



OPEN
Compute Project



OCP U.S. SUMMIT 2017

Santa Clara, CA



Backpack: Facebook's 100G Modular Switch

Zhiping Yao

Network Hardware Engineer, Facebook

OPEN HARDWARE.

OPEN SOFTWARE.

OPEN FUTURE.



Backpack System Introduction



Backpack

From Wedge 100 to Backpack

→ Open

- Hardware: fully contributed to OCP
- Software: FBOSS, OpenBMC

→ Disaggregated

- Software and hardware disaggregation
- Based on a simple building block: switch element, and replicated 12 times
- Physically separated System Control Module from Switch ASIC

→ A bunch of servers

- Managed by BMC
- Deployed and provisioned like servers
- Much faster software push to support new features

Backpack : An Open Modular Switch

- Hardware architecture: ethernet only, fully open
- Network topology: dual stage spine-leaf
- Switch software: FBOSS and OpenBMC
- Manageability: operated like server from BMC
- Density: 128 x QSFP28 100G support

Switch Element (SWE)

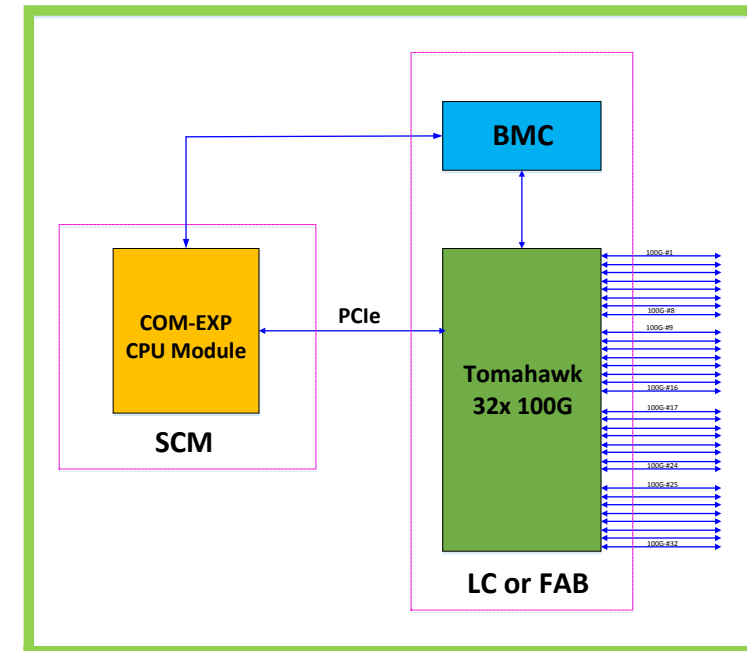
→ Switch element consists of three components

