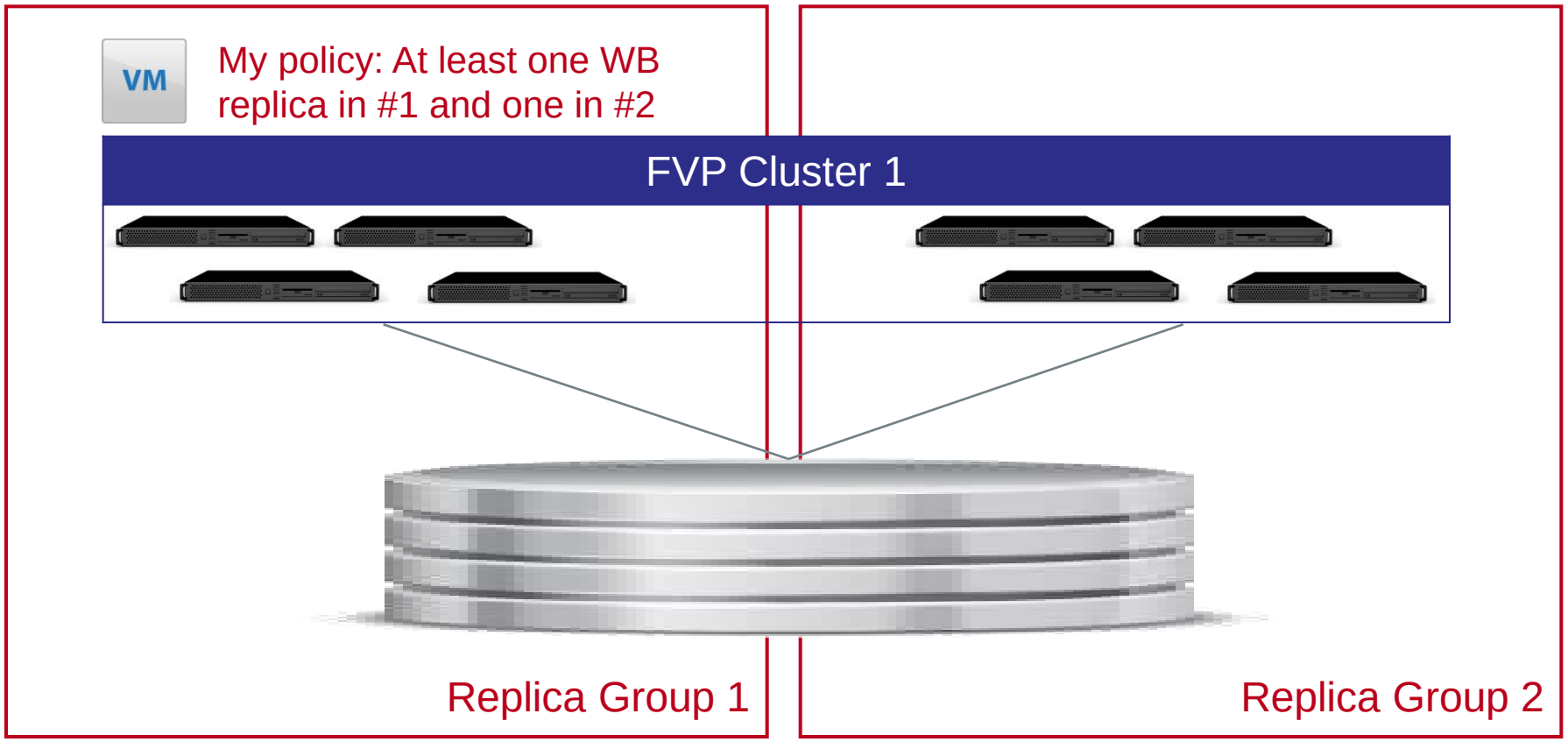
Decoupled Architecture



DFTM Motivation

- Does the exact acceleration media type matter?
  - Perfect physics
  - Resource pooling
  - Fault tolerance
- Flash is great. Why bother with Memory?
  - Denser and cheaper memory (overprovisioned memory)
  - Easy to test, deploy and manage (business operation).

