Next Generation Data Center Connections in China

Liang Guo

CHAIR OF ODCC NEW TECHNOLOGIES AND TEST WG

Jie Li

VICE CHAIR OF ODCC

http://www.odcc.org.cn/introduction-en.html

OUTLINE

- **ODCC DCCNG Introduction**

- **Datacenter Development in China**

- **Requirements of next generation DC connections in China**

- **Technology Trends**

- **Conclusion**

http://www.odcc.org.cn/introduction-en.html

- **ODCC DCCNG Project**
- Datacenter Development in China
- Requirements of next generation DC connections in China
- Technology Trends
- Conclusion

http://www.odcc.org.cn/introduction-en.html

DCCNG Project

DCCNG is a joint project of New Technology & Testing WG and Network WG in ODCC, aimed to investigate the requirements of next generation data center connections in China and the potential technology directions.

Market Demands

1. 800G（100G access） time；
2. Connection distance；
3. Any new network architecture or device interconnection?
4. Type of module form factor
5. Power consumption and temperature
6. Fiber/cable choice
7. End to end latency；
8. ......

Participants of DCCNG:

Technology Trends

1. 8*100G or 4*200G or others?
2. Trends of components；
3. Direct detection vs. coherent；
4. New FEC?；
5. OBO ? Co-packaging；
6. ......

4

http://www.odcc.org.cn/introduction-en.html

Status of the project

- The whitepaper of this project has been published at 2019 ODCC Summit, September 4th in Beijing.

- It has also been invited to present in 2019 CIOE, September 5th in Shenzhen.

- The group will consider further research on specific market and technology direction based on feedbacks from the industry.

http://www.odcc.org.cn/introduction-en.html

OUTLINE

- ODCC DCCNG Project
- **Datacenter Development in China**
- Requirements of next generation DC connections in China
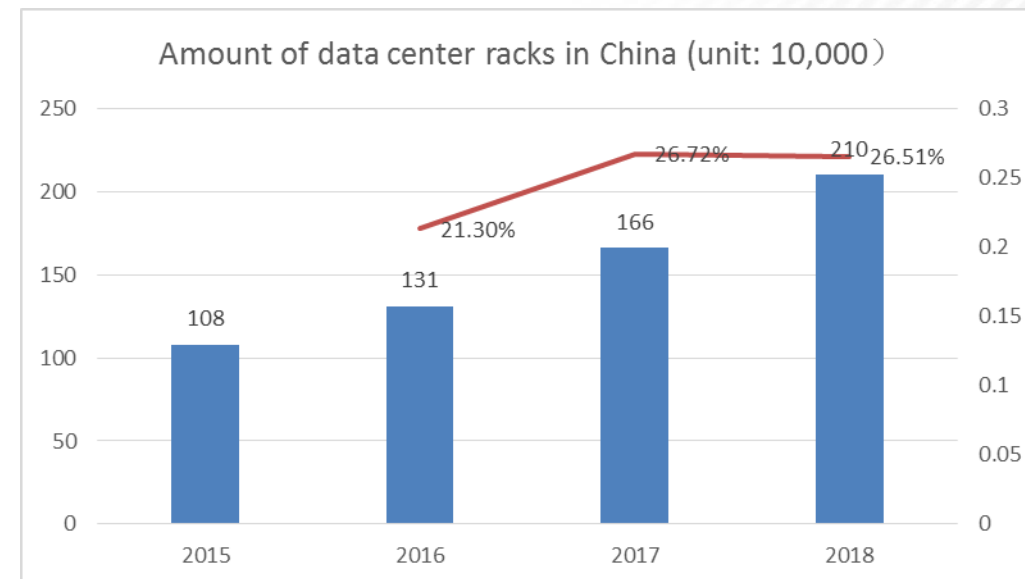- Technology Brief
- Conclusion

