



# Using Distributed Fault Tolerant Memory in Virtualized Data Centers

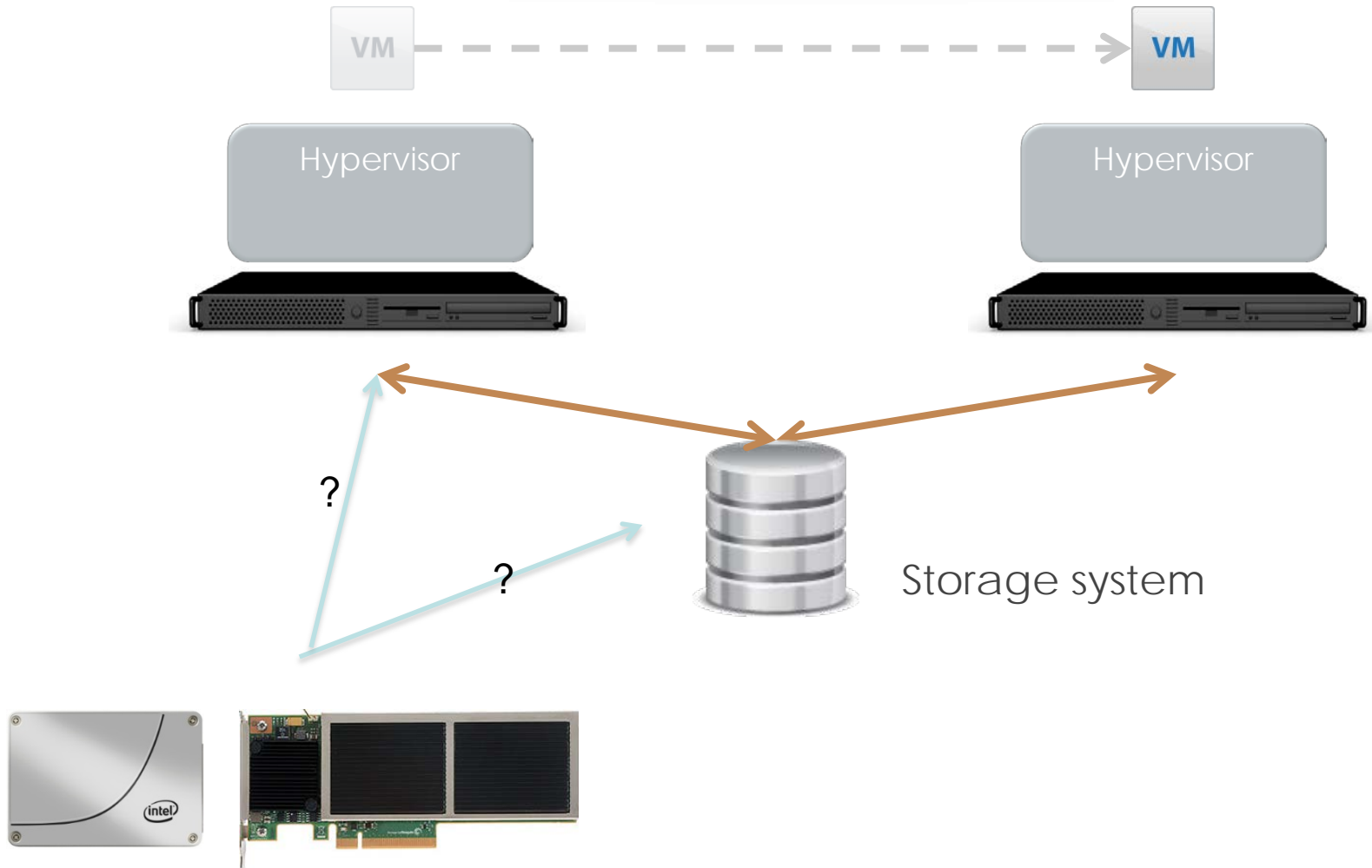
Woon Jung  
PernixData



# Agenda

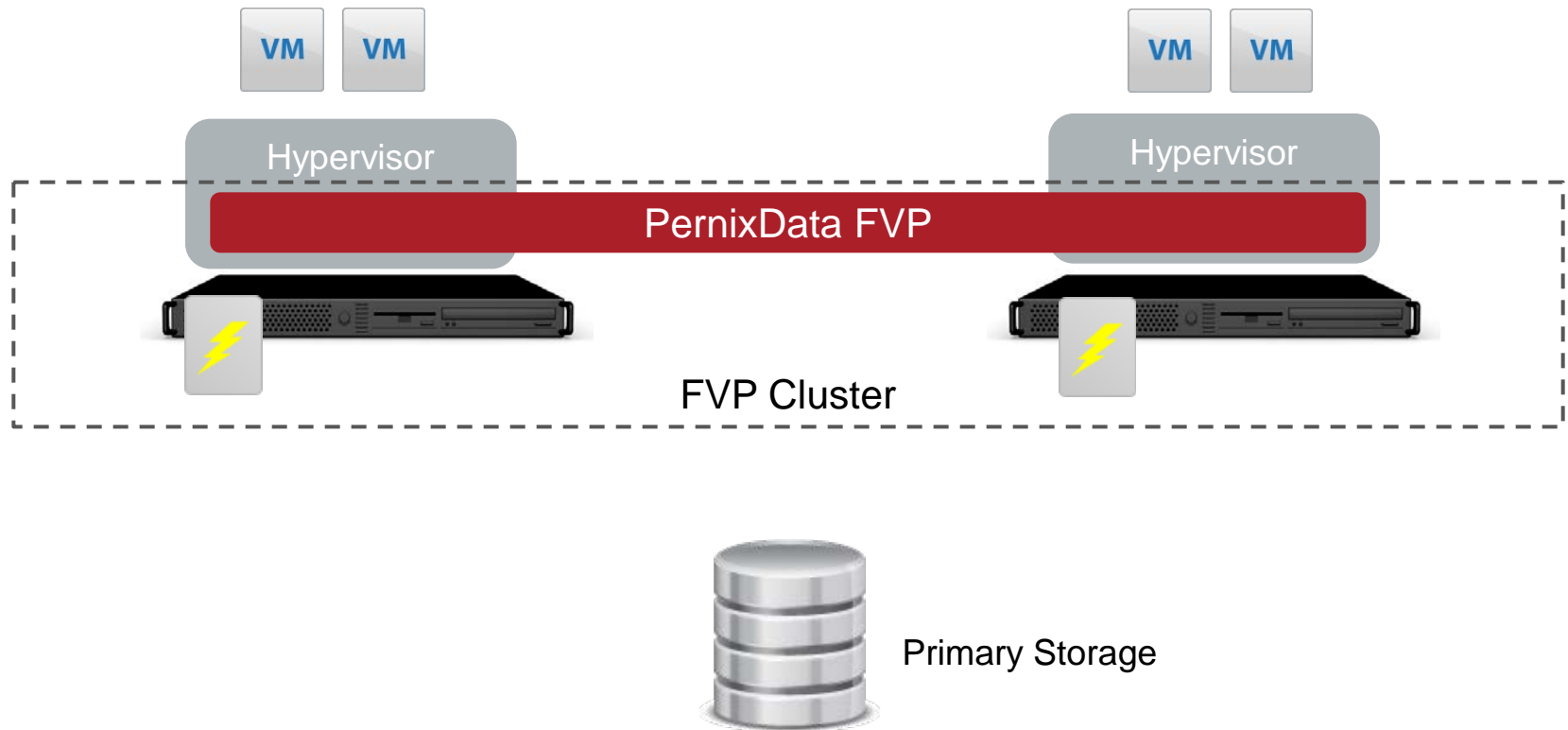
- Background
- Motivation for DFTM
- Survey of challenging problems with DFTM
- Q&A

