




# Managing Flash In OpenSDS For Cloud Native Frameworks

**STEVEN TAN**, OpenSDS TSC Chair, VP & CTO Cloud Storage Solution - Huawei @stevenphtan

**ANJANEYA 'REDDY' CHAGAM**, OpenSDS TSC, Chief SDS Architect – Intel



# Overview

 Open SDS Platform	<b>Control/ Management Plane</b>	Standard REST API, Single Pane Management, Policy-Based, Storage and Data Services, Orchestration and Automation
	<b>Data Plane</b>	Data Reliability, High Availability, Data Protection, Data Mobility, Data Reduction, I/O Performance
	<b>Storage/Service Layer</b>	DAS (HDD/Flash), SAN, NAS, AFA, Commodity, Private Cloud, Public Cloud

**Decoupled control plane from data plane and storage/service layer**

# The Projects

## SUSHI

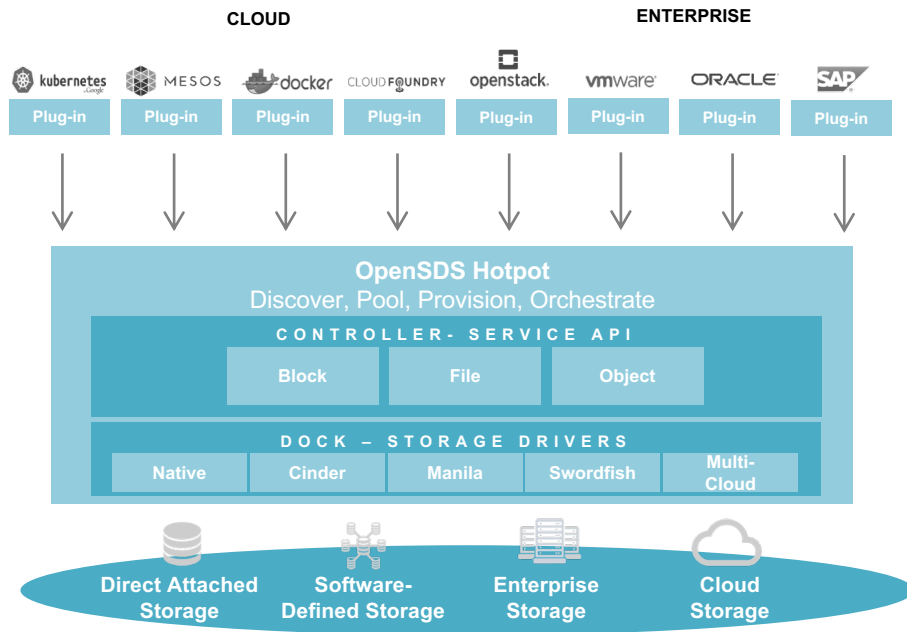
### The Northbound Plug-ins Project

Common plug-ins to enable OpenSDS storage services for cloud and application frameworks