http://www.odcc.org.cn/introduction-en.html

The Growth of data center racks in China
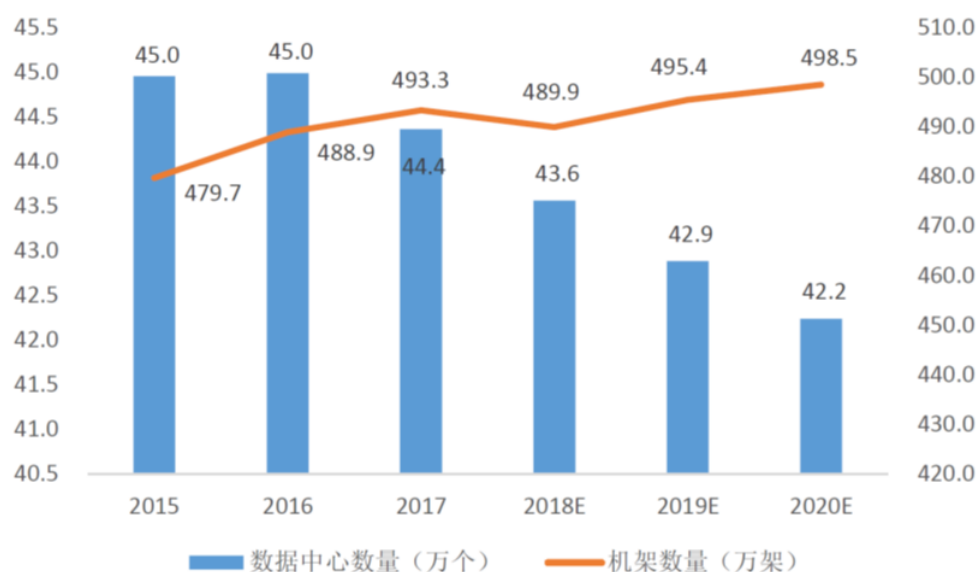
图 3 2015-2020 全球数据中心和机架数量统计及预测

数据来源：Gartner

Amount of data center racks in China (unit: 10,000）

Source: China Academy of Information and Communications Technology (CAICT)

Since 2015, the annual growth of racks in China data center market is over 20% and reached 2.1 million by the end of 2018, which is significant higher than that in 2015.
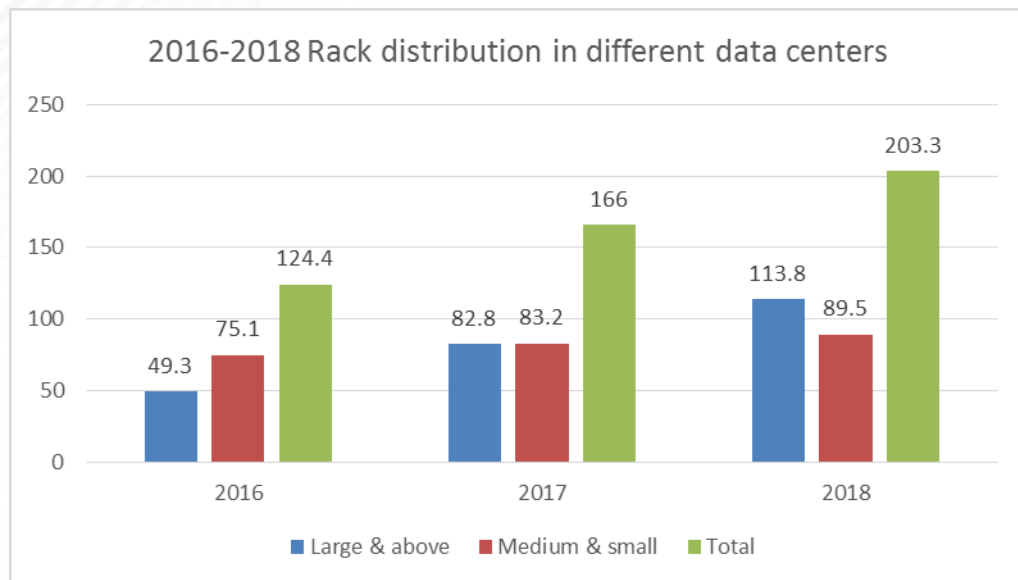
http://www.odcc.org.cn/introduction-en.html