# Background



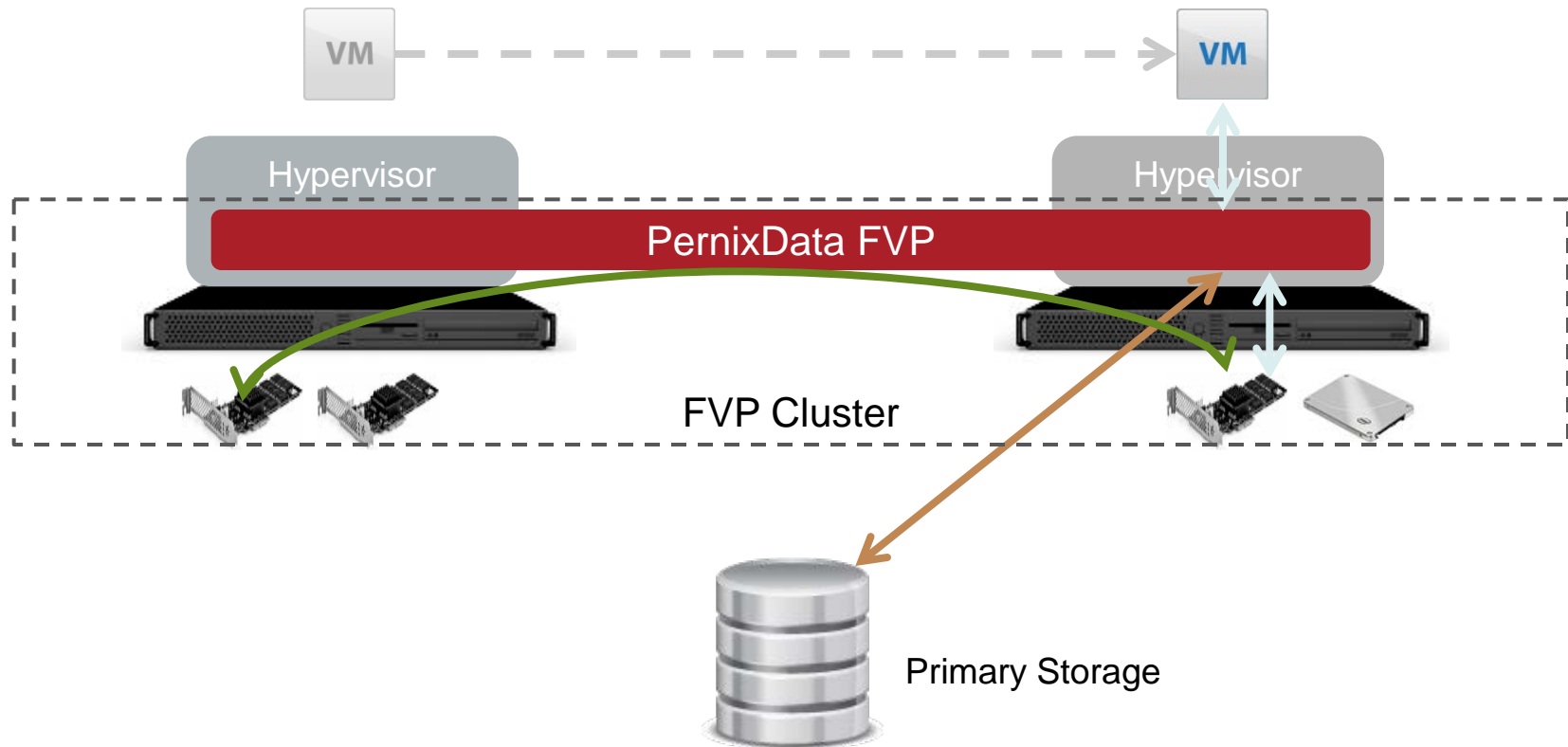
# It started with FVP...