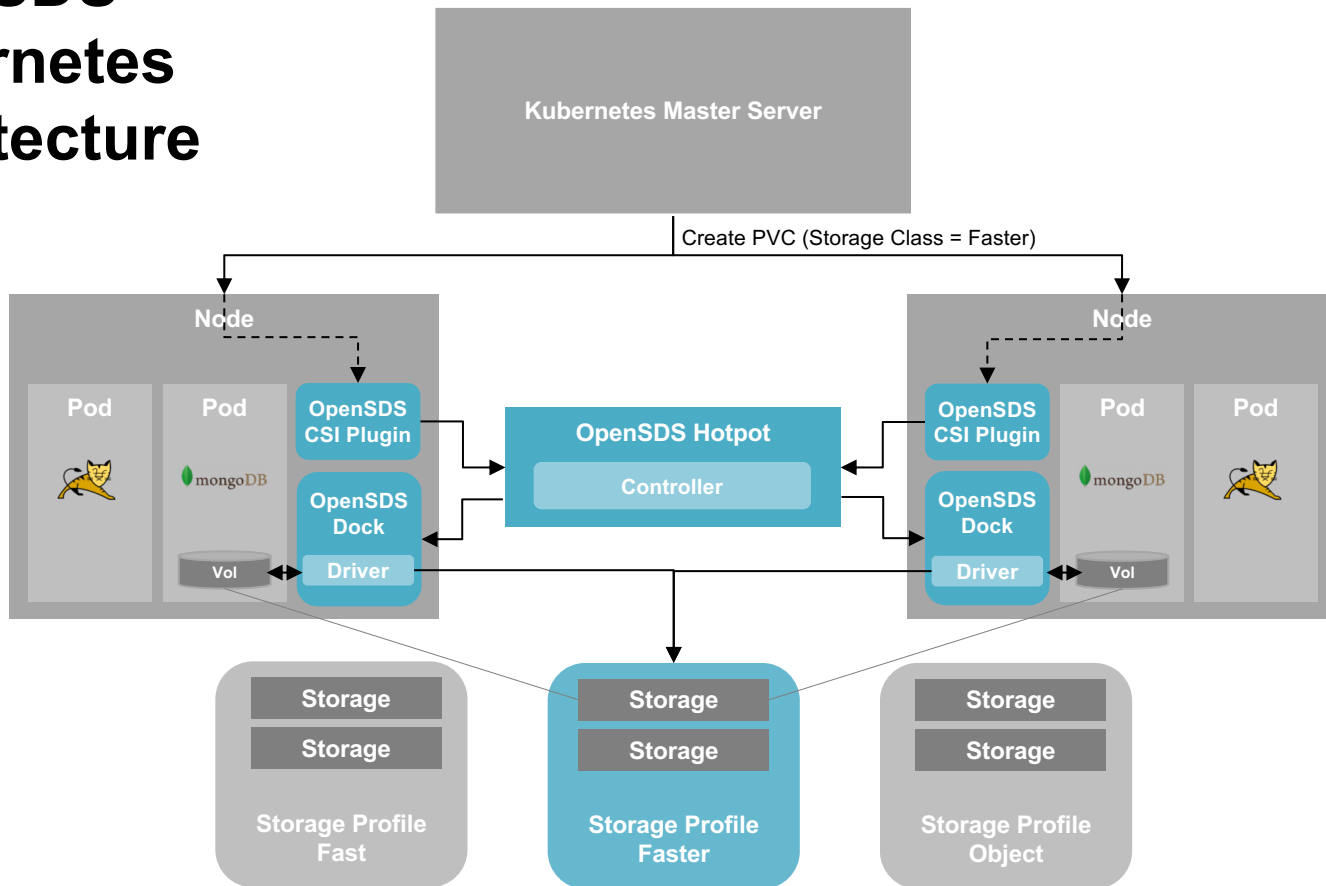
## HOTPOT

### The Storage Controller Project

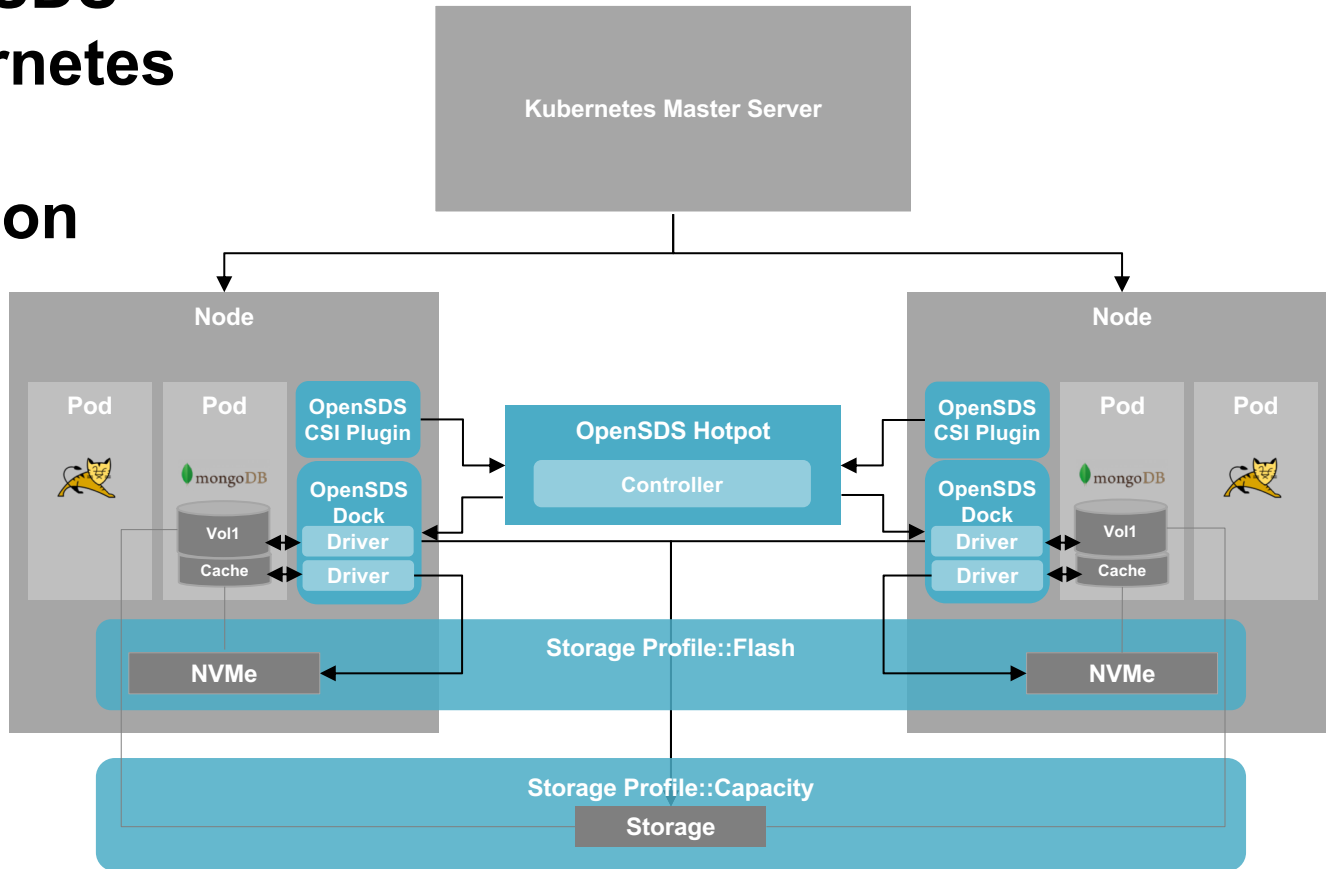
Single control for block, file, and object services across storage on premise and in clouds



# The OpenSDS Kubernetes Architecture



# The OpenSDS Kubernetes Flash Solution



# NVM Express (NVMe)

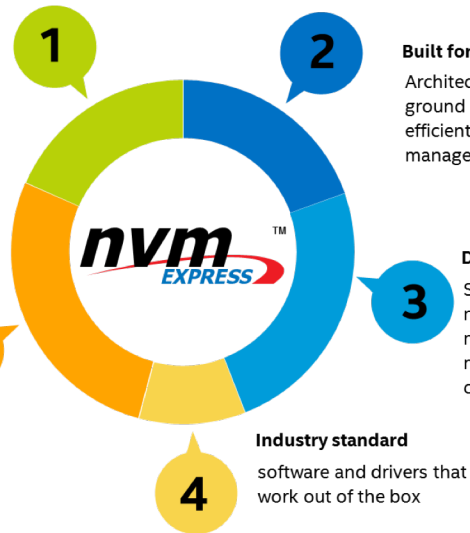
Standardized interface for non-volatile memory, <http://nvmexpress.org>

## What is NVMe?

NVM Express<sup>®</sup> (NVMe) is a standardized high performance software interface for PCI Express<sup>®</sup> Solid State Drives