Usage of Datacenter racks in the region

2016-2019 Distribution of data center racks in use

Source: China Academy of Information and Communications Technology (CAICT)
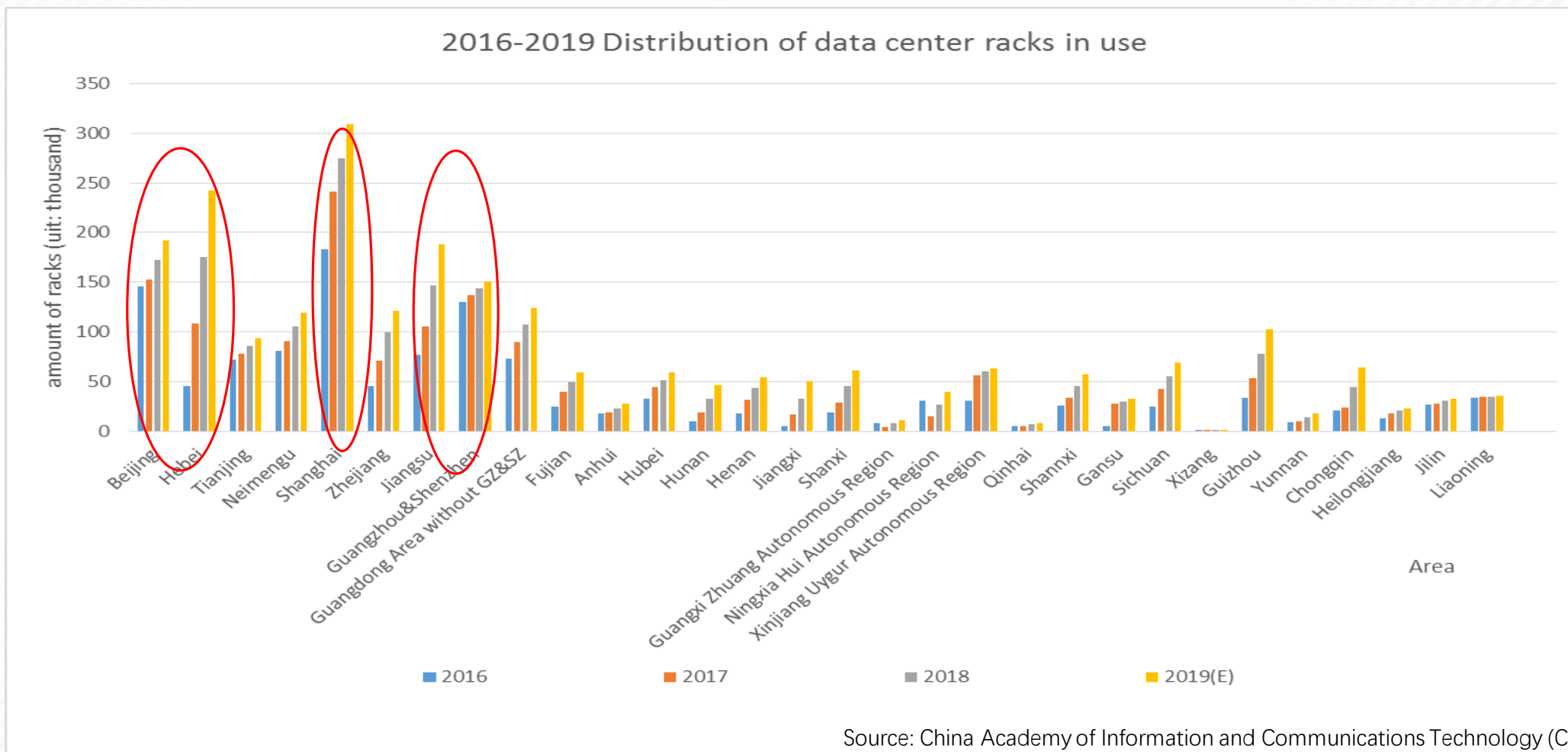
The Shanghai area and its surrounding Jiangsu Province, Beijing area and its surrounding Hebei Province, as well as Guangdong and Shenzhen are active areas of data center usage. The distribution of data centers in various regions of the country is unevenly distributed.

