SONiC at 200/400GbE
- Ready For Hyperscale

David Iles – Sr. Director Ethernet Switching

March 2019
Mellanox Leadership Across Industries

- 5 of Top 6 Global Banks Use Mellanox
- 10 of Top 10 Automotive Manufacturers Use Mellanox
- 3 of Top 5 Pharmaceutical Companies Use Mellanox
- 9 of Top 10 Oil and Gas Companies Use Mellanox
- 9 of Top 10 Hyperscale Companies Use Mellanox

Mellanox Interconnect Solutions Deliver Highest Data Center Return on Investment
25G and faster speeds growing to 63% of the market
Data Center Optics Market

100G:
- The hottest market today

200G & 400G:
- Will surpass 100G in 2022

Data from LightCounting, Sep. 2018
Cloud Without Compromise: Spectrum-2

Best in Class Buffer Architecture
- Fair & Predictable Bandwidth sharing
- 42MB Packet buffers - available to any port
- 10x Higher Microburst Absorption

Best in Class Virtualization
- VXLAN Routing without loopbacks
- 12K VTEPs for EVPN
- 500K+ VLXAN Tunnels

Best in Class Telemetry
- What Just Happened
- Streaming & Inband Network Telemetry
- Histograms – better than watermarks
- Deep Packet Inspection (512B)
High Performance with Fully Shared Buffers

**Fully Shared Packet Buffers**

- **Bursty traffic**
- **33Gbps**
- **33Gbps**
- **33Gbps**

All 42MB packet buffers available to absorb **Microburst**

**Superior Microburst Absorption**

- **4-6X larger effective buffer size!**

**Competitor’s Split-buffers**

- **50Gbps**
- **24Gbps**
- **25Gbps**

- **Bursty traffic**

Only 17% of buffers available to absorb **Microburst**

© 2019 Mellanox Technologies
Can You Afford an Unfair/Unpredictable Cloud?

Best in Class Buffer Architecture

Pepsi
Coca Cola
Snapple
Red Bull
Monster
Starbucks
Can You Afford to Lose a Rack Full of Servers?
ISSU for SONiC

- Problems with Commercial ISSU
  - NOS downgrades
  - SDK change
  - Reflash firmware/CPLD/FPGA
  - Only reliable for patch fixes

- A Better ISSU for SONiC:
  - Fast Boot:
    - Requirement: Data plane disruption under 30 seconds
    - Status: Supported today
  - Warm Boot:
    - Requirement: Data plane disruption under 1 second
    - Status: ~1 sec downtime with upgrade of SAI/SDK - done
      demonstrated warm boot in 70 msec
  - No Hit Boot - WIP
Protocols vs Telemetry

Legacy Mindset

- PIM
- HSRP
- LACP
- VPC
- OSPFv2
- RIPv2
- EIGRP
- SNMP
- TACACS
- UFD
- PVRST/ MSTP
- Private VLAN
- Loop/Root/BPDU Guard
- QOS
- VRRP
- VTP
- GVRP
- IGMP
- SNMP
- RMON
- SPAN
- CDP
- Packet Brokering
- LLDP
- TRILL
- SPB
- FabricPath
- VCS
- Qfabric
- BGP
- FCoE
- BFD
- FEX
- OVSDB/VTEP
- MLAG
- QinQ
- SNMP
- SPAN
- ERSPAN
- sFlow
- IPFIX
- SYSLOG
- Packet Brokering
- LLDP
- Watermarks
- Mirror Drops
- Buffer Histograms
- Mirror Congestion
- In Band Telemetry
- Streaming Telemetry (GPB)
- Real-time Visibility Snapshots

Webscale Mindset

- EVPN
- LACP
- BGP/BFD
- WJH
- LLDP
- SPAN
- sFlow
- ERSPAN
- SYSLOG
- In-situ OAM
- Mirror Drops
- RoCE Telemetry
- Packet Brokering
- Buffer Histograms
- Mirror Congestion
- In Band Telemetry
- Streaming Telemetry (GPB)
- Real-time Visibility Snapshots
Why Do We Need Telemetry?

- Improve Time to Innocence
- Understand our Network
- Get more out of the Network
SONiC – Critical Resource Monitor

- Monitors utilization of ASIC resources by polling SAI attributes
  - Syslog message if utilization exceeds thresholds

- Resources monitored by CRM:
  - IPv4 and IPv6 routes
  - IPv4 and IPv6 nexthops
  - IPv4 and IPv6 neighbors
  - Nexthop groups and group members
  - ACL groups, tables, entries, counters
  - FDB entries

- Low and High thresholds per each resource

- Percentage used and free threshold types
Event Driven Telemetry

- CRM logs when any resources exceed LOW or HIGH threshold

- SYSLOG Message Format:
  - <Date/Time> WARNING <resource>: THRESHOLD_EXCEEDED for <type> <%> Used count <value> free count <value>
  - <Date/Time> NOTICE <resource>: THRESHOLD_CLEAR for <type> <%> Used count <value> free count <value>
SONiC – Server Load Balancing Demo

- **Client**
  - Generates 100K connections per second
  - Average connection live time 10 sec

- **Server Load Balancing box**
  - Load balancer
  - Single VIP
  - 1K DIP

- **Controller**
  - Create DIP change in average every 10 sec

- **Server**
  - Receive and monitor connection

---

100K Connections per second
Key Take Away’s

- SONiC is gaining momentum
- SONiC is gaining functionality
- Go try SONiC – It’s free
Thank You