Introduction of Meta network hardware

Wedge 400: 16x400G + 32x200G ToR Switch
Introduction of Wedge 400

Lingjun Wu, Hardware Engineer, Meta
Vignesh Vijayanath, Technical Program Manager, Meta
Johnson Liu, Engineering Director, Celestica
Motivation for Wedge 400

• Increased adoption of 100G NICs on Compute and flash storage platforms
• Need for higher front panel port density
• Enabling AI/ML platforms with higher NIC speeds
• Enabling future uplink bandwidth expansion
Key Improvements over Wedge 100s

- 2RU 48 Port Switch
  - 32 QSFP56 ports support up to 200G per port
  - 16 QSFP-DD ports support up to 400G per port
- 4X the switching capacity (3.2T → 12.8T)
- 8X burst absorption performance
- RS-FEC support for lower speeds
- Field replaceable CPU sub-system
## ToR Evolution

<table>
<thead>
<tr>
<th>Switch Platform</th>
<th>Wedge 40</th>
<th>Wedge 100</th>
<th>Wedge 400</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Year Introduced</strong></td>
<td>2015</td>
<td>2017</td>
<td>2020</td>
</tr>
<tr>
<td><strong>ASIC</strong></td>
<td>Trident 2 (BRCM)</td>
<td>Tomahawk (BRCM)</td>
<td>Tomahawk 3 (BRCM)</td>
</tr>
<tr>
<td><strong>Switching Capacity</strong></td>
<td>1.28 Tbps</td>
<td>3.2 Tbps</td>
<td>12.8 Tbps</td>
</tr>
<tr>
<td><strong>Supported Speeds</strong></td>
<td>10GE/40GE</td>
<td>10/25/40/50/100 GE</td>
<td>10/25/40/50/100/200/400 GE</td>
</tr>
<tr>
<td><strong>Physical Port Config</strong></td>
<td>16x 40G QSFP</td>
<td>32 x 100G QSFP28</td>
<td>16 x 400G QSFP-DD + 32 x 200G QSFP56</td>
</tr>
</tbody>
</table>
Product Overview

AC PSU
DC PEM
Switch Control Module (FRU)
Fan Module (FRU)
BMC Storage Module (FRU)

• 21 inch ORv2 Tray Adaptor
Wedge 400 System Introduction

- Wedge400 HW Description
- Wedge 400 Top/Side View
- Wedge400 Front/Rear View
- SWE (Switch Element)
- Out-of-Band Ethernet
- PCIe Assignment
- UART and Rackmon
- SPI Diagram
- BSM (BMC Storage Module)
- Power Evaluation
Wedge 400: HW Description

- 19-inch rack, 2RU
- ORv2 adaptor tray
- Uplinks: 16xQSFP-DD (400/200/100G)
- Downlinks: 32xQSFP56(200/100/50/25/10G)
- Switch Software: FBOSS
- Single stage / Single ASIC (TH3)
- Minilake COMe Module (OCP)
- BMC (AST2520)
**Wedge 400 Top/Side View**

- FRU: SCM, FAN modules, PSU/PEM.
Wedge 400 Front/Rear View

QSFP-DD

QSFP56

PEM (12V)

FANs

SCM

OPEN POSSIBILITIES.
Switch Element (SWE)

- Switch element consists of three components
- TH3 switch ASIC
- Minilake COMe module
- BMC
- Single switch element
Out-of-Band Ethernet

- BCM5389 as OOB switch
- Front port RJ45
- uS (Minilake COMe)
- BMC

Reserved for RackMon

10Gbase-KR

TH3

Configure SPI

BCM5389

SGMII 3

SGMII 6

SGMII 7

SGMII 0

SGMII 5

SGMII 1

SGMII 2

MC0

MC1

NETWORKING

MUX
PCIe Assignment

- PCIe Gen3
  - x4 lanes to TH3 (switch ASIC)
  - x4 lanes to NVMe SSD

- PCIe Gen1
  - x2 lanes to DOM FPGA#1
  - x2 lanes to DOM FPGA#2
COMEmon

- BMC access
- Sol.sh to uS
- OCP debug card access
- Rackmon
- Manage PSU/BBU inside rack
- USB access through BMC or uS

