New Infrastructure Architectures for Optimizing the Modern Data Center

Nigel Alvares
Vice President of SSD & Data Center Storage Solutions

Marvell Semiconductor
Growing Waves of Innovation & Productivity Growth

1st Wave
1983
Personal computing

2nd Wave
1995
Internet era

3rd Wave
2007
Mobile era

4th Wave
2011
Cloud era

5th Wave
2018
Data era

SW-Defined Composable Infrastructure
Data Era is **Driving** Need to Transform Infrastructure

Unprecedented demand growth & velocity

Low latency & data volume will decentralize the Cloud

Optimized architecture solutions needed

5G & Artificial Intelligence will accelerate disruption

“One Size Fits All” data centers will no longer work
“One Size Fits All” Data Center Will No Longer Work

Modern data centers migrating to composable infrastructure

Architecture solutions supporting open standards & innovation are critical

Compute  Networking  NVMe
Storage  Class  Memory  SATA & NVMe
SSD  SSD  Storage  SAS & SATA
HDD Storage  SSD  m.2  SSD  GPU
Accelerators  Application  Accelerators

Open. Together.
Introducing Composable Data Center Fabric Switch Family: Marvell Prestera® CX 8500 Family

Complete portfolio for 25G & 50G composable data center architectures
Integrates innovative SAFE & FASTER technology to optimize infrastructure solutions
Prestera® SAFE Enables Virtual Storage Orchestration

**Storage Aware Flow Engine (SAFE)**

- Ability to track flows real-time to enable network resource usage visibility
- Pin points congested disks & aggressor hosts
- Can limit bandwidth of a given aggressor or can move data from the congested disk to another location in the pool
Prestera® FASTER Reduces Power Consumption & Latency

Reduces network layers from 4 to 2 for up to 500K network nodes
- Reduces power
- Minimizes Cost
- Lowers Latency
- Increases performance

Reduces overall network costs by greater than 50%

High-radix switch core enables scale out of high performance modular nodes up to 1K ports

Legacy Networks

FASTER Enabled Network
Marvell Prestera® CX Reference Design with SONiC

- Scalable design from 12.8Tbps to 51.2Tbps
- Demonstrates Prestera® FASTER technology
- Supports SONiC 2019.3
- Free Range Routing (FRR) is default stack
- RDMA & Access Control List (ACL) enhancements

Come check it out in action at Marvell OCP booth
Network Connectivity for Composable Infrastructure: FastLinQ® NICs

Complete product family spanning 10, 25, 40, 50 & 100 GbE
Multiple protocols supported with offloads & accelerators
Optimized NVMe-oF offloads across RDMA over Ethernet (RoCE), TCP/IP & iWARP

OCP 2.0 & OCP 3.0 Form Factors Available
ARM Server Processors for Composable Infrastructure: ThunderX2®

>30% smaller die size
Up to 2X more efficient
Up to 50% lower cost
Most widely deployed & supported ARM-based server processor

2-4X better TCO than alternative x86 solutions
ThunderX2® Offers Composable Infrastructure Benefits

Better application throughput

- Maximum cores: ThunderX2 > Xeon Skylake

Lower application latency

- Memory bandwidth: ThunderX2 > Xeon Skylake

Flexible networking & storage

- Number of I/O lanes: ThunderX2 > Xeon Skylake

Compelling performance/$

- Full SoC perf/$: ThunderX2 > Xeon Skylake

Source: Intel & Marvell data sheets
OCP Project Olympus with ThunderX2®

Dual socket Arm®-based server processor of choice

Enabling multiple composable data center infrastructure architecture solutions
Today’s Composable SSD Storage (JBOF) Challenge

Today’s Composable SSD Storage (JBOF)

Latency
Power
Oversubscribing $$Ds
Significant Bottleneck
How Significant is Bottleneck?