## Hypervisor-based non-disruptive data acceleration

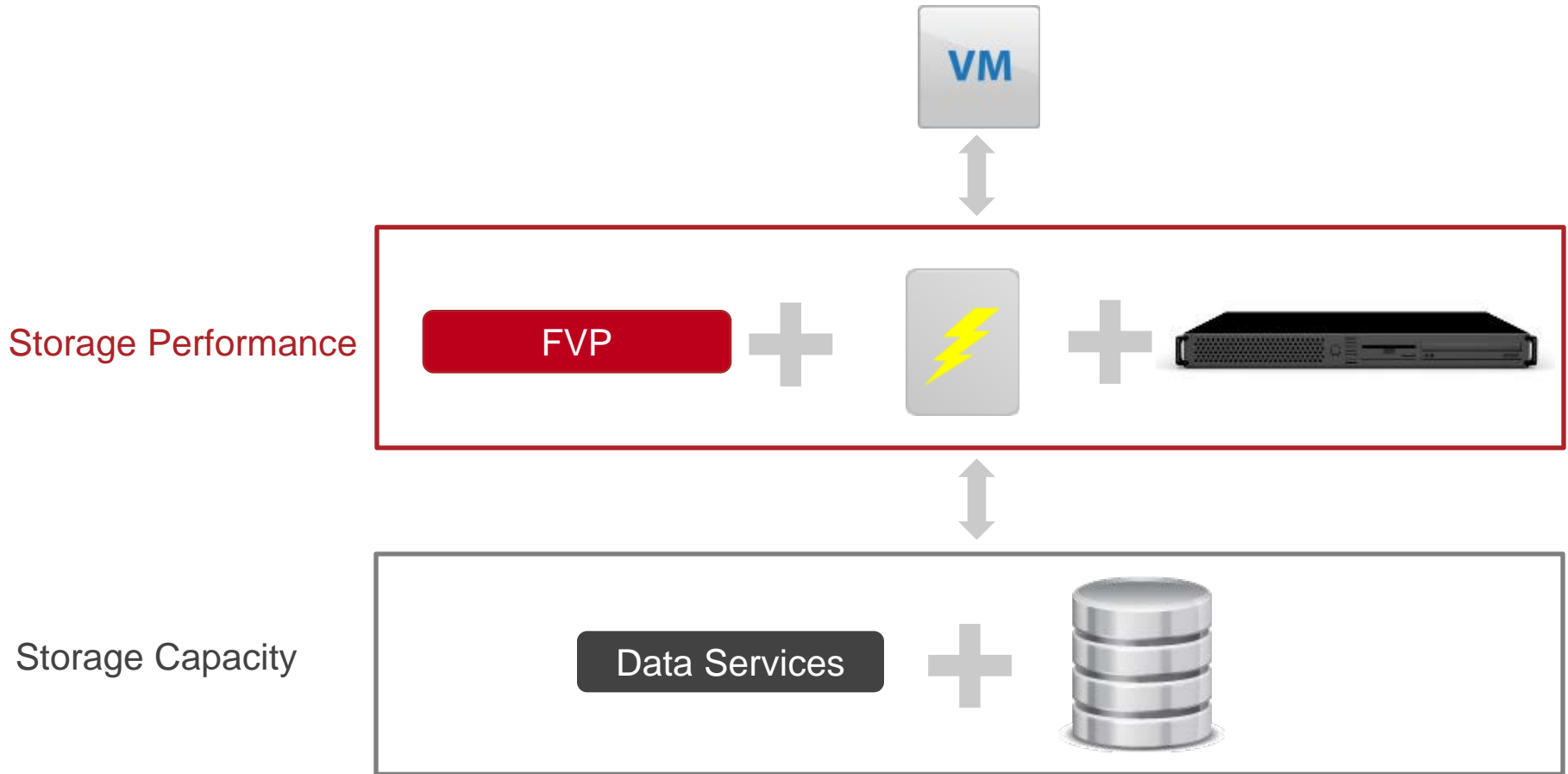


# Clustered Data Tier

## Server flash aggregation into cluster-wide pool

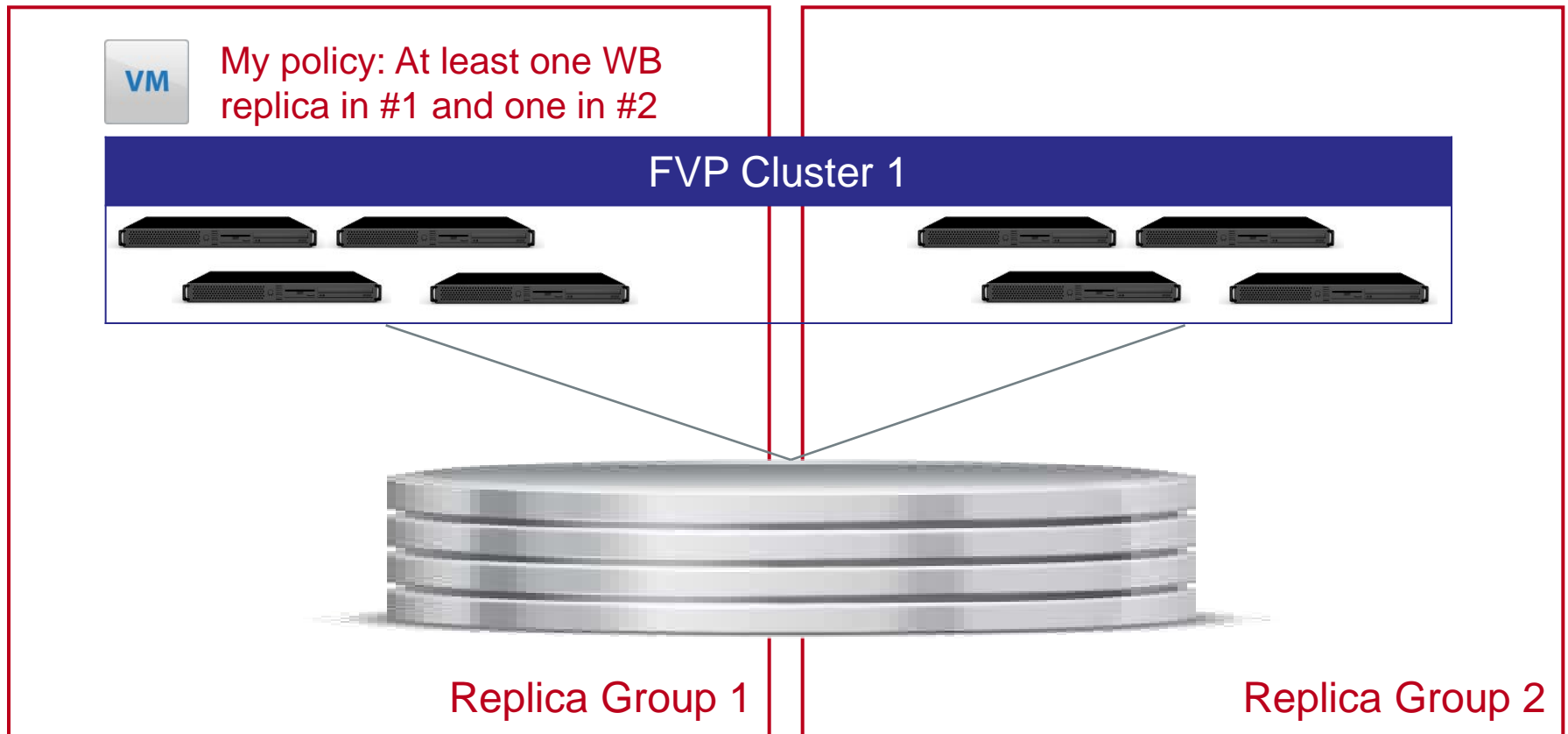


# Decoupled Architecture



- 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

Summary Virtual Machines Hosts DRS Resource Allocation Performance Tasks & Events Alarms Permissions Maps Profile Compliance Storage Views vShield PernixData