Challenge : Fault Tolerance



Challenge : Software Overhead

- Software deficiencies show up… with ultra low latency devices

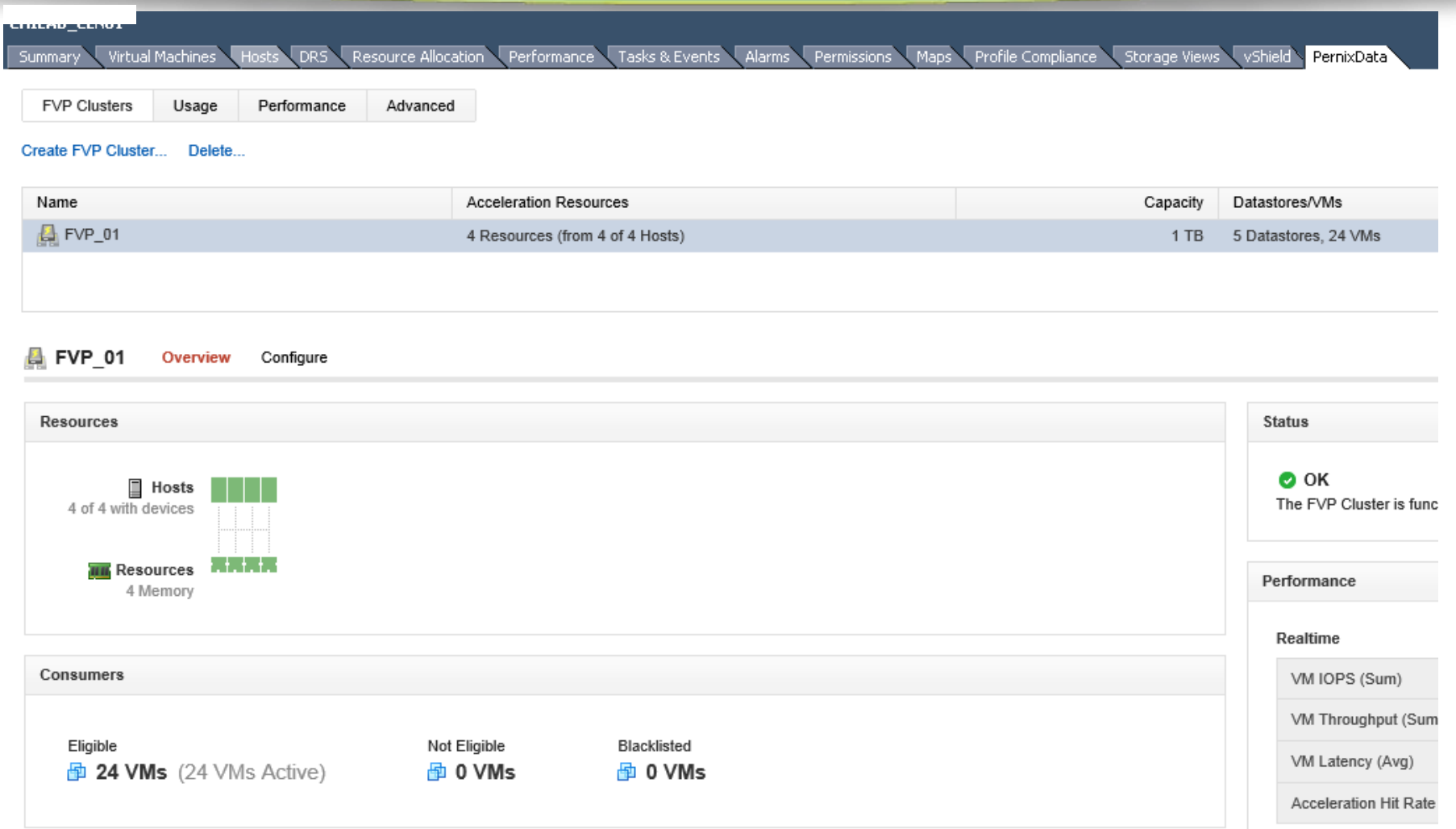
Time to issue an IO via storage stack	15us ~ 20us
Acknowledging IO completion	20us ~ 350us
4KB IO to local PCIe flash device	15us ~ 75us
4KB IO to new NVMe device	1us ~ 3us (Claimed)
4KB IO to DRAM	1us

- Dedicated contexts to issue and complete IOs
- Scrutiny over every single lock.
- Memory Allocation.

