Developing NVMe-oF Virtual All-Flash Array

Alan Lin
Chief Solution Architect, AccelStor Ltd.
Data Center Is Transforming

**Hyper Converged Infrastructure**
- Application
- Management
  - CPU / Memory / Storage
  - CPU / Memory / Storage
  - CPU / Memory / Storage
- Network / SDN

**Challenge**
- Simply Deployment
- Remove Complexity for the SAN
- Virtualized CPU and Storage

**Software Defined Data Center**
- Application
- Management
- Compute Farm
- Network / SDN
- Storage / SDS

**Software define everything**
- Compute farm can be use application and various of purpose
- Storage/Network virtualize.

**Challenge**
- CPU/Storage as a block unit, utilization will be an issue
- Performance might
NVMe-oF Realize
The Composable Infrastructure

- Leverage the existing compute node
- Virtualize the storage controller in compute node

- Ethernet NVMe JBOF can reduce the overall cost
- More flexible for the storage controller.
NVMe-oF Decouples Compute and Storage Nodes

Scale out for Performance

Scale-up for Capacity

RDMA over Converged Ethernet

Server Virtualized, VDI, Database, AI/ML

Feature-Rich Data Services

VM

NeoTopaz™

Application server

Compute Farm

Virtual AFA Controller

Storage Nodes

VM

VM

VM

SSD

SSD

SSD

SSD

NVMe-oF JBOF

Application server

Compute Farm

Virtual AFA Controller

Storage Nodes
Flash Memory Summit 2019
Santa Clara, CA

Proof-of-Concept Topology

- 2 Clients for fio benchmark
- 4 NVMe SSD as one group
  Each group has 2 namespaces
- Create 4 vAFA to connect with Ethernet JBOF
Each vAFA hardware spec.
- CPU Intel Scalable CPU 6126 12 core
- Memory 192GB
- Mellanox 100GbE CX5
- Each of the vAFA can have the stable 4KB random write at 750~850K
- Total aggregated*4 vAFA performance is 3.1M IOPS
- 8 Ethernet NVMe SSD.
- 2 namespaces for each array group
Performance Result and Dashboard
Data Center and Enterprise Storage Feature Requirement

- FlexiRemap - FMS 2016 Best of Show Technology / software
- Implement the interrupt-free Asynchronous IO (IF-AIO) between SPDK and NeoTopaz
- Deliver up to 1M IOPS in 4KB and under 100us

I/O Request

NVMe-oF with SPDK

IF-AIO

NeoTopaz with FlexiRemap

IF-AIO

NVMe SSDs

NeoTopaz NVMe-oF vAFA

NeoTopaz
FlexiRemap

2(N +1) Protection
Volume Mgmt.
Data Reduction
RoW Snapshot
Link Clone
Thin Provisioning
High Availability
Remote Replication
NeoTopaz™ High Availability Solution

Data flow of vFlexiArray™

1. User
2. Node A
3. RoCE
4. Node B
5. Data Storage
6. Data Storage
7. FlexiRemap®
8. RDMA over Converged Ethernet
9. User
10. User
11. User

Shared-Nothing HA

Data Storage

Controller

VirtualMachine

NVMeof@SPDK

NeoTopaz Controller

IF-AIO

FlexiRemap®

NVMeof@SPDK

VirtualMachine

NVMeof@SPDK

NeoTopaz Controller

IF-AIO

FlexiRemap®

NVMeof@SPDK

Data Storage

NVMeof JBOF

NVMeof JBOF

NVMeof JBOF

NVMeof JBOF
De-centralized management design
Key Benefits of Virtual NVMeoF AFA

- Leverage Existing Compute Nodes
- On-demand
- Flexible Deployment
- HW-independent