NVMe over Fabrics (NVMe-oF) For Containers

Scaling Cloud-Native Applications With Elastic And High-Performance Storage

Sudhakar Mungamoori
Director Storage Solutions, Toshiba Memory America Inc.

Venkat Ramakrishnan
VP Engineering, Portworx
What is NVMe over Fabrics (NVMe-oF™)?

A new storage networking protocol, optimized for flash, exclusively for NVMe commands

Based on Industry-standard Networks

Networked flash, with the performance and latency of native NVMe

Flash Memory Summit 2018
Santa Clara, CA
No matter what size SSD you choose, it may miss the mark; tough to optimize configs.

Pool the physical drives, dynamically create any size “logical drive” for each node

With NVMe-oF - Any service can run on any server
Each service gets “just the right amount” of high performance, low latency storage
Toshiba KumoScale™ Shared Accelerated Storage Software

- **Management Engine**
- **Virtualization / Abstraction**
  - **Fabric Module** (front end)
  - **NVMe Module** (back end)

- **Network Driver**
- **NVMe Driver**

- **NVMe-oF Network**
- **NVMe SSDs**

- **Kubernetes®**
- **OpenStack®**
- **Mesosphere®**
- **Intel® RSD**

RESTful API
Container Orchestration Transforming Datacenter Cloud Architectures

- Modernize on Commodity Hardware
- Enable Operational Agility
- Build Elastic Private and Hybrid Clouds
- Efficient Utilization of Infrastructure
Challenges With Stateful Containers

- Modern Applications are Data intensive and Require Stable, Persistent and Performance Storage
- Expensive to Migrate Stateful Applications Across Nodes
- DAS Single-Node Solution for Persistent Storage is Inelastic
- Dynamic Storage Scaling is Impossible in DAS Architectures
NVMe-oF Solves Container Storage Problems

- Dynamic provisioning of NVMe storage namespaces per container.
- Fast, isolated, highly available storage volumes for containers – Scheduler Integrated.
- Storage orchestration that optimizes for performance and cost
- Application aware live migration for performance and scalability.

Flash Memory Summit 2018
Santa Clara, CA
Portworx Introduction

Storage for DevOps

Portworx takes existing infrastructure
integrates with schedulers
and enables self-service IT for stateful containers

Fingerprints, aggregates, tiers, and monitors storage media (elastic container storage)
Integrates provisioning, control, and cluster scale (up to 1000 servers)

Integrates with schedulers:
- Mesosphere
- Kubernetes
- Swarm
- Any Scheduler

STATEFUL + STATELESS

Volumes

Dynamically creates volumes with schedulers, resizes, encrypts, and moves across clouds
KumoScale NVMe-oF : Portworx API Integration

Application
(1) Volume create request
   to_priority=high

Container Orchestrator
(2) Volume create request
   to_priority=high

Portworx
(5) Create volume in 'n1'

Host 1
(4) Attached numa namespace 'n1' to host

KumoScale

KumoScale APIs -
Create namespace 'n1'
Grant host 1 access to the namespace

Host 2
(2) Attached numa namespace 'n1' to host

KumoScale APIs -
Grant host 2 access to the namespace 'n1'

When Host 1 goes down
Compute Node
- One or more Pods w/ containers per physical server.

Portworx Container
- PX dynamically manages pool of storage (Kumoscale NVMe-oF volumes) to hold data for stateful containers
- PX provides replication of data across volumes.

K8S Master
- Manages scheduling and orchestration
  Container replication and restart to defined policy.

KumoScale Node
- One or more Kumoscale NVMe-oF Storage Nodes.
Kubernetes integrated with Portworx and Kumoscale
Thank You

Visit Toshiba Booth #307 For Live Demo of NVMe over Fabrics

Sudhakar.Mungamoori@taec.toshiba.com
Venkat@portworx.com