FVP Clusters Usage Performance Advanced

[Create FVP Cluster...](#) [Delete...](#)

Name	Acceleration Resources	Capacity	Datstores/VMs
FVP_01	4 Resources (from 4 of 4 Hosts)	1 TB	5 Datstores, 24 VMs

**FVP\_01** [Overview](#) [Configure](#)

**Resources**

4 of 4 with devices

4 Memory

**Status**

✔ **OK**  
The FVP Cluster is func

**Consumers**

Eligible <b>24 VMs</b> (24 VMs Active)	Not Eligible <b>0 VMs</b>	Blacklisted <b>0 VMs</b>
---	------------------------------	-----------------------------

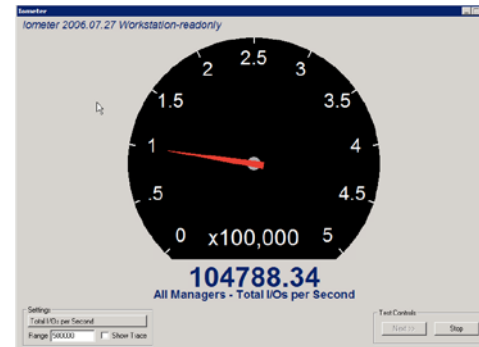
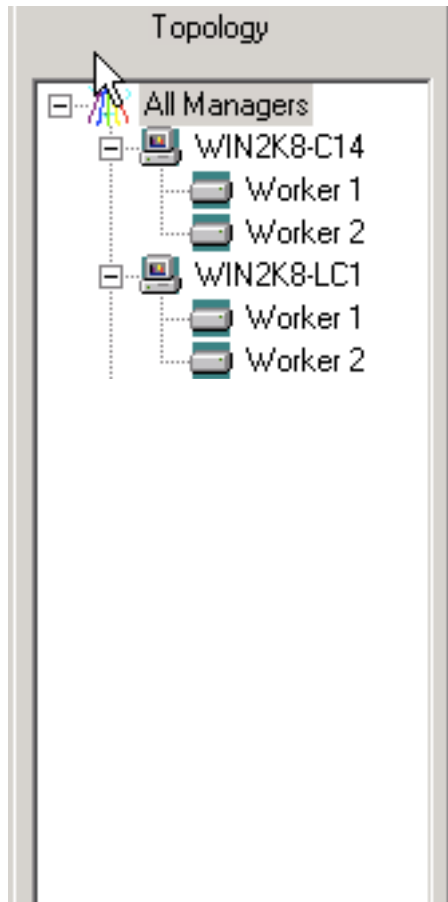
**Performance**

**Realtime**

- VM IOPS (Sum)
- VM Throughput (Sum)
- VM Latency (Avg)
- Acceleration Hit Rate

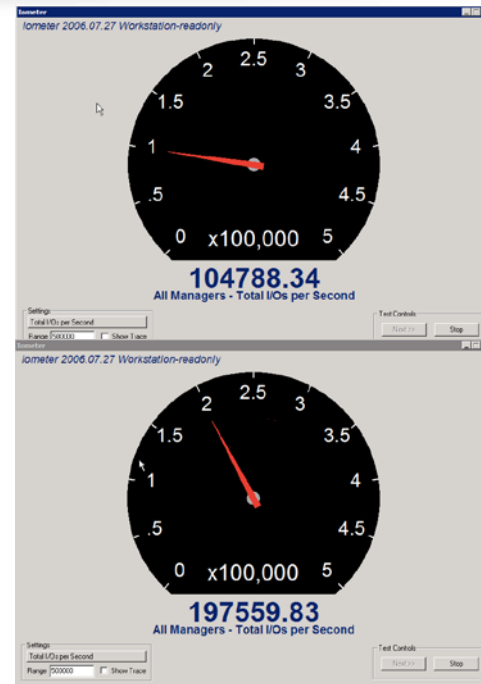
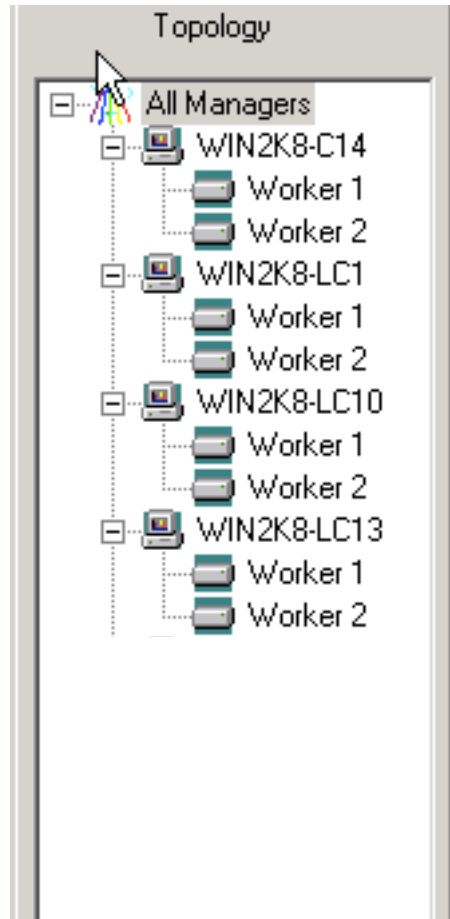
# RAM based I/O Acceleration

