Benefits and Challenges of Using SmartNICs in Distributed Shared Storage

Kirill Shoikhet
Chief Architect, Excelero
Who is Kirill Shoikhet?

Chief Architect at Excelero

- 20 years of software development and architecture experience in various aspects of high performance and distributed systems including storage, networking, performance analysis and diagnostics.
- MSc in Computer Sciences from the Technion.

www.excelero.com
Why Distributed Shared NVMe Storage?

- Flash is expensive
- Local drives are underutilized
- NVMe allows minimal overhead compared to local access
- RDMA becomes ubiquitous especially in rack-scale environments

But many existing, virtual & legacy environments don’t support NVMe
SoC/RDMA NIC + ARM Cores + NVMf = ?

Disaggregated A

Disaggregated B

Converged – the focus of this talk
SmartNIC + NVMf = Benefits!

1. Allows supporting NVMf in existing, virtual or legacy environments
2. Offload storage stack to SmartNIC - higher application compute density
3. Better security and support model
   1. Applying security patches independently
   2. Separate support matrix for host and SmartNIC environments
... = Benefits + Challenges 😞

1. How to connect host storage stack to NVMf stack on a SmartNIC?
2. How to prevent data copying?
3. How to handle errors?
4. How to tackle large scale?
Assumptions

• Legacy or virtualized hosts – NVMe stack but not NVMf
• A SmartNIC has access to the hosts RAM via PCIe
• Shared drives are exclusively accessed via SmartNICs
1. SmartNIC 1 exposes NVMe i/f to the host
2. SmartNIC 1 implements NVMf Initiator
3. SmartNIC 2 implements NVMf Target to the NVMe drives on the host
   1. Possible HW offload to prevent passing through ARM cores on the target
4. Problems:
   1. Data is copied twice!
   2. Number of SQ+CQ pairs in both Host 1 & SmartNIC 1 is \#Drives*\#Cores
1. **Host NVMe → SmartNIC NVMf (Opt 2)**

   1. SmartNIC 1 accesses the host’s RAM, creates and maintains NVMf queues for it in the host’s RAM.
   2. SmartNIC 1 implements NVMf Initiator control and partial data paths.
   3. SmartNIC 2 implements NVMf Target to the NVMe drives on the host.
      1. Possible HW offload to prevent passing through ARM cores on the target.
   4. **Problems:**
      1. SmartNIC 1 needs to poll completions in the host’s RAM via PCIe hierarchy.
      2. Requires special device driver to allocate host memory for the SmartNIC.
Host NVMe → SmartNIC NVMf (Opt 3)

1. SmartNIC 1 accesses the host’s RAM, duplicates the memory key and uses the host’s memory key for data transfer.
2. SmartNIC 1 implements NVMf Initiator control and data paths – completions arrive to SmartNIC 1’s memory.
3. SmartNIC 2 implements NVMf Target to the NVMe drives on the host.
   1. Possible HW offload to prevent passing through ARM cores on the target.
4. Benefits:
   1. Data is not copied and is transferred directly between the hosts.
   2. Protocol details (NVMf or Excelero’s RDDA) are hidden within SmartNIC.
Q & A

Thank you!

www.excelero.com