- Tomahawk switch ASIC
- COM-e CPU module
- BMC

→ Logically wedge100 is one SWE

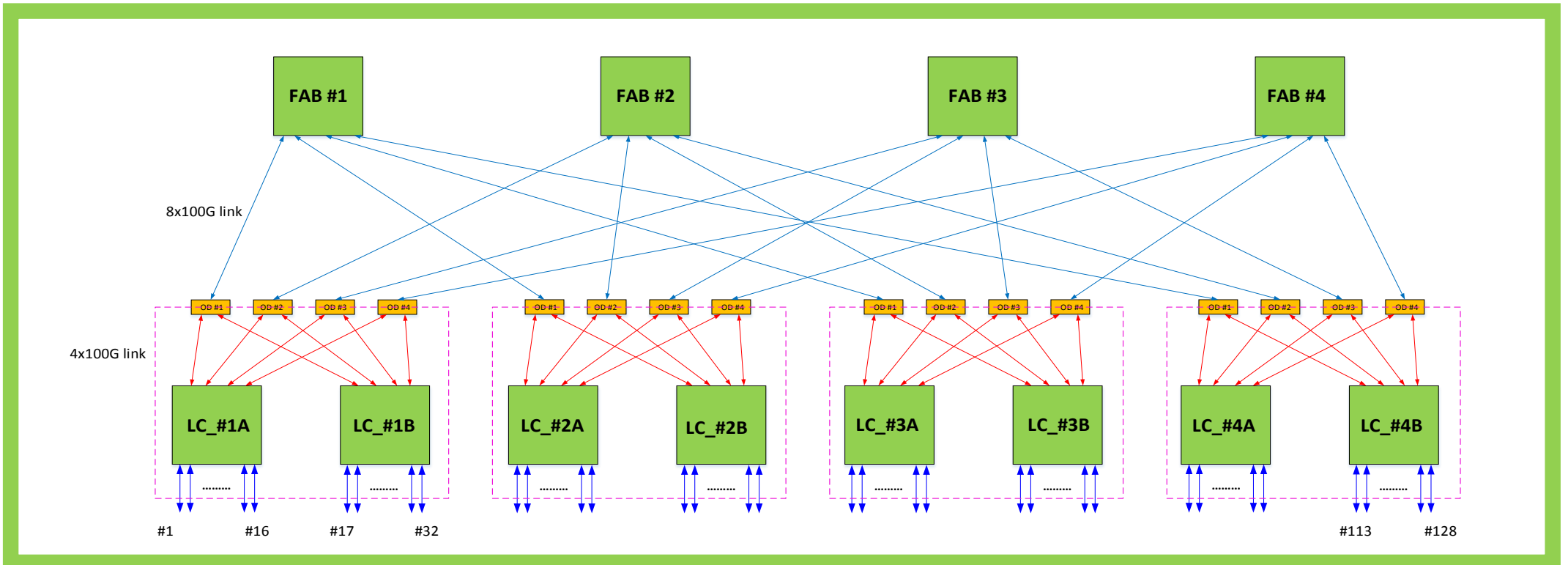
→ 12 switch elements in Backpack

- Each LC has 2 SWE
- Each FAB has 1 SWE



Fabric Topology: CLOS

- Two Stage Spine-leaf architecture
- Fully non-blocking
- 12 Switch Elements

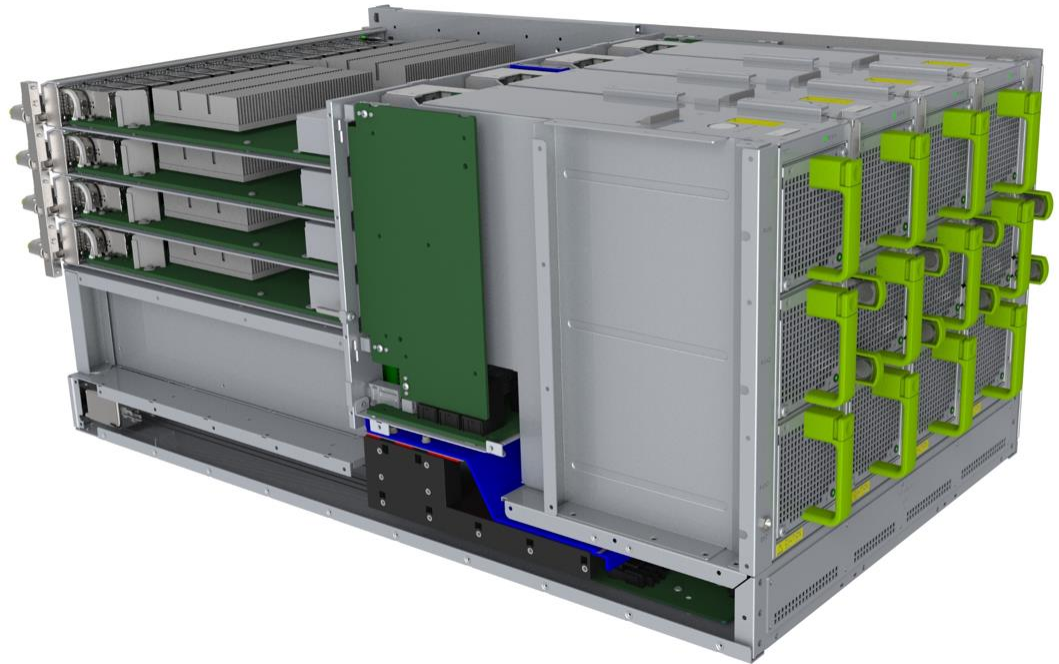
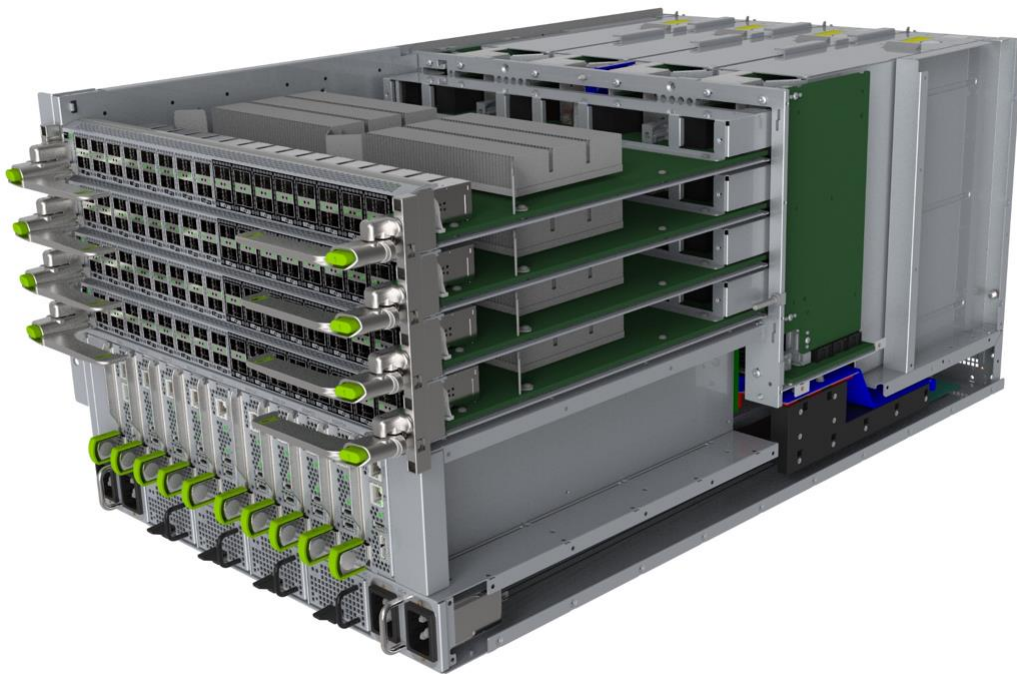


