FLASH MEMORY SUMMIT 2018
The True Performance of Flash Storage
DEVELOPING LOW-LATENCY DATA SERVICES ON NVME-OF SHARED STORAGE

Presented by

Chaan W Beard
TOSHIBA

• Toshiba and AccelStor have been working jointly on NVMe-oF technology – we are grateful to Toshiba for the wonderful work and contributions they have made

• Our other SSD partners use Toshiba Memory in their products
NVMe-oF Changes the Rules for All-Flash Arrays

- NVMe-oF enables efficient data transfers over a network, changing the rules for all-flash arrays.

Diagram showing the transition from traditional storage protocols (SCSI, TCP, IP, Eth) to NVMe-oF for improved performance.
NVMe-oF Decouples Compute and Storage Nodes

- Scale **Performance** or **Capacity** on demand
- RDMA over Converged Ethernet

**Feature-Rich Data Services**
- NeoTopaz™
- NeoTopaz™
- NeoTopaz™

**Compute Nodes**
- Virtual AFA Controller
  - Scale-out for Performance

**Storage Nodes**
- Scale-up for Capacity

**NVMe-oF JBOF**
**NVMe-oF Tools and Development Kit Overview**

4KB Random Write IOPS (Million)/Per Compute Node

- **Fast and Low Latency**
  - SPDK without data services
  - SPDK + AIO + LVM + Linux NVMe Driver in Physical Machine

- **Flexible Architecture**
  - FlexiRemap® + SPDK + AIO + Linux NVMe Driver in Physical Machine

- **Feature-Rich Data Services**
  - SPDK + AIO + LVM + Linux NVMe Driver in Virtual Machine
  - FlexiRemap® + SPDK + IF-AIO in Virtual Machine

---

Flash Memory Summit 2018
Santa Clara, CA
Virtual All-Flash Array, vFlexiArray™ Introduction

- NeoTopaz™ Virtual Controller
- Virtual NVMe SSD Namespace
- Virtual AFA creation
- Data Services same as physical Array (Snaps, Clones, thin/thick provision etc.)
- Fast, Flexible, and Feature-Rich
- Hyperscale now becomes a reality
- High Performance and Data Protection can now be a Guaranteed Service
NeoTopaz™ Virtual Controllers

- Real virtualization not logical
- All the data services run in NeoTopaz™ Virtual Machine Storage OS
- Isolated, secure and reliable
- Communication via 100GB RoCE
- Delivers 1M IOPS @4K random writes, and scales up to 40M IOPS
NeoTopaz™ High Availability Solution

Shared-Nothing HA

Data Storage

Node A
Controller

Data

Node B
Controller

Data Storage

1. Data flow of vFlexiArray™

2. RDMA over Converged Ethernet

3. FlexiRemap®

4. RDMA over Converged Ethernet

5. NVMeoF JBOF

6. NVMeoF JBOF

7. NVMeoF JBOF

8. NVMeoF JBOF

VirtualMachine

NVMeoF@SPDK

NeoTopaz Controller

IF-AIO

NVMeoF@SPDK

VirtualMachine

NVMeoF@SPDK

NeoTopaz Controller

IF-AIO

NVMeoF@SPDK

Data Storage

Flash Memory Summit 2018
Santa Clara, CA
Virtual SSD Namespace

NeoTopaz™

NVMe-oF JBOF

NeoTopaz™

NVMe-oF JBOF

NVMe-oF JBOF
vFlexiArray™ Creation

Need Performance or Capacity?

Server Farms with NVMe-oF Initiator

Feature-Rich Data Services
NeoTopaz™ NeoTopaz™ NeoTopaz™

Virtual AFA Controller
Scale-out for Performance

RoCE
Scale-up for Capacity

NVMe-oF JBOF
NeoTopaz™
Fast, Flexible, and Feature-Rich
Fast, Flexible, and Feature-Rich

Build a high capacity - low IOPS vFlexiArray™ using Virtual SSD Pools with required SSD characteristic (Fewer SSD with more capacity)

Add in HA via configuration

Build a high IOPS, small Capacity vFlexiArray™ with tailored, high performance SSD

So granular even a single large capacity and high IOPS SSD can be shared by multiple vFlexiArray™ via the Namespace feature

Tailor the vFlexiArray™ any way you desire with any SSD spec

Feature-rich data service: (FlexiRemap®, FlexiDedupe™, FlexiVirtualArray™, FlexiSnap™, FlexiClone™ etc.)
Hyperscale Becomes a Reality
Hyperscale Becomes a Reality

Relationship between controller and SSD liberated by NVMe-oF network itself

Unlimited extension of Capacity

A vFlexiArray™ can dynamically add any number of SSD in the pool

No limit of the number of JBOF ports on a vFlexiArray™ since it is virtualized

Expansion Simple and easy, just plug in the new JBOF to the RDMA bus, the management system detects and updates using REST API automatically!
High Performance & Data Protection

Guaranteed!
High Performance and Data Protection Guaranteed

Our technology guarantees there will be no IOPS degradation using our vFlexiArray™

Performance also tailored by adding computing nodes as required

Every vFlexiArray™ can provide up to 1 Million IOPS with either WRITE or READ random operations (@4K) with FULL Data Protection and HA built in

No longer need expensive storage hardware to service full HA needs
vFlexiArray™ Use Cases

Multi-tenant data centers

ISPs

CSPs
User Interface
Thank You
Free Software Resources

- Install Linux Centos 7.2 or above for target-side
- Install Linux Centos 7.2 or above and upgrade kernel to 4.8 for initial-side
- Download SPDK: lock-free software
- Multiple modules in SPDK: nvme driver, nvmeof target, iscsi target.
- User-layer application
- Replace the role of the Linux driver for nvme
Missed Features

Initial-side NVMf driver for popular os dist.

In-cap read response packet

Runtime attaching LUN/ns to target

LUN across multiple drives: using namespace

Host (Initial-side)

All Flash Array (Target-side)

SPDK NVMf target

 Влад Мемори Сумит
Flash Memory Summit 2017
Santa Clara, CA