8

http://www.odcc.org.cn/introduction-en.html

The Scale of China's datacenter

2016-2018 Rack distribution in different data centers

	Large & above	Medium & small	Total
2016	49.3	75.1	124.4
2017	82.8	83.2	166
2018	113.8	89.5	203.3

Source: CAICT, MIIT

2016 Proportion of data center scale in China

16, 1%
109, 7%
1516, 92%

■ Super ■ Large ■ Meduim&small

2017 Proportion of data center scale in China

36, 2%
166, 9%
1642, 89%

■ Super ■ Large ■ Meduim&small

Source: China Academy of Information and Communications Technology (CAICT)
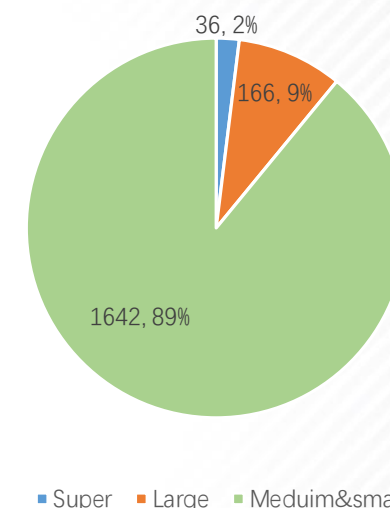
In 2017, the proportion of China's data center is that 11% large data centers occupy 50% of the number of total racks, while 89% of small and medium-sized data centers occupy the other half. From the perspective of the global data center growth, China's large and super data centers still have great development room to accommodate more racks and improve the overall data center construction and usage efficiency.
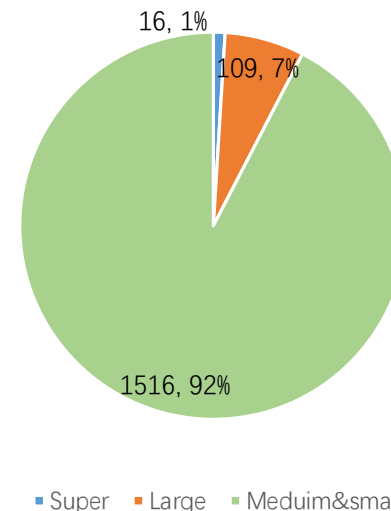
From the perspective of the proportion of data centers in China, in 2016, there were only 16 super large data centers which can support over 80,000 servers, and 106 large data centers with over 20,000 but below 80,000 servers. This number has doubled and increased 56% in 2017, respectively.

http://www.odcc.org.cn/introduction-en.html

Summary

- In recent years, China's data centers have maintained rapid development in terms of quantity, number of racks, distribution of regions and scales, and have room for sustained and rapid growth.

