



Go further, faster™

Flash on Compute Servers

Steve Kleiman
Senior Vice President & Chief Scientist





Agenda

I'd like to have an argument, please.
- Michael Palin



Technology Trends

- Multi-core CPUs
 - Virtual machines
 - More BW and IOPS required per node
- Flash
 - >5X better \$/IOP
 - Enough IOPS to service a node
 - <100usec read latency (vs >5,000 usec for disk)
 - Significantly worse \$/GB than disk
 - ~\$3/GB vs ~\$.10/GB



The (Original) Case for Networked Storage

- Benefit of aggregating disks
 - More random IOPS available than DAS
 - Additional access latency insignificant compared to disk access time.
 - Capacity sharing
 - Centralized data management
- Networks are fast enough
- A single app/host complex has limited ability to consume BW
 - A single commodity-based server can serve many hosts



Flash in Networked Storage

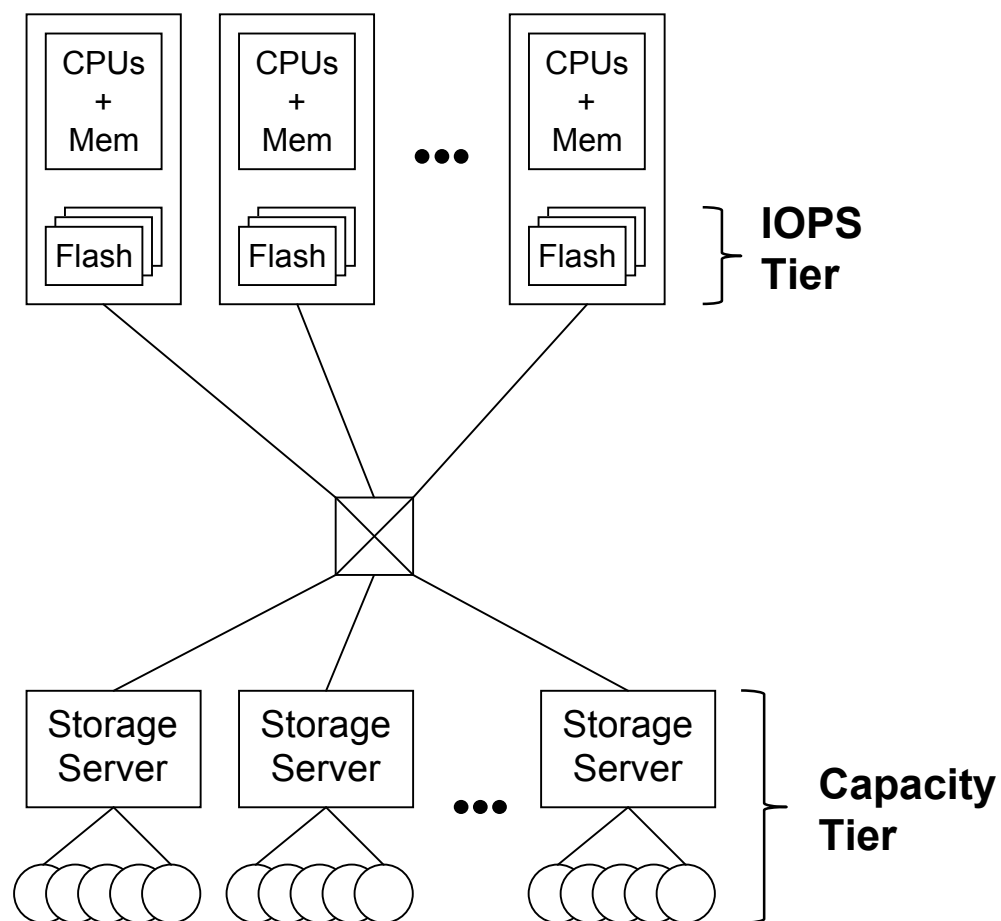
- Benefit of aggregating Flash
 - No performance gain from aggregating flash
 - IOPS are not scarce
 - Capacity sharing
 - Centralized data management
- Is networked storage fast enough?
 - Multi-core nodes with VMs: larger load
 - 5 SSDs: 40,000 IOPS, 250MBps
 - Storage controllers can add significant latency compared to Flash



Flash on Compute Servers

- Servers can already accommodate flash
 - Enough Disk or PCI slots for >250GB
 - 2-8X 6Gbps
 - System vendors will provide servers with flash built-in
- Flash is cheap enough to dedicate
 - 250GB < 10% of server cost
 - Even 250GB can be enough!
- Transition can be rapid
 - Enterprise grid trend already established

- IOPS Tier
 - Best \$/IOP
 - Low latency
 - Enough capacity for active data
 - The new “primary storage”?
- Capacity Tier
 - Best \$/GB
 - Deduplication, compression
 - Good serial performance
 - Conflicts with dedup?





Host-based Flash: DAS

- Single point of failure
 - Use normal backup for non-volatile data
 - Use application supported replication
 - Exchange, Oracle
- Booting, VM, temp files
- Distributed data management
- No capacity sharing



Host-based Flash: Primary Storage Service

- File or block
- Requires mirroring or RAID to peers
- Allows capacity sharing
- Microsoft & VMware will likely compete here
 - Complex integration otherwise
- Reliable primary storage is hard
 - Changes IT roles?
- Data movement between IOPS & Capacity tiers
 - Manual
 - Rules
 - On Demand



Host-based Flash: Cache

- File or block
- Write-through
- Write-back: non-zero RPO
- No performance penalty vs. standalone for reliability
- Maintains centralized data management
- Maintains storage management IT role
- Automatic “tiering”
- Can use standard protocols, but enhancements can be helpful



Data Management

- Spans IOPS and Capacity tiers
 - Automated data movement and global access
 - Manual placement is impractical at scale
- Need a technology-independent language to specify desired properties
 - Say what you properties you want not how do it.
- “Service Level Objectives” (SLOs)
 - Specifies properties such as:
 - Max latency, min BW, redundancy, DR etc.
 - Systems work to meet the objective



Summary: A New Storage Model?

- Host-based flash is the “IOPS tier”
 - Replaces high-performance primary storage in clouds. Not economical for secondary/archival
 - Provide software to allow reliability and easy access to all stored data
 - Cache “plug-ins” to: VMware, Hyper-V
- Networked storage is the “Capacity tier”
 - Optimize for best \$/GB with reliability
 - Optimize for rapid cache fill
 - Serial access of entire datasets (like tape HSM)
 - Use flash in Capacity tier as well
- Data management between IOPS and Capacity tier is important