Software Defined Storage meets NVMe Over Fabrics

Tom Lyon
DriveScale, Inc

Flash Memory Summit
August 2019
Who is Tom Lyon?

Co-Founder and Chief Scientist, DriveScale

Distinguished Engineer, Cisco

Distinguished Engineer, Sun Microsystems [employee #8]

Systems Architect, Amdahl Corp

Co-Founder and Chief Scientist, Nuova Systems [-> Cisco UCS]

BSE in EECS, Princeton Univ

CTO, Nokia Internet Communications

Hacking on UNIX/Linux since 1975

Founder & CTO, Ipsilon Networks [-> Nokia IP Routing & Security]

UNIX, UTS, SunOS, NFS, SPARC, SunLink, IP Switching, ATM, UCS, VFIO...

@aka_pugs
What is Software Defined Storage

Clustered/Distributed software providing storage services
• (sure, there’s a few non-clustered)

Block, File, Object, or Database services
Provide their own data protection/replication/management
• Data placement critical for availability – rack awareness

Hundreds of Examples:
• Ceph, StorageOS, OpenEBS, … (aka Server SAN)
• GlusterFS, Quobyte, RozoFS, …
• Scality, Swiftstack, Minio, …
• Hadoop, Cassandra, CouchBase, Greenplum, …
Clusters can grow to thousands of nodes
ALL of these written for generic servers with DAS – Direct Attach Storage
NVMe Over Fabrics – Major Use Cases

1. SAN back-end fabric
   • Replace SAS or FC in large SAN systems back-end fabric
   • No visible changes to rest of datacenter
   • RoCE/RDMA viable here

2. SAN front-end fabric
   • Typically NVMe over FC between servers and SANs
   • Breathing new life into FC, but nobody new is buying it
   • Gradual iSCSI to NVMe/TCP transition

3. DAS replacement
   • DAS is cheap and fast, but inflexible
   • Rampant over-provisioning
   • Never the right storage-to-compute ratios
Direct Attached Storage

- DAS is chosen because of low cost and scalable performance
- Per IDC, “internal” storage market is growing twice as fast as “external"
- $$$s are 2x, unit growth is much greater
- To replace DAS, you must meet cost and performance goals
- But that’s not all!

I Love DAS!  I Hate DAS!
Data Placement

For SDS applications, one cannot replace DAS volumes with arbitrary target volumes

- Can completely destroy resilience guarantees
- All 3 copies of data on a single drive? A single JBOF? A single rack

Assignment of hosts to targets must be topology aware and application aware

- Some SDS have placement plugins (e.g. Hadoop) to help this

Most storage vendors have zero understanding of this problem

- Their answer: don’t use SDS
Out-of-the Box: Network Resilience

- In-the-box “fabrics” are way more reliable than out-of-the-box
  - Net-work is an oxymoron
  - Beware of “Fabric” thinking – real world networks are ugly
    - Dedicated storage fabrics are dead and buried
- Link aggregation
- Multi-pathing
- Load balancing
  - Storage flows are *heavy*
  - End-to-end host/target coordination
Security

• Standards never come with security - until “later”

• DAS is private to server
  – Only one way to get to the storage

• Authentication/Authorization
  – Who are the parties?
  – What are they allowed to do?

• Data privacy
  – At rest: encryption is easy, key management is hard
  – In flight: encryption? VLANs?
  – Connection hijacking
Composable Infrastructure: Definition

Source: Moor Insights & Strategy
What is DriveScale

DriveScale was born to radically transform how modern applications are deployed in the data center.

Seeing how hyperscale companies were reaping the benefits of analytics and AI with massive data sets on their scale-out infrastructure, DriveScale set out to make it viable for any enterprise to deploy this infrastructure on-premises and reap the same reward.

DriveScale designed a software platform that creates a way to run these same data-intensive applications unchanged with cloud scale and efficiency at a significantly lower cost.
Questions?

Stay in touch!

@aka_pugs

pugs@drivescale.com