- From the perspective of global data center rack growth and data center scale, China's large data centers have great room for development, by accommodating more racks to improve the overall data center construction and efficiency.

http://www.odcc.org.cn/introduction-en.html

- ODCC DCCNG Project

- Datacenter Development in China

- **Requirements of next generation DC connections in China**

- Technology Brief

- Conclusion

http://www.odcc.org.cn/introduction-en.html

Evolution of server & GPU Interconnects

Router

Core

Leaf / Spine

TOR / Leaf
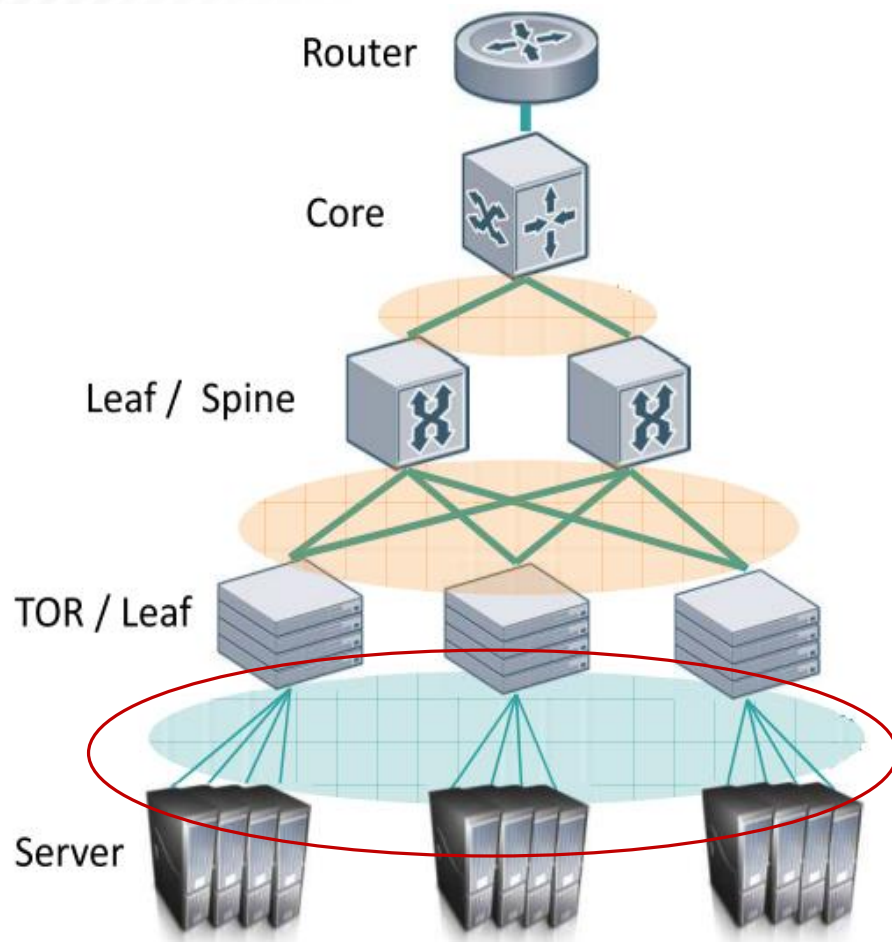
Server

1. Bandwidth

- Datacom companies started deploying 25G servers in 2017 with 100G ToR switches as the uplinks. Enterprise customers delay for 2~3 years.
- As for now, 25G access still dominates the market however 100G servers are expected to be deployed early next year.
- It is expected that the scale will be increased around 21Q4.
- The main application will be GPU clusters.
- 50G is not being considered.