Innovative Design

- Fully disaggregated architecture
 - 12 switch elements (SWE)
 - Separate SCM module for control CPU
- An Orthogonal Direct Chassis Architecture
 - All major module cards are designed to be mated orthogonally
 - Open up more air channel for a better thermal performance
 - Reduce the PCB trace length for better signal integrity
 - Support future 8x16 OD connector to double port and speed
- A sophisticated thermal design to support low cost 55C CWDM4 optics

Backpack : FB Modular Switch Platform

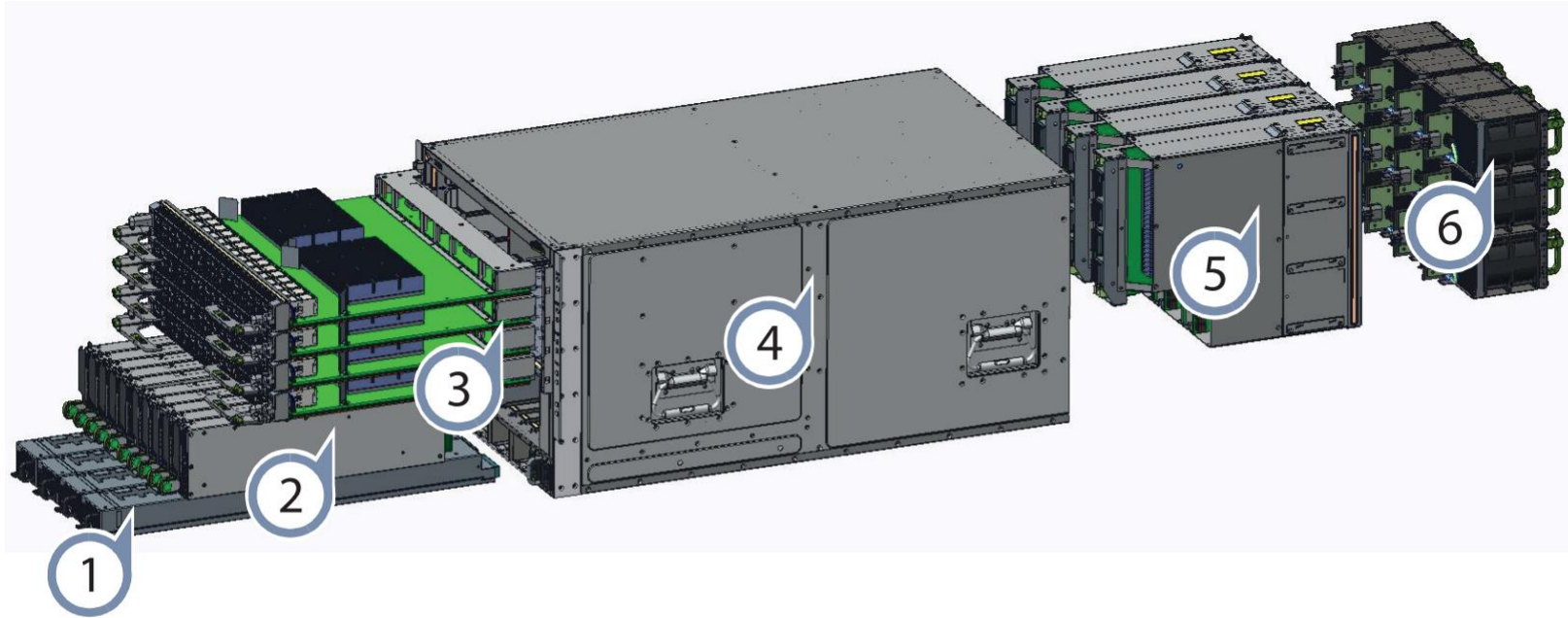
- Orthogonal Direct Architecture
- Disaggregated data, control, and management plane design