Challenge : Memory is Precious

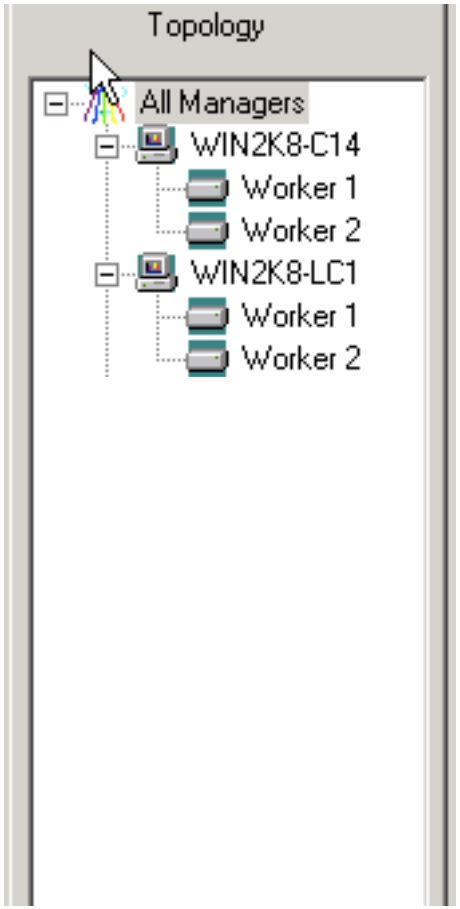
- Dynamic re-sizing
  - Memory is a “flexible” resource.
- Reduce Metadata overhead
  - This works great even with Flash.
- DFTM-z (Compression)
  - Do more with the same amount of RAM.