2. Connection (server to ToR)

- 5m within cabinet; a small number of cross-rack interconnects up to 20m.
- DAC interconnect is used for 25G access.
- For 100G access, due to constraints of distance and deployment (the diameter becomes thicker, the degree of buckling and the compatibility interoperating testing between vendors become complicated), server connections may turn to AOC or multi-mode transceivers.
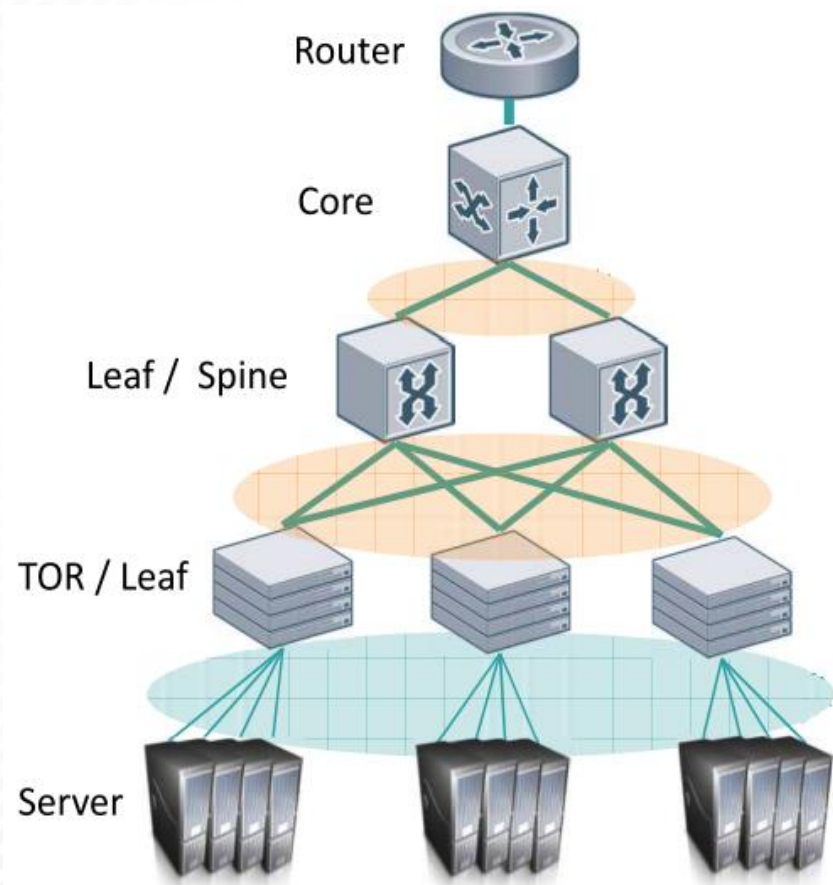
12