Backpack System Components

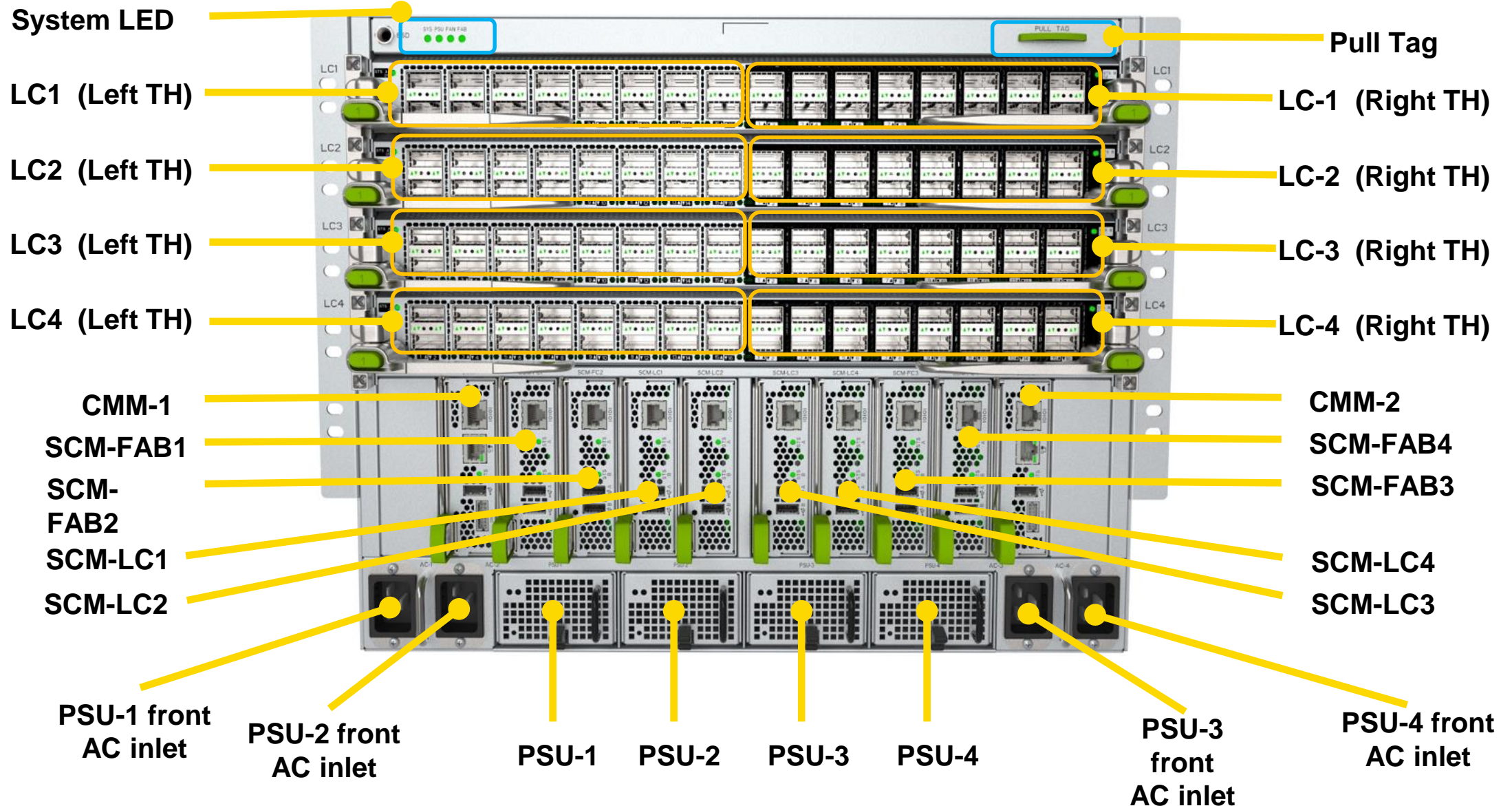
- Line Cards (LC)
- Fabric Card (FAB)
- System Controller Module (SCM)
- Chassis Management Module (CMM)
- Horizontal Control Plane (HCP)
- Vertical Control Plane (VCP-L, VCP-R)
- Bus Bar Assembly (BBA)
 - Horizontal Bus Bar (HBAR)
 - Vertical Bus Bar (VBAR)
 - Horizontal Power Distribution Board (HPD)
- Fan Control Board (FCB)
- Power Distribution Board (PDB)