## 4KB random reads



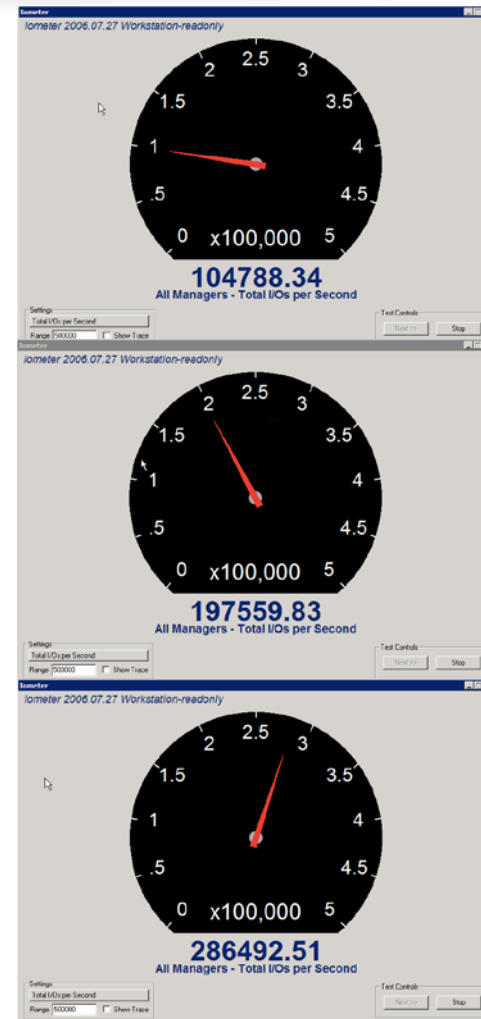
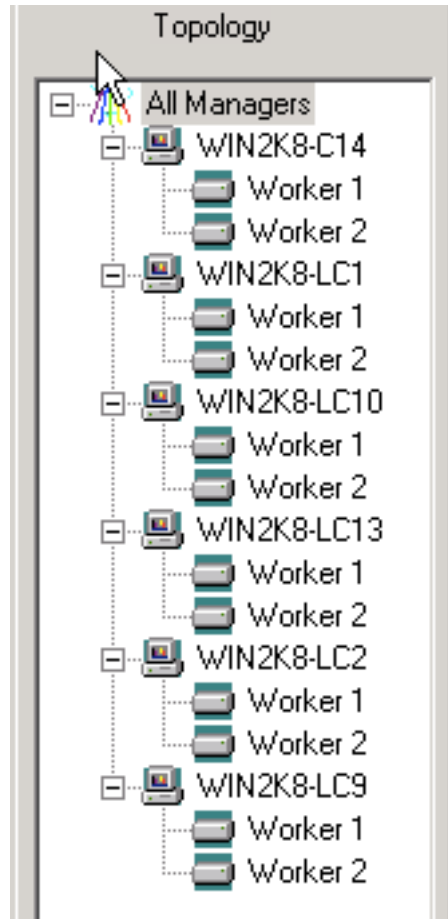
# RAM based I/O Acceleration