Today’s Composable SSD Storage (JBOF)

24 SSDs = PCIe Gen3x4 32Gbps → PCIe Gen4x4 64Gbps → PCIe Gen5x4 128Gbps

768Gbps → 1.5Tbps → 3.0Tbps

Huge bandwidth pipe bottleneck combined with CPU+DRAM+NIC+PCIe switch power, cost & latency increase TCO

New high-performance, low-latency scalable architecture needed
New Composable SSD Storage Solution: Marvell End to End NVMe-oF Ethernet Bunch of Flash (EBOF)

End-to-End NVMe-oF EBOF: simple, scalable linear native performance!
Optimizes $ per IOPS & IOPs per GB
24 NVMe-oF SSDs = upto 16M IOPs

TOR Ethernet
Comparing Today’s Composable JBOF vs EBOF

Today’s Composable SSD Storage (JBOF)

- SSD SSD SSD SSD SSD SSD SSD SSD

Limited performance, high CPU power & high BOM

End-to-End NVMe-oF Ethernet Bunch of Flash (EBOF)

- NVMe SSD NVMe SSD NVMe SSD NVMe SSD NVMe SSD NVMe SSD

Simple native scalable performance with extremely lower power consumption

>65%* TCO Savings excluding SSDs

*Boshiba & Marvell TCO analysis

High TCO

Low TCO

Bottleneck
Comparing OCP Olympus FX-16 PCIe-JBOF vs EBOF

OCP FX-16 JBOF (PCIe fabric)

<table>
<thead>
<tr>
<th>ToR (Switch)</th>
<th>1</th>
<th>Utility Node (Server)</th>
<th>Storage Node (2S 500W)</th>
<th>JBOF (50W)</th>
<th>Storage Node (2S 500W)</th>
<th>JBOF (50W)</th>
<th>Storage Node (2S 500W)</th>
<th>JBOF (50W)</th>
<th>16</th>
<th>Storage Node (2S 500W)</th>
<th>JBOF (50W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Future EBOF (Ethernet fabric)

| ToR (Switch) | 1 | Utility Node (Server) | Storage Node (2S 500W) | JBOF (50W) | 1 | Storage Node (2S 500W) | JBOF (50W) | 8 | Storage Node (2S 500W) | JBOF (50W) | 24 | EBOF (100W) |
|--------------|---|------------------------|------------------------|------------|----|------------------------|------------|----|------------------------|------------|    |                |
|              |   |                        |                        |            | 1  | EBOF (100W)            |            |    |                        |            |    |                |

- **20% Lower Rack Cost**
- **30% Less Power Consumption**
- **50% More Rack Capacity**

Eliminates 1:1 Storage Node to JBOF Ratio
Enabling Higher Storage Efficiencies, Utilizations & Capabilities
New Composable HDD Storage Solution: NVMe-oF Ethernet Bunch of Disks (EBOD)

- Enables common management & network fabric for HDDs, SSDs & SCM
- Supports SAS, SATA & NVMe HDDs with offload capabilities
- Eliminates need for CPUs, SmartNICs, SAS controllers & expanders/switches
- Minimizes TCO with best-in-class low-power optimized chipset
Marvell Composable Data Center Infrastructure Architectures in Action at our OCP Booth

**Initiators:**
3x 2S ThunderX2® Compute shelf with 200Gb FastlinQ® RNIC
Total: 600Gb/s

**EBOF:**
2U 24 SSD shelf with Marvell Prestera® & NVMe-oF Storage Controllers
Input: 600Gb, Output: 24x 25Gb SSDs
Total: 1.2Tb/s

**EBOD Target:**
2U 16 HDD shelf with Marvell Storage Processor & NVMe-oF Storage Controllers
Input: 25Gb, Output: 16x 12Gb HDDs
Total: 217Gb/s
Summary & Call to Action

- Data era is upon us & morphing data center architectures
- “One Size Fits All” data centers no longer work
- Composable data center infrastructure solutions required
- Marvell offering multiple open chipset solutions to address
- Call for action:
  - Develop OCP reference designs for EBOFs & EBODs
  - Develop OCP specifications for composable data center fabric
  - Help standardize Ethernet SSD connector
  - Attend tomorrow’s “Ethernet Is The New Fabric-of-Choice for Storage Expansion” session
  - Visit Marvell booth to see composable data center solutions in action!