SFD8 - Day 3 - Violin Memory

http://www.violin-memory.com/resources/

*Steve Dalton - SVP Product Strategy & Development (looks a little like Greg Behrendt)

Violin Product and Positioning Overview

NYSE: VMEM

“Flash Fabric Architecture”

Despite what people may think, we’re in the conversation on every flash deployment opportunity …

“AFA Market”?
Tier0 - performance
Tier1 - Primary
Data Reduced, VDI/VSI

Then Tier2, Tier3 and Tier4

“evolutionized”

1 platform, 3 segments - ViolinMemory_1platform

FSP Performance - extreme (no dedupe) - (sustained Usec latency, sustained high IOPS)
Thick / Thin only, async replication, snapshot, clone, simplicity

FSP Enterprise - mission critical, primary storage (performance, DR, HA, resiliency, cost)
All inclusive data services and data protection, search, CDP

FSP Capacity - price and performance (dedupe only) - (Cost, Optimised, Effective Capacity)
No thick, thin, async replication, snapshot, clone, simplicity

7300 Performance - 17TB - 140TB, >1M IOPS
7300 Enterprise - 17TB - 70TB
7300 Capacity - 17TB - 200TB

7700 Performance, Enterprise or Capacity - 17TB - 2PB, >2M IOPS, 0 RPO, 0RTO

Key Technology
Symphony - Single pane of glass mgmt console for flash arrays
Cloud Support - VMware, OpenStack
Concerto OS - Mature, feature rich and flash optimised

Flash array - mature, purpose-built and high performance

*Vikas Ratna, VP ATG and Product Strategy - https://www.linkedin.com/pub/vikas-ratna/0/329/376

Technology Deep Dive
- Flash Array Hardware
- Concerto Software
- Aria (Platform) Software
- Symphony Software

Building blocks
- NAND Flash (Toshiba)
- VIMM (Violin Intelligent Memory Module)
- vRAID Groups
= Violin 7300 (3RU) or the Violin 7700 (Controller and shelf model)

Flash Memory Fabric (VCMs / VIMMs)
- flash and flash mgmt
- delivers high sustained backend performance, high density and high resiliency
Array control modules
- controls flash memory fabric
- system level PCIe switching and health monitoring
- delivers high resiliency and availability
Active/Active Flash Gateways
- Intel-based - runs all SW mgmt
- PCIe connect to Flash Fabric
- Delivers all in one flash optimised enterprise features
IO Modules
- FC, 10GbE interfaces

*James (Jim) Bowen, Technology Director
System Hardware Architecture - [picture]

Modular, highly available system design
redundant and parallel data paths
independent control plane
Violin Flash Fabric
- multiple paths between each VIMM and VCM
- VIMM Tree dynamically reconfigures to handle component failures and / or upgrades

Flash Fabric Architecture
Multipath VIMM Fabric
- 60 Data VIMMs + 4 spare VIMMs
- triple-ported VIMMs provide multiple data paths between VIMM and VCM
- Any VCM can address any VIMM
- Each VCM manages 3 vRAID groups
VCM Failure Protection

- in event of VCM failure, ownership of VIMMs is transferred to another VCM
- no disruption of data access
- system will operate with 1-4 VCMs
- failover also used during VCM NDU

VIMM Failure protection
- in event of VIMM failure, available spare is allocated for immediate use
- vRAID handles rebuild of spare without manual intervention
- 4 spares can be used for any spare

Fabric can handle failure of up to 3 VCMs and 4 VIMMs, with vRAID can tolerate up to 16 VIMM failures without data loss

Performance through parallelism
- multiple HBAs connect to customer SAN
- Flash optimised driver breaks inbound IO requests of any size (4K, 8K, etc) into multiple 4K requests
- 4K IO requests are distributed across vRAID groups and VCMs
- HW vRAID controller in VCM splits each parallel 1K IOs across all VIMMs in vRAID group
- 12 vRAID groups x 5 VIMMs per RG = 60 parallel IOs
- system hw capable of 2M IOPS in 3RU
- Backend currently supports 760K sustained IOPS (100% writes); capable of 1.5M sustained IOPS

*Chris Zhang, Technology Director

“Sustained performance matters”
Peak IOPS vs sustained IOPS
- write cliff after 20 minutes writes. 5x - 10x difference
- better sustained performance = predictable IOPS and low latency
- many factors affect sustained performance (IO load, R/W mix, SW and HW interaction, FTL, RAID algorithm, etc)

How does Violin ensure sustained IOPS and latency?
1. Intelligent GC (garbage collection) and App integration [notes in photo]
2. VCM (vRAID - optimised for Flash) [notes in photo]

vRAID vs SSD Based Array? [photo?]
Violin Read Handling
- Reads never blocked by GC or write activity
- system level orchestration enables sustained low latency for mixed workloads

SSD-Based read handling
- arrays get a read in one RAID stripe and have to wait until current Write completes to read
- garbage collection also inflicts latency spikes

[Peak and Sustained Performance - numbers table photo]

VIMM Architecture [photo has more notes]
Designed for flash optimised data path for storage arrays

Patented violin IP on Flash endurance and wear levelling
Designed for high resiliency
Designed for quick litho transition
Hot swappable

Concerto Software Tech Details

Simplified Concerto Software Stack [photo]
Full Concerto Software Stack [photo]

UI runs on HTML5

*Tim Stoakes - Technology Director

Thick / Thin Data Path

Dedupe Modular Data Path [photo of Tim presenting]

Data Reduction Write Process - Buffering
- stage incoming IOP data in system DRAM and protect with vRAID in VIMM DRAM (no need for UPS, special NVRAM HW, etc)
- provides low and consistent latency
- reduces performance impact of sub-block and mis-aligned writes
- dedupe processing continues from DRAM, no flash reads

Data Reduction Write Process - Inline
- small fixed dedupe chunk size (smaller chunks give better dedupe ratio, no requirements to reconfigure apps)
- all in-memory chance lookup, no IO
- stub update (update the mapping of client LBA to point to existing data location, aggregated with nearby updates, IO is vRAID friendly)

Data Reduction Write Process - Unique Data
- compression (tuned for balance of compression / decompression speed and ratio)
- chunk aggregation (improved locality, no vRAID sub-stripe write, VIMM GC friendly)
- data LUN write (single large flash write of aggregated chunks)
- stub update (update the mapping of client LBA to point to new, unique data location)

Data Reduction Read Process
- data services layered on top
- read unique data location from stub LUN metadata (may not exist - thin provisioning, no secondary hash lookup required)
- many read in flight at once

*Vikas - Concerto Software - DR Services

Replication Overview
- 2 modes: Periodic (Delta) replication and CDR
- Over IP

- App consistent recovery points when using snapshot agents
- policies for security and bandwidth optimisation

Stretch cluster setup review (7700 only) - [photo]

Don’t scale infinitely, but can scale “tremendously”

*Pat Balakrishnan, Sr director platform sw and cloud - Aria Software - skipped

[demo on mgmt software - Symphony]

vCenter plugin
VASA provider

The dashboard looks pretty cool (and configurable)