Backpack : Chassis Assembly

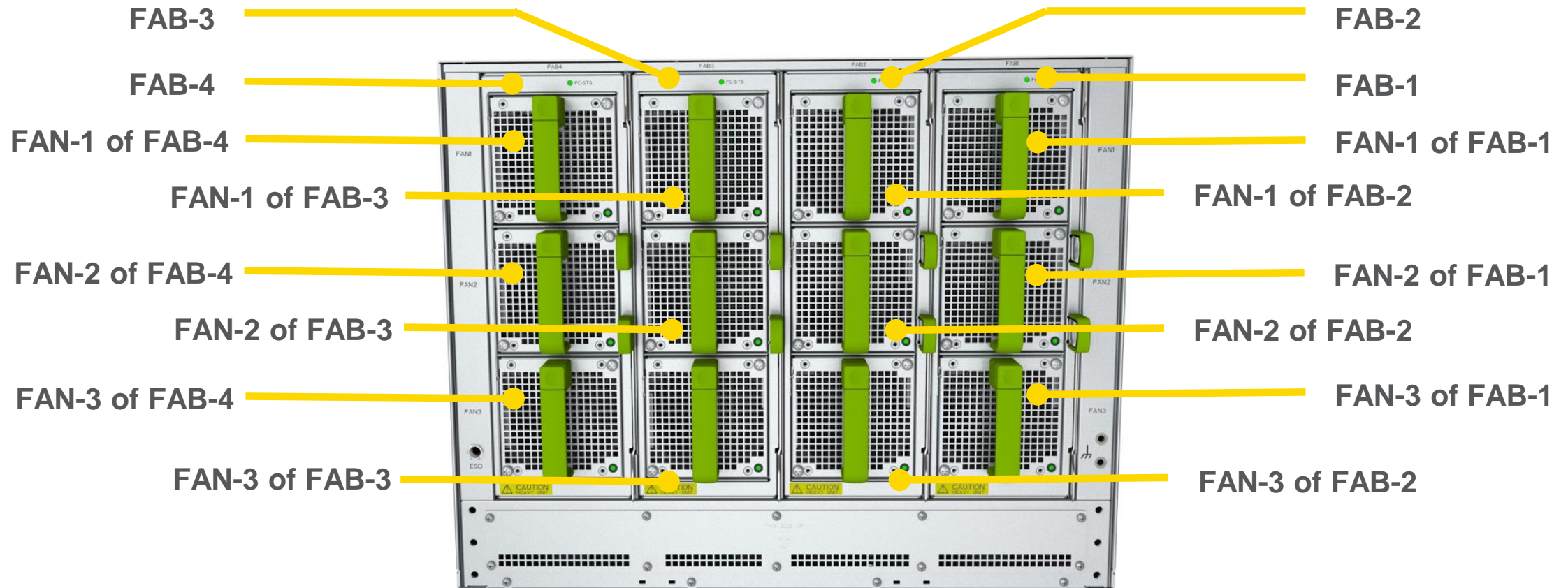


- | | | |
|---------------|-------------------|---------------|
| ① PSU modules | ② SCM&CMM modules | ③ LC modules |
| ④ Chassis | ⑤ FAB modules | ⑥ FAN modules |