Hello, I am FVP with DFTM



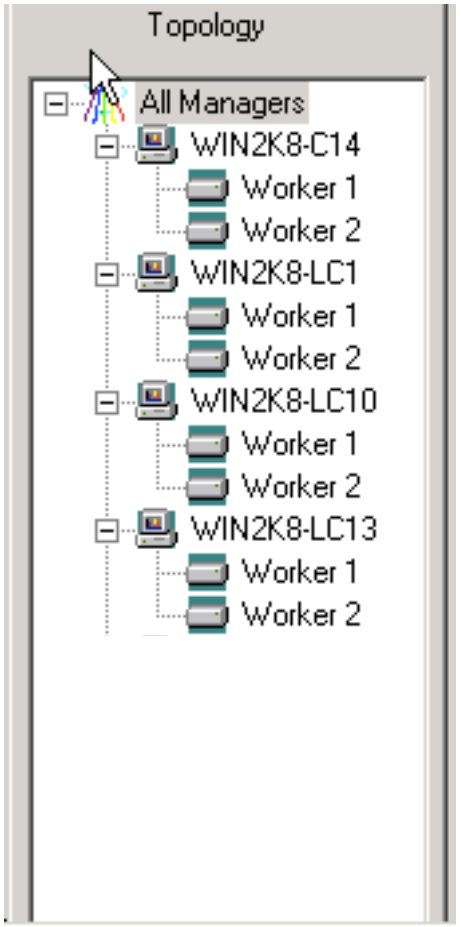
RAM based I/O Acceleration

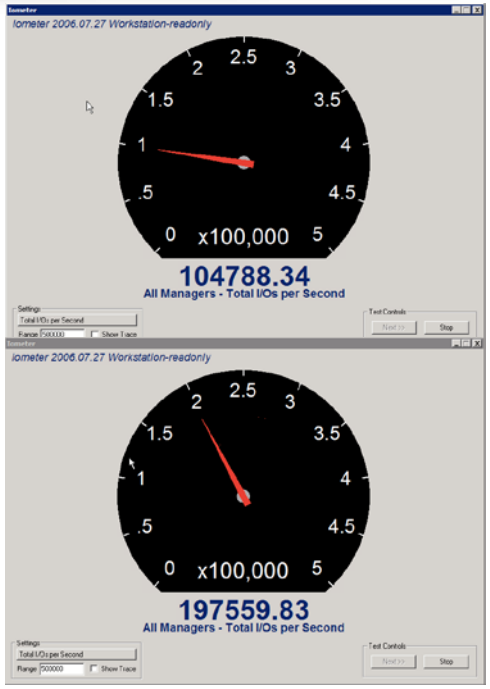
4KB random reads



RAM based I/O Acceleration

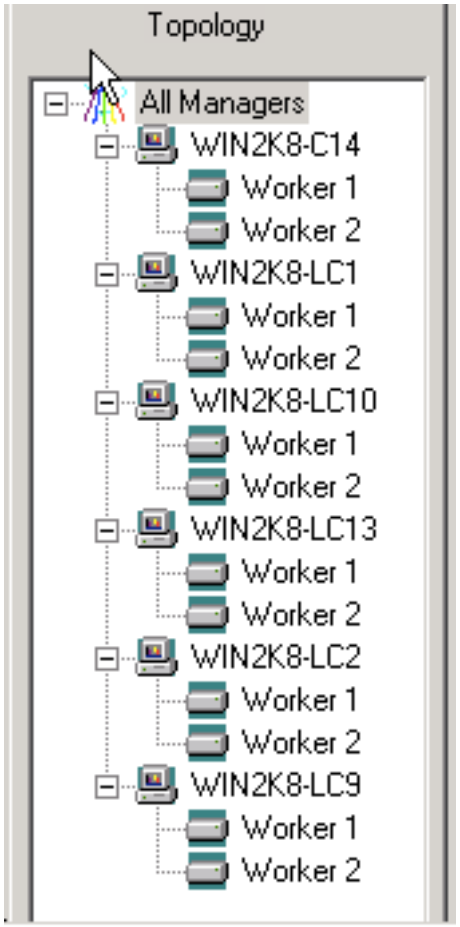
4KB random reads

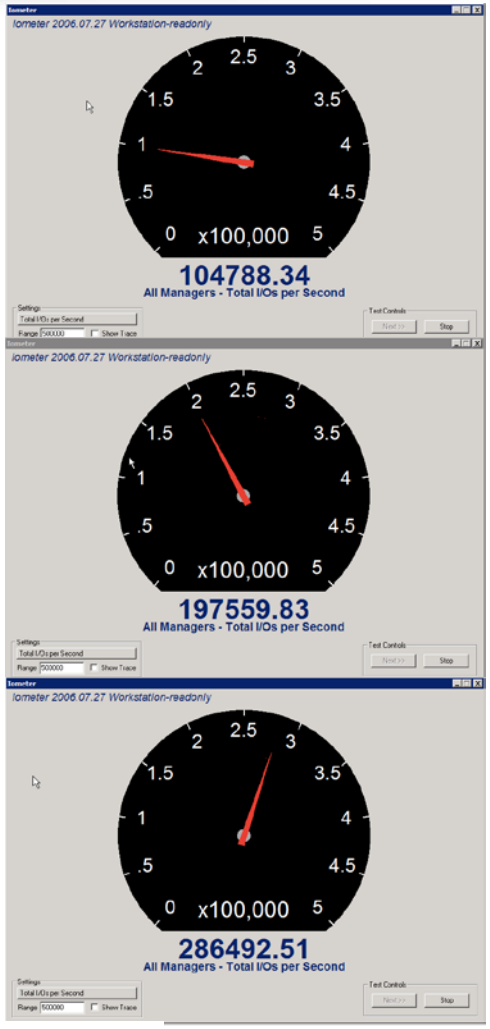




RAM based I/O Acceleration

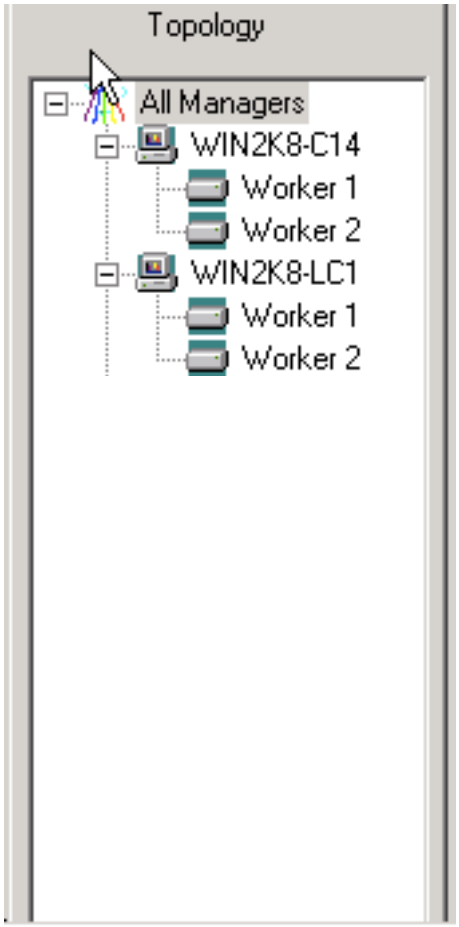
4KB random reads





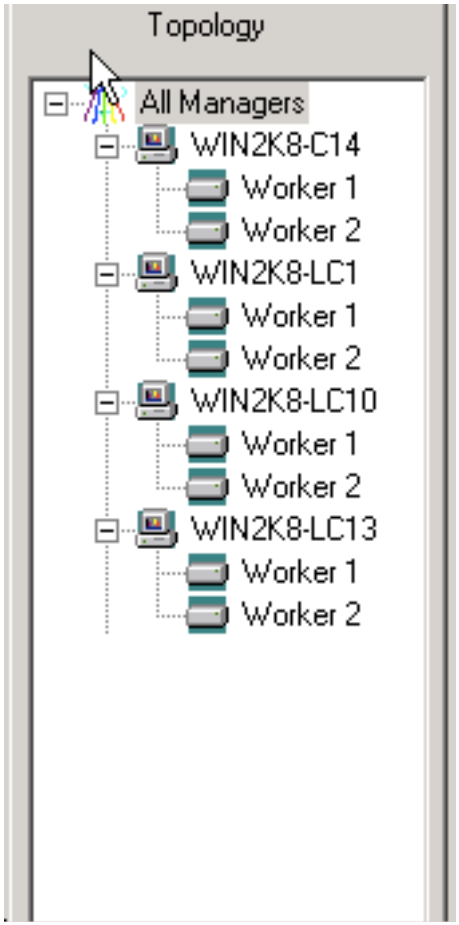
RAM based I/O Acceleration

64KB sequential reads