## Ready for next generation SSDs

New storage stack with low latency and small overhead to take full advantage of next generation NVM



## Built for SSDs

Architected from the ground up for SSDs to be efficient, scalable, and manageable

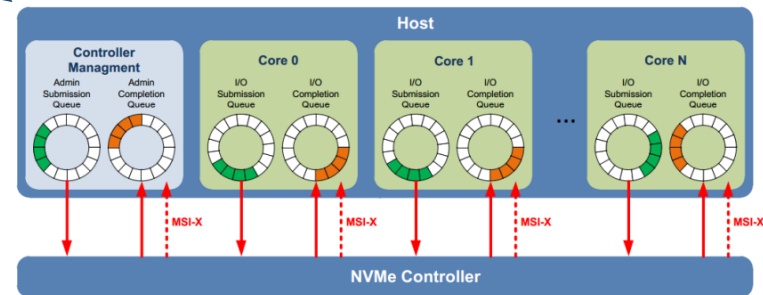
## Developed to be lean

Streamlined protocol with new efficient queuing mechanism to scale for multi-core CPUs, low clock cycles per IO

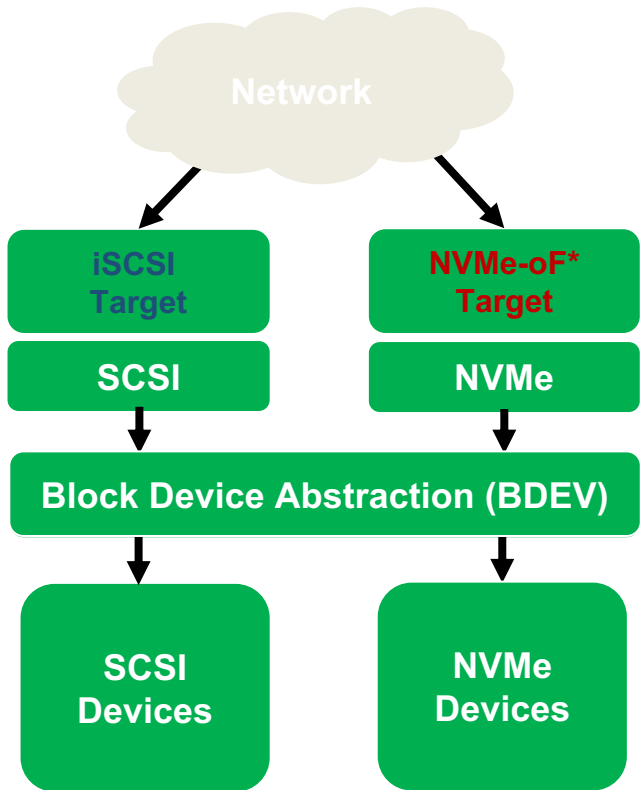
## Industry standard

software and drivers that work out of the box

- Performance: 1 GB/s per lane.. 4 GB/s, 8 GB/s, 16 GB/s per device..
- Lower latency: Direct CPU connection
- No host bus adapter (HBA): Lower power ~ 10W and cost ~ \$15
- Increased I/O opportunity: Up to 40 PCIe lanes per CPU socket
- Form factor options: PCIe add-in-card, SFF-8639, M.2, SATA Express, BGA