## 4KB random reads



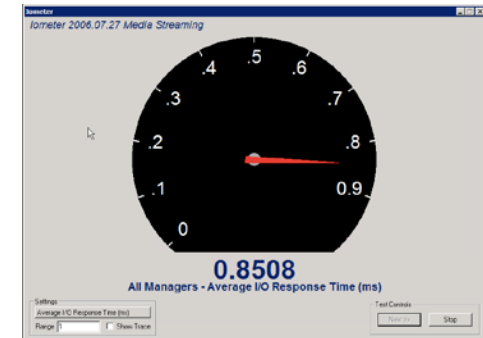
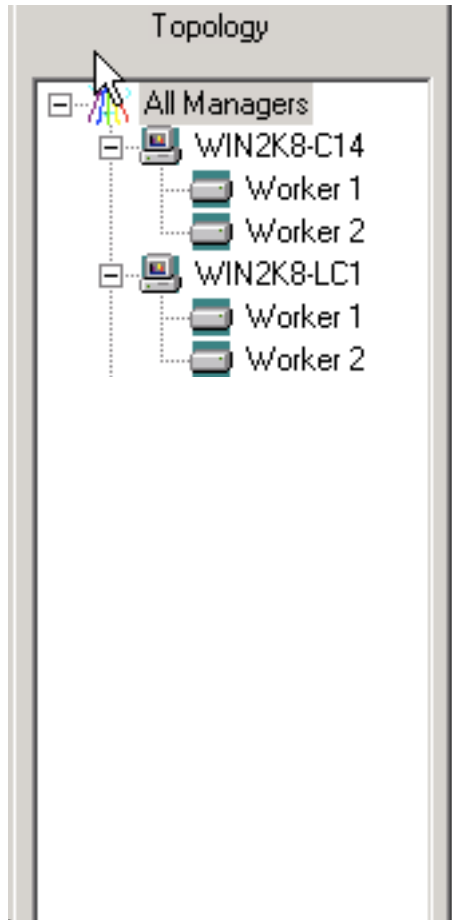
# RAM based I/O Acceleration

## 4KB random reads



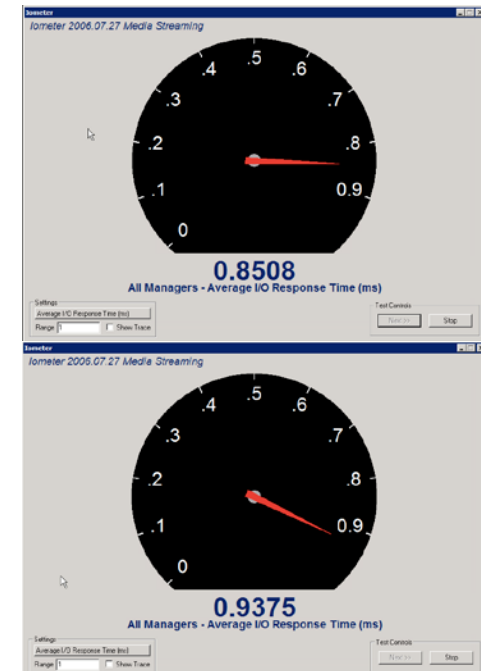
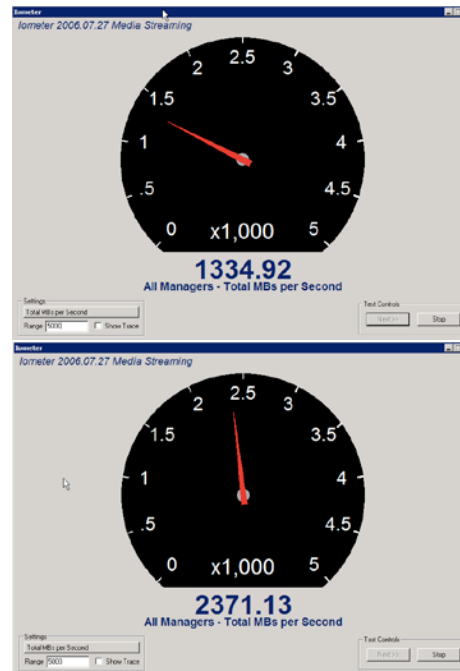
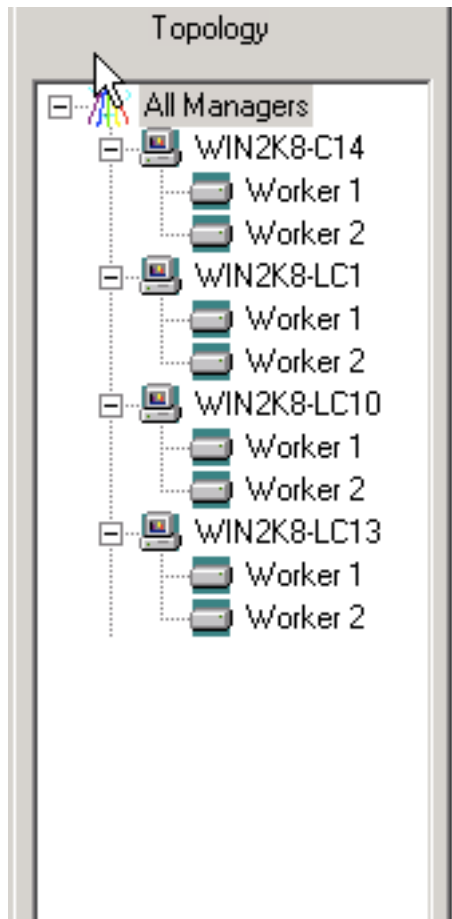
# RAM based I/O Acceleration