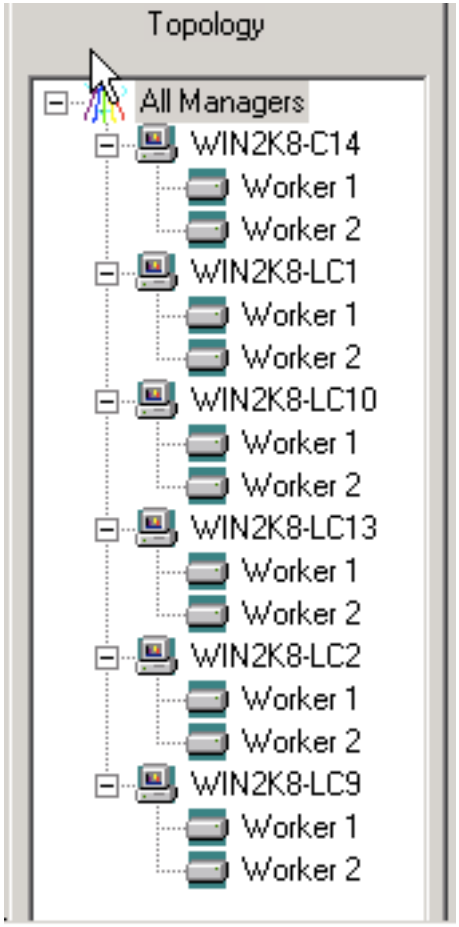
RAM based I/O Acceleration

64KB sequential reads



RAM based I/O Acceleration

64KB sequential reads



Potential Impact

- Allows up to 1TB of RAM per host
  - Cluster of 32 host, we can create a 32TB DFTM tier
- Infrastructure Level play
  - Applies to all type of applications running inside a VM, without any modification.
- Extremely easy to configure and manage.

Q&A

Thank You!

@unnojung
whj@pernixdata.com