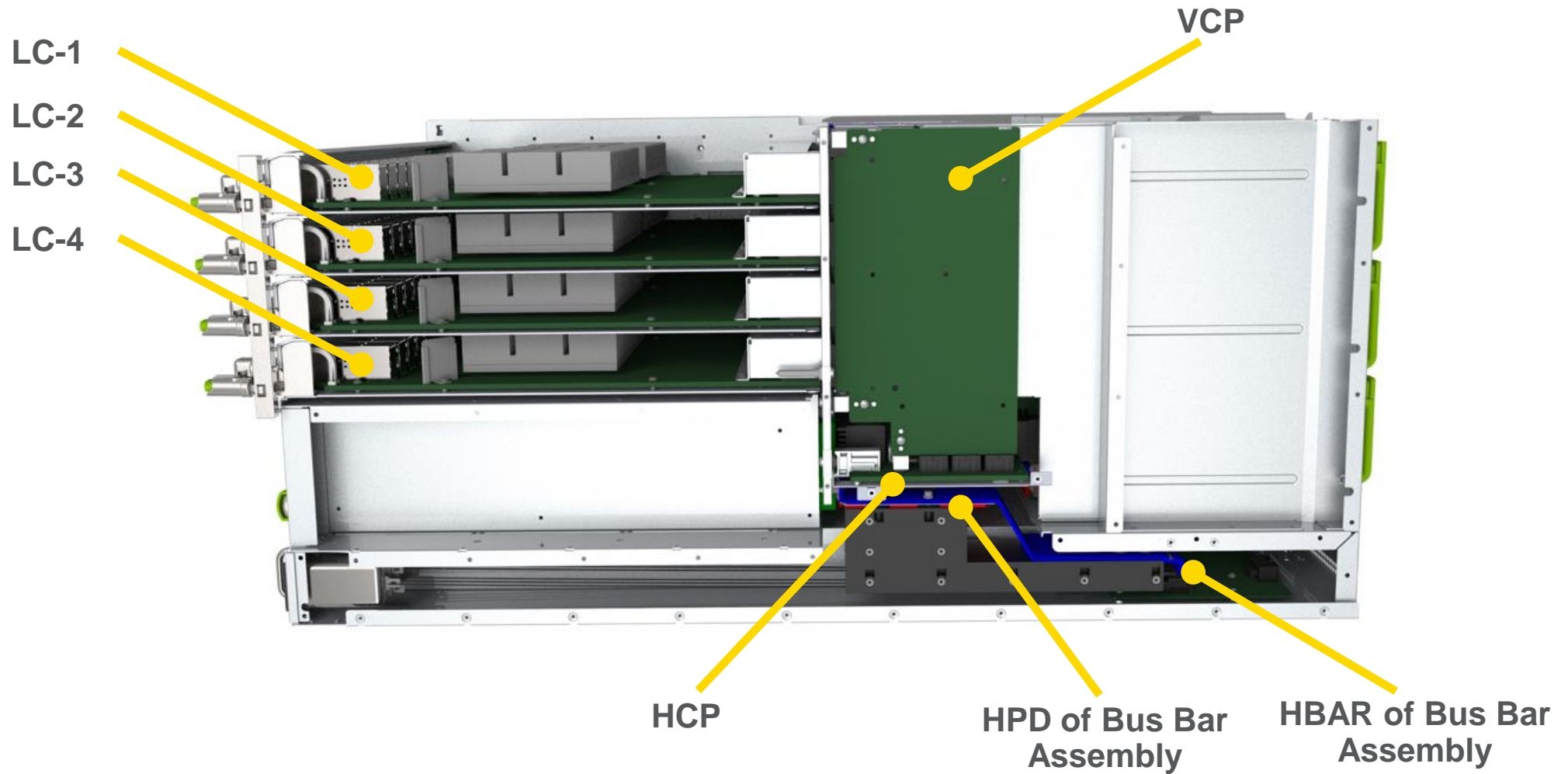
Backpack Front View



Backpack Rear View



Backpack Side View

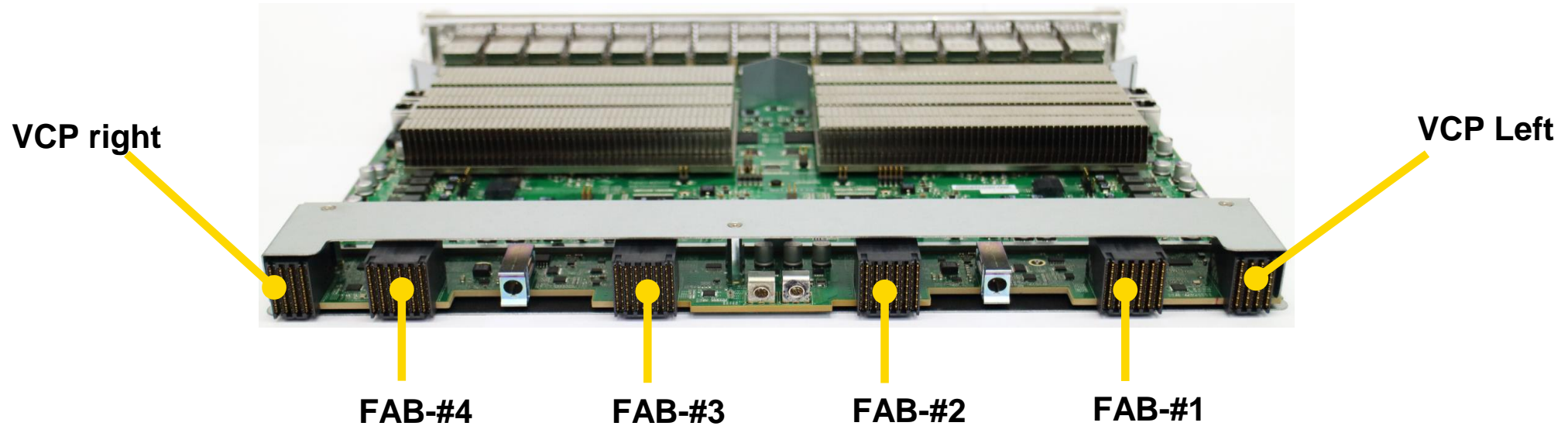
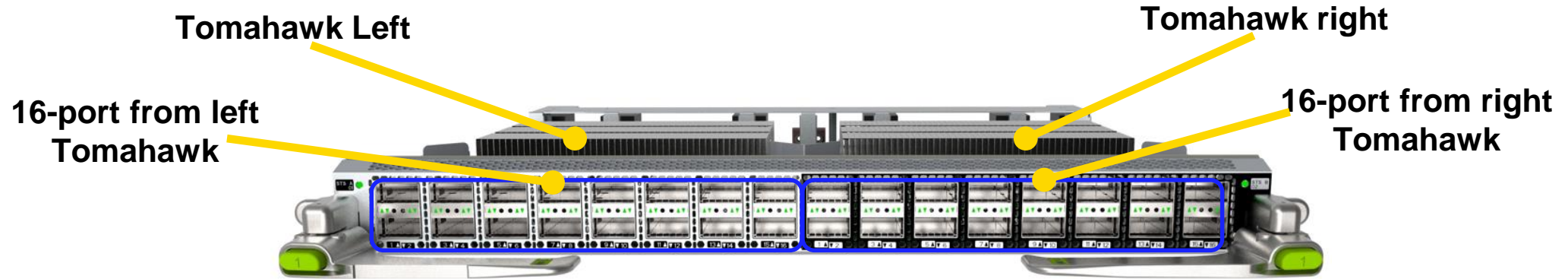


