Event Builder Upgrade

S.Y.Suzuki (KEK, CRC)

Assumption

- Total throughput on EB ↑
- Throughput per ROPC ↑
- **# of ROPC ↓**
- **Throughput per HLT unit →**
- # of HLT unit may ↑

Software

- If the throughput per HLT unit will be kept, not problematic

- Currently multi-in & single-out parts consume CPU power
  - eb0, eb1rx, eb2rx
  - Reducing # of input nice
  - Is eb0 necessary in upgraded readout? (probably NO)

Hardware

- It is just choice of network switch for EB
- needs 10G-T or SFP+?
- Deep buffer or not?
- Link speed between EH and Server Room

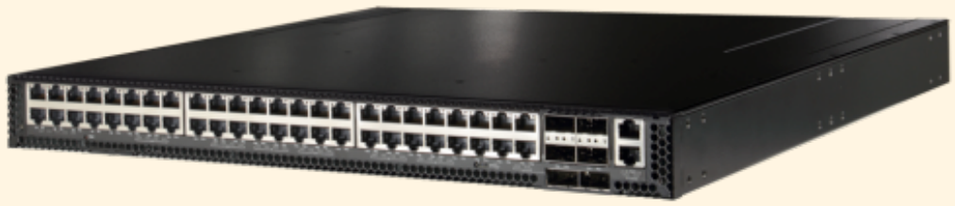
Prosptect

- needs 10G-T or SFP+? → **NOT SURE**
- Deep buffer or not? → **depends on budget, but will be no.**
- Link speed between EH and Server Room → **40G**

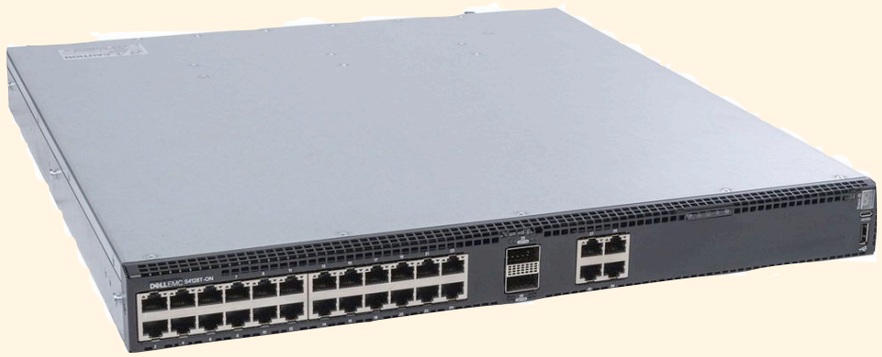
10G-T or SFP+ ?

- Currently SFP+ module for 10G-T is not yet popular because of the power consumption.
- Support is very limited.
- If we purchase 10G-T model switch, number of SFP+ is very limited.

Switch for 10G-T option

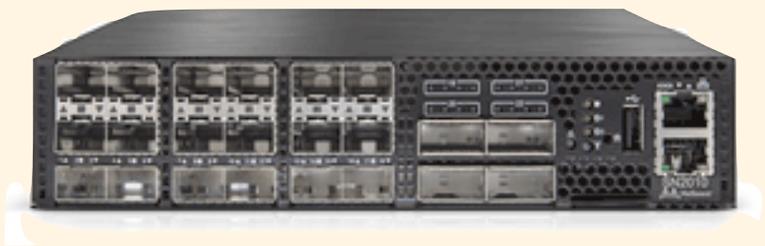


Mellanox AS5812
(48x10G-T + 6xQSFP+)



DELL S4128T
(48x10G-T + 2xQSFP28)

Switch for SFP+ option



Mellanox SN2010
(18xSFP28 + 4xQSFP28)

10G-T vs SFP+

10G-T option

- PCIe slots of ROPC = 1
  - 4port NIC (like Intel X710-T4) for data, SLC, NSM
- cost of EH switch ↑↑ (except DELL)
- cost of cable (cat7) ↓↓

SFP+ option

- PCIe slots ROPC = 2 (1 for data, 1 for SLC+NSM)
- cost of EH swtich ↓
- cost of cable (AOC, 25k for 30m) ↑

Hard to decide, so NOT SURE

How about direct connection?

- Directly connect upgraded ROPCs to existing 10G switch in the server room.
- It will require 10G-LR
  - 10G-NIC for upgraded ROPC
  - Optics for the server room side (120k per port)
  - Optics for the ROPC side (50k per port)

10G-T / SFP+ option

10G

ROPC

ROPC

ROPC

ROPC

New switch in EH

40G-LR

Existing switch in server room

10G

HLT

HLT

HLT

HLT

EH | Server room for HLT

Direct connection

10G

ROPC

ROPC

ROPC

ROPC

10G (LR)

Existing switch in server room

HLT

HLT

HLT

HLT

Three candidates

1. **10G-T → 3.5M + 80k x # of ROPC**
    - ROPC NIC (80k yen)
    - Short buffer switch in EH (2M)
    - 40G QSFP+ (for EH 500k, for Server Room will be 1M ~ 1.5M)
2. **SFP+ →3.5M + 65k x # of ROPC**
    - ROPC NIC 40k (without 1000T)
    - AOC cable 25k
    - Short buffer switch in EH (2M)
    - 40G QSFP+ (for EH 500k, for Server Room will be 1M ~ 1.5M)
3. **Direct connection → 220k ~ 250k x # of ROPC**
    - ROPC NIC without 1000T (40k)
    - 10G-SR for ROPC (50k)
    - 10G-LR for Server room (120k)

When # of ROPC is 10 ~ 20

- 10G-T: 4.3M ~ 5.3M
- SFP+: 4.2M ~ 4.8M
  (and additional 1000T NIC will be needed)
- direct: 2.5M ~ 5M
  (and additional 1000T NIC will be needed)

About flow control

- There is no deep packet buffer in 10G-T and SFP+ options
  - Without flow control, switch may discard packets silently.
  - Relying TCP retransmission causes performance degradation.
- **We have to study flow control on cascaded connection with candidate switches.**
  - We did it to choose switches for EB1, 10 years ago.
  - **Probably Mellanox is OK**
  - **We are not sure about low-cost switches by DELL**
- Typical network switch doesn't send pause frame actively.
- ARISTA 7048T-A does, but its successor 7020 doesn't. 7280 doesn't also.
  - They use 802.1Qbb instead of former 802.3x
  - We have to study it works with existing ARISTA and with Linux
- **If the behavior is not what we expected, 10G-T and SFP+ options are bad.**

Summary

- Software will be usable without heavy rewrite
- Hardware choice is still difficult.
- Does anybody have strong motivation about 10G-T?
  - I don't have
  - as it requires expensive 40G-LR4 optics for switch in server room