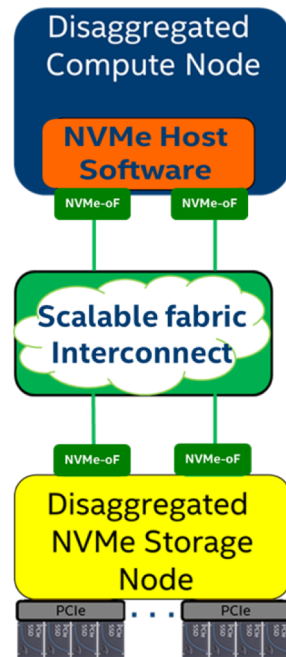


# Remote Access To Storage – iSCSI and NVMe-oF



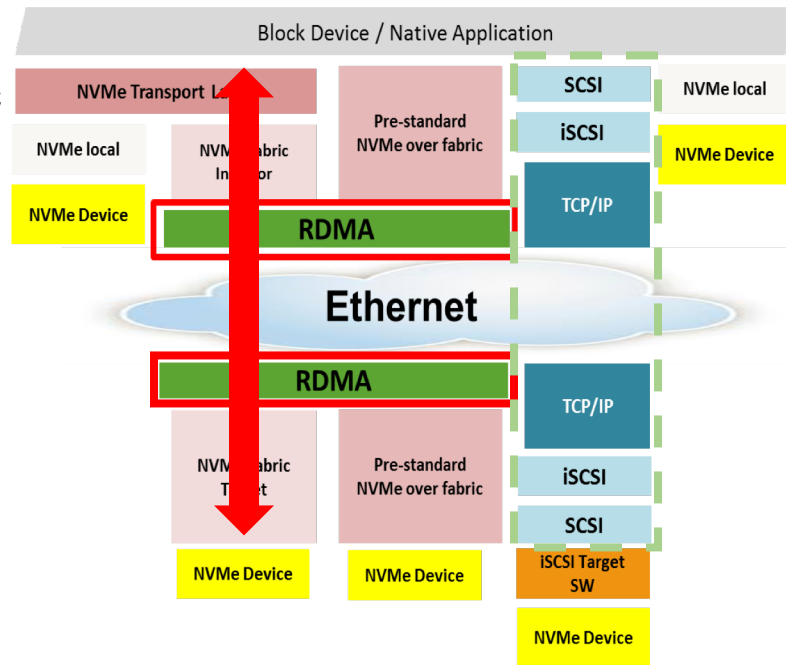
- NVMe-over-Fabrics
  - NVMe commands over storage networking fabric
- NVMe-oF supports various fabric transports
  - RDMA (RoCE, iWARP)
  - InfiniBand™
  - Fibre Channel
  - Intel® Omni-Path Architecture
  - Future Fabrics

## Disaggregated Cloud Deployment Model



# NVMe-oF: Local NVMe Performance

- The idea is to extend the efficiency of the local NVMe interface over a network fabric
  - Ethernet or IB
  - NVMe commands and data structures are transferred end to end
- Relies on RDMA for performance
  - Bypassing TCP/IP
- For more Information on NVMe over Fabrics (NVMe-oF)
  - [http://www.nvmexpress.org/wp-content/uploads/NVMe\\_Over\\_Fabrics.pdf](http://www.nvmexpress.org/wp-content/uploads/NVMe_Over_Fabrics.pdf)