Line Card (LC)

- Two tomahawk 3.2T ASIC
- 32 QSFP28 100G Ports
- Four DMO Connector to FAB for data plane signals
- Two DMO Connector to VCP for control plane signals



Interface of Line Card (LC)



Fabric Card (FAB)

- One Tomahawk 3.2T ASIC
- Fan Control Board (FCB) is mezzanine card of FAB
- Four 6 x 12 DMO Connector to Four LC for data plane signals
- One 6x8 DMO Connector to HCP for control plane signals



Chassis Management Module (CMM)

- AS2540 BMC
- OOB 16 port Switch BCM5396
- Console UART MUX for all LC and FAB COM-e CPU and BMC CPU
- Chassis Management I2C bus



System Control Module (SCM)

- Two COM-e CPU Module Cards
- COM-e CPU Module has one to one mapping to SWE
- 6x12 DMO connector to HCP
- SCM-LC has two COM-e mounted, SCM-FAB has only one COM-e mounted



System Control Module (SCM)

→ SCM-LC:
SCM for Line Card



→ SCM-FAB:
SCM for Fabric Card



Optic Transceiver

→ Backpack support QSFP28 100G optic

- CWDM4-Lite
- Support other 100G MSA, such as SR4, LR4, CLR4, etc

→ Backpack support QSFP+ 40G optic

- QSFP+ 40G SR4 optic (multi-mode fiber OM3/OM4)
- QSFP+ 40G LR4 optic (single mode fiber)



Backpack Power Consumption

- Measured with 3.5W eload, unit is Watt
- Real power consumption of CWDM4 optic is between 2.6W to 3.1W
- System is running at <40% under normal DC scenario

PWM	20%	30%	40%	50%	60%	70%	80%	90%	100%
Normal condition	2670	2697	2726	2800	2896	3124	3290	3550	3729
One fan failure condition	2550	2627	2703	2768	2822	3056	3223	3428	3645

Thermal design

- Support 55C optic at ambient 35C environment
- 12 CR fan-tray on the rear panel
- Thermal Improved LC front panel design for better QSFP28 cooling
- Multiple on-board temperature sensor to monitor thermal healthy status of the system



Efficiency

- Excellent thermal design enable backpack fan running at close to 30% to achieve much better power efficiency
- Fully disaggregated architecture enable much easier software development and integration: simply replicate 12 times and scale
- The concept to design modular switch like a bunch of servers makes it much faster to integrate with our data center tooling environment.
Enabling Backpack to be seamlessly integrated into our infrastructure