Network:

- COME board
- LPC
- UART
- UART 1
- UART 2
- UART 5
- BMC
- Mux
- System CPLD
- Rackmon
- Rack RJ45 x4
- BMC
- USB 2.0
- USB Hub
- MaxUSB x3
- MaxUSB x4
SPI Diagram

- BMC Upgrading FWs
- TH3 FW flash
- FPGA flashes
- Secondary BIOS
- OOB switch config EEPROM
BSM (BMC Storage Module)

- Meta defined BMC storage module
- Primary & secondary flashes
- eMMC and EEPROM

M.2 Type A key 2260 H=8.5mm
<table>
<thead>
<tr>
<th>Rack Type</th>
<th>Downlink NIC Speed</th>
<th>NIC Qty.</th>
<th>Uplinks</th>
<th>W400 Power (Typical)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref #1</td>
<td>50G</td>
<td>16</td>
<td>8*100G CWDM4</td>
<td>240W</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4*100G CWDM4</td>
<td>222W</td>
</tr>
<tr>
<td>Ref #2</td>
<td>100G</td>
<td>4</td>
<td>4*100G CWDM4</td>
<td>212W</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>8*100G CWDM4</td>
<td>242W</td>
</tr>
<tr>
<td>Ref #3</td>
<td>100G</td>
<td>32</td>
<td>8*100G CWDM4</td>
<td>285W</td>
</tr>
</tbody>
</table>
Use cases inside Meta

• Deployed in production since July 2020
• Onboarded onto all new rack platforms (Compute, Storage, AI/ML)

Optics Supported
• 100G (Native, 40G & 10G mode)
• 200G (Native, 100G mod)
• 400G

Physical Media Supported
• DAC Q-Q (Multi-host & Single host)
• DAC Q-2Q
• DAC Q-4S
Celestica’s Role in OCP

- Celestica is a leading hardware platform solutions provider to both service provider and enterprise markets, with a focus on leading technologies such as 400G
- We provide both standard solutions or customized solutions by collaboratively engaging with customers like Meta to develop, deliver, and support their platforms such as **Wedge400**
Collaboration Model

HW Development
- Celestica aligns with Meta on requirements and finalizes system level architecture.
- Celestica responsible for HW design, and complete prototype bring-up and the following functional/reliability tests.
- Meta will review Celestica engineering design and test reports.

SW Development
- Joint design and source sharing
- From third-party or community, like ASIC SDK/BIOS from vendor, and standard open source Linux distribution.
- Celestica responsible for diagnostic software development.
Wedge 400 Test Strategy

- **Comprehensive test coverage:**
  board level beyond 90% (ICT plus Boundary test) and 100% test coverage for function

- **Closed loop and continuous improvement:**
  test case mapping to product requirement, RCCA process

- **Management system:**
  issue management, issue lifecycle, DI criteria, test progress transparency

- **Automation:**
  efficiency, quality, and software release lead time improvement, real time closed loop, large scale rack orchestration test

- **High test standard:**
  stress test (margin, cycle, long duration), sample size (>35), CICT with dedicated test setup

- **Manufacturing test:**
  criteria for each test phase, design and MFG co-design MFG test process, customized diagnostics
Call to Action

• Timeline for Contribution Availability
  • Wedge400 specs and design package for OCP Accepted in Nov. 2021
• Timeline for Product Availability
  • To be provided by Celestica
• Where to find additional information
  • For more information, please visit: https://www.celestica.com/our-expertise/hardware-platform-solutions/overview
  • For further inquiries, please contact: ccsonovation-cls@celestica.com
Thank you!
Open Discussion