## 64KB sequential reads



# RAM based I/O Acceleration

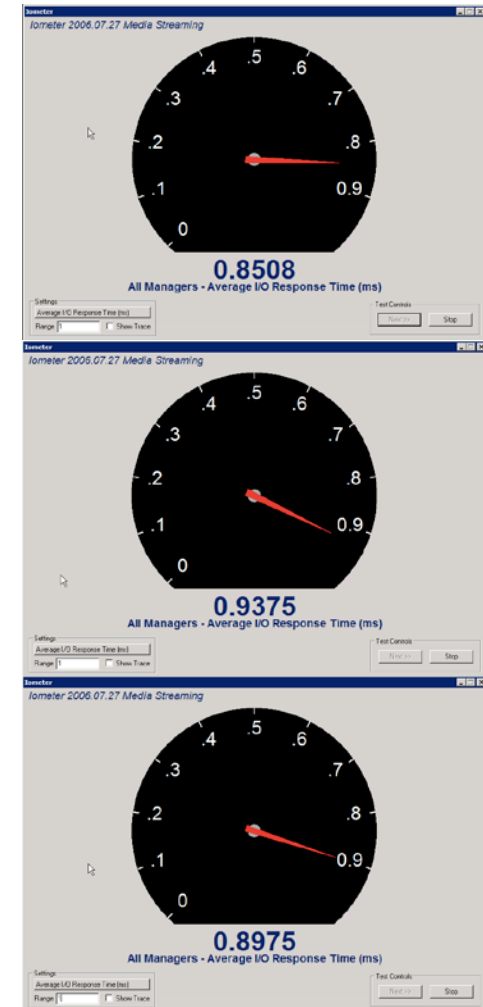
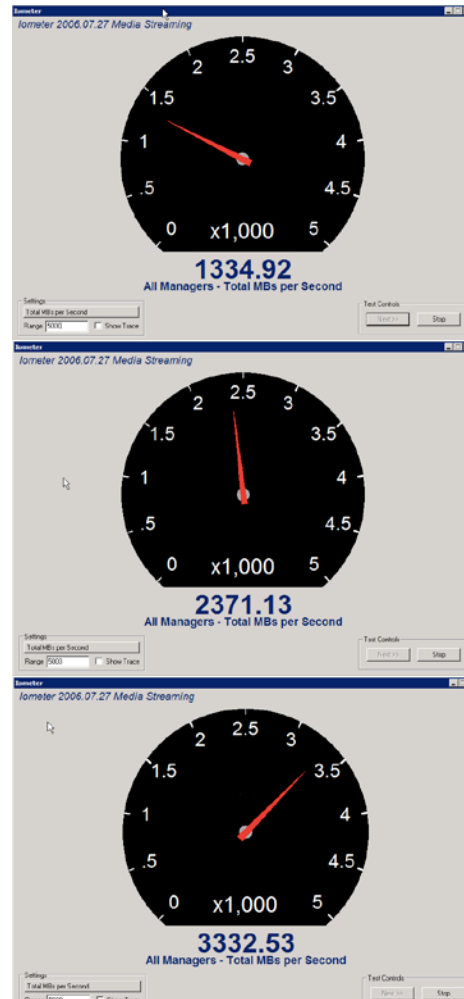
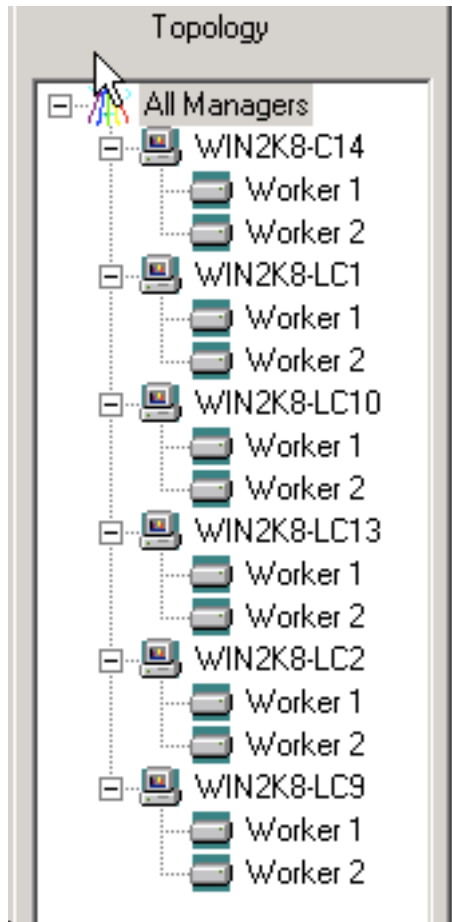
## 64KB sequential reads





# RAM based I/O Acceleration

## 64KB sequential reads



- 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.

**Thank You!**

**@unnojung**  
**whj@pernixdata.com**