Serviceability

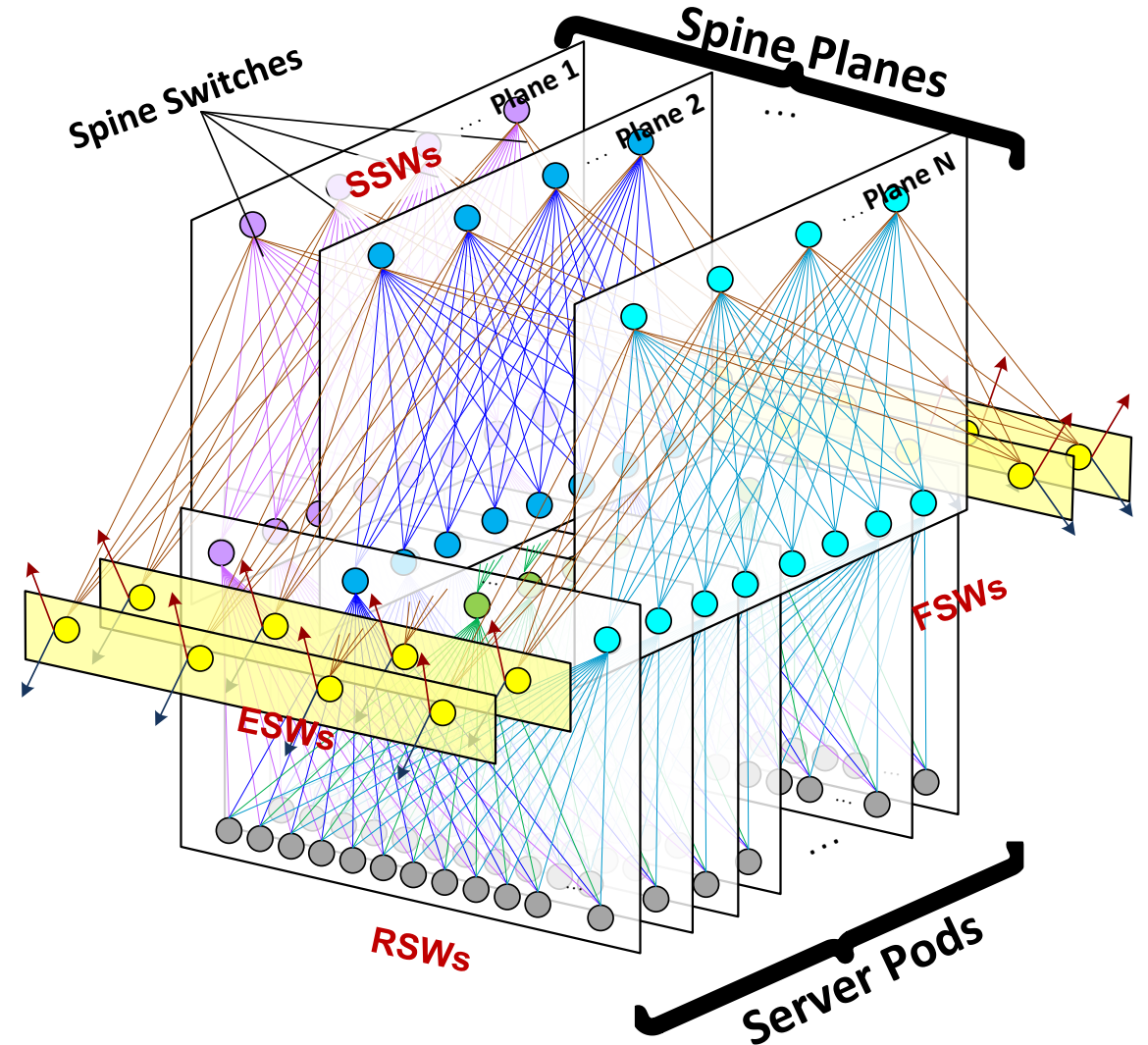
- Hot swappable CPU module(SCM) makes the maintenance work of DC much easier and smooth
- Screw-less chassis design
- Fully installed chassis directly from factory, plug and play
- Every switch element is managed by BMC
- Seamlessly integrated to FB DC tooling
 - Automatic provisioning
 - Fully monitor and managed by FB DC tool

Backpack at Scale

- Automatic provisioning
- 12 SWE Disaggregated design makes it much easier for FBOSS to scale to much larger scale.
- Chassis mechanic is hard tooled and mass production ready
- Manufacture is ready to ramp up

Facebook Data center Network

- 3-tie fabric
 - Spine Switch Layer: SSW
 - Fabric Switch Layer: FSW
 - Rack Switch Layer: RSW
- Backpack can be deployed in FSW, ESW and SSW!
- Ready to scale!



IN_USE!

- Dozens of Backpack switch have been carrying live traffic in FB data center for few months already
- Used in both our 40G and 100G data center fabric
- Ready to scale!



Open Community of Backpack

- ➔ OCP submission on 11/08/2016
- ➔ Our ODM partner: Celestica
- ➔ CWDM4-Lite: Finisar, ColorChip
- ➔ FB Open software: FBOSS, OpenBMC
- ➔ 3rd Party Switch Software:
 - Cumulus
 - Snaproute
 - Broadcom

OCP Contribution



Backpack OCP contribution

- Backpack OCP specification
- Backpack design package



Our ODM partner: Celestica

- Backpack 3rd Party Orderable PN: **D8020**



CWDM4-Lite

- ColorChip
- Finisar



OCP submission ftp

- <http://www.opencompute.org/wiki/Networking/SpecsAndDesigns>

Backpack OCP Document and Design Package



Electrical

- /Electrical/released schematic
- /Electrical/released layout
- /Electrical/BOM
- /Electrical/Released image



Mechanic

- /Mechanic/2D
- /Mechanic/BOM
- /Mechanic/3D (Plan to release after MP)



Specification

- /Backpack_OCP_Spec 1.0.pdf
- /Backpack_Deep_Dive_OCP_v03.pdf

Key Milestone of Backpack

- ➔ Initial announcement and submission to OCP: 11/08/2016
- ➔ The first backpack to carry FB live traffic: 09/30/2016
- ➔ Mass Production at Celestica: Q1-2017

Q&A





OPEN

Compute Project