# NVMe-oF: Kernel Initiator

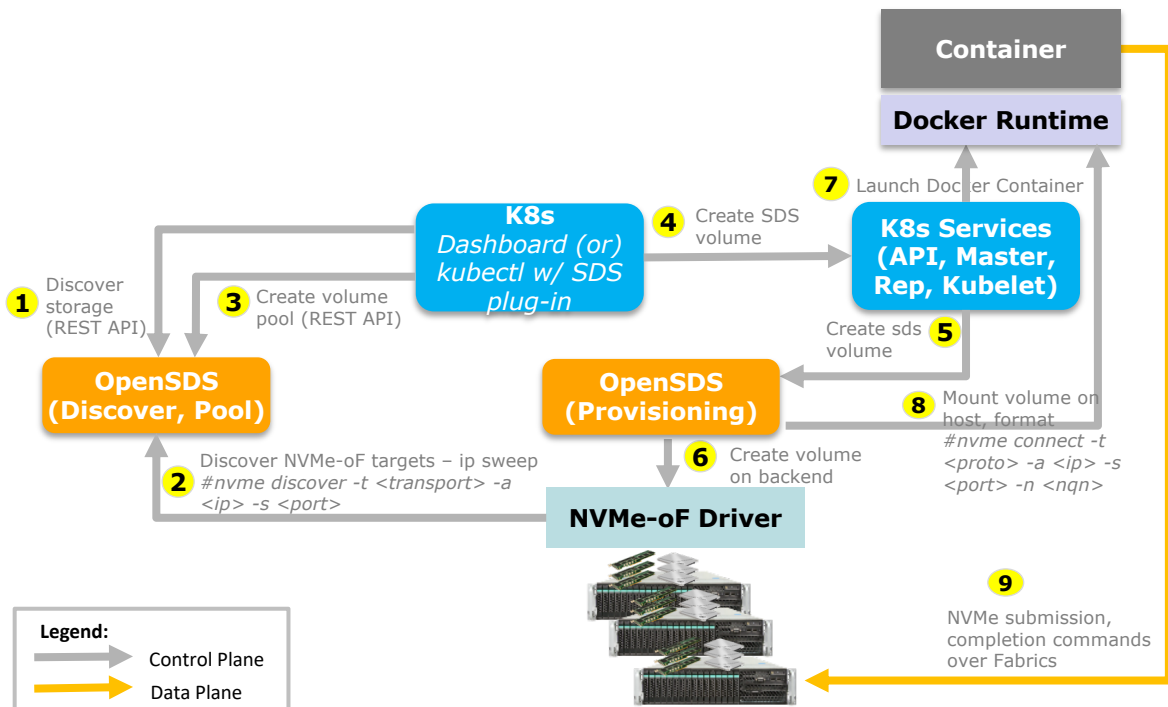
- Uses nvme-cli package implement the kernel initiator side
- Connect to remote target
- `nvme connect -t rdma -n <conn_nqn> -a <target_ip> -s <target_port>`
- `nvme list` - to get all the nvme devices

# NVMe-oF: Kernel Target

- Uses nvmetcli package implement the kernel target side
- nvme save <file\_name>– to create new subsystem
- nvme restore – to load existing subsystems

```
''  
"subsystems": [  
  {  
    "allowed_hosts": [],  
    "attr": {  
      "allow_any_host": "1"  
    },  
    "namespaces": [  
      {  
        "device": {  
          "nguid": "ef90689c-6c46-d44c-89c1-4067801309a8",  
          "path": "/dev/nvme0n1"  
        },  
        "enable": 1,  
        "nsid": 1  
      }  
    ],  
    "nqn": "testnqn"  
  }  
]
```

# NVMe-oF in OpenSDS



## Work In Progress

- Linux Kernel Driver
- *nvme* commands for connect
- Drive assignment
- Specs include target info

## 2019 Plans

- Pooling
- Rack aware scheduling
- User mode target (SPDK)
- NVMe over TCP/IP

# OpenSDS Roadmap v0.17

## 2017H2 ZEALAND

- Kubernetes FlexVolume
- Vol CRUD
- Standalone Cinder Integration
- CSI Support
- Ceph, LVM

## 2018H1 ARUBA

- OpenStack
- Replication Array-Based, Host-Based
- Dashboard
- Storage Profiles
- Enumeration
- Block Storage
  - Cinder Drivers
  - Ceph
  - LVM
  - Huawei: Dorado

## 2018H2 BALI

- S3 Object
- Multi-Cloud Data Control
- Multi-OpenStack
- Monitoring
- Storage Groups Snapshots, Replication
- Southbound Swordfish\*
- NVMeoF Preview

## 2019H1 CAPRI\*

- File Share
- Analytics
- Lifecycle
- Migration
- Data Protection
- NVMeoF


## 2019H2++

- Optimization
- Tiering
- Security
- Sharing
- Networking
- SCM



# THANK YOU

 <https://www.opensds.io>

 <https://github.com/opensds>

 [info@opensds.io](mailto:info@opensds.io)

 [@opensds\\_io](https://twitter.com/opensds_io)

FIND OUT  
MORE

OpenSDS @ SNIA SDC  
Santa Clara, Sep 24-27

BE A  
MEMBER

Accepting New Members  
Vendors And End Users Welcome