Switch

Router

Core

Leaf / Spine

TOR / Leaf
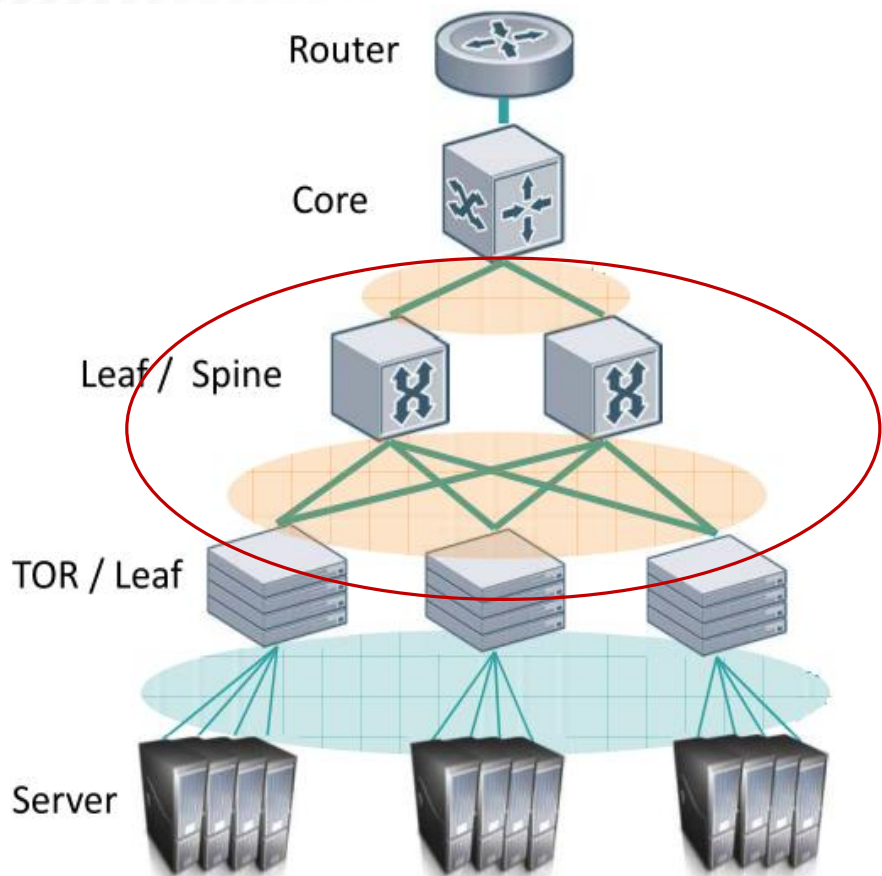
Server

3. ToR and ASIC

- It is expected that the ToR maintains the 48+8 configuration. For 100G server access, the corresponding ToR uplink would be 400G (cost efficient).
- When the leaf/spine exchange capacity reaches 20T+, it will support 800G ports. The requirement of upgrading to 800G is based on the data capacity of increased services and is expected to be reviewed by 2023. The data requirements for 1.6T are expected at a latter stage.
- Data centers interconnects may be the first area to deploy 800G.
- Further, each layer is expected to be a uniform design using the same box with the same chips.

4. Switch interconnections

- **ToR-to-Leaf**: mostly MMF deployments in existing data centers; 70m is a basic requirement for current deployments; while 100m reach is expected depending on cost-efficiency.
- **Leaf-above:** using CWDM SMF, 500m~2km.

13

http://www.odcc.org.cn/introduction-en.html

Package

Router

Core

Leaf / Spine

TOR / Leaf

Server

5. Package

- The form factor of 800G for data center networks is still expected to be pluggable.

- QSFP-DD and OSFP are potential options for 800G.

- For 1.6T, COBO or co-package might be considered.

- <u>Low power consumption is one of the key requirements.</u>

http://www.odcc.org.cn/introduction-en.html

- ODCC DCCNG Project

- Datacenter Development in China

- Requirements of next generation DC connections in China

- **Technology Brief**

- Conclusion

http://www.odcc.org.cn/introduction-en.html

Technology Brief

- **Electrical**： 100G SerDes IO would be the electrical interface for both next generation 100G server access and 800G switch interconnect.
- **MMF:** Several directions to achieve 100G per wavelength are being considered.
- **Package options:** pluggable is still the choice for 800G, while co-package seems to be the inevitable trend for high bandwidth interfaces, but more work is required.
- **High-capacity ASIC:** manufacture improvements and chiplet might be the key.
- **Direct detection vs. Coherent**: for the cost-sensitive data center market, 800G (<2km) may still be covered by IM-DD while coherent optical communication becomes a new possibility for data center inter-connections.

http://www.odcc.org.cn/introduction-en.html

- ODCC DCCNG Project

- Datacenter Development in China

- Requirements of next generation DC connections in China

- Technology Brief

- **Conclusion**

http://www.odcc.org.cn/introduction-en.html

Conclusion

- The industry needs to be ready in advance for next generation 100G server access and 800G interconnect technologies to support the evolution of cost-efficient connections for data center networks in the next 5~6 years, allowing them to cope with the development of internet services, HPC, AI and distributed storage applications.

- Due to its large market share, MMF might still have a cost advantage and is expected to evolve to 100G per wavelength.
  - 30m is a reasonable first goal for MMF links for AOC or transceivers in server interconnects.
  - Longer distance MMF transceivers are needed for switch interconnects with transceivers.

- Pluggable is preferred in 800G while co-package is an inevitable trend, however technical challenges should be resolved by the entire industry before adoption.

- With low complexity and low power consumption, IM-DD might be the choice for 800G below 2km, while Ethernet coherent might be a new possibility for data center inter-connects.

18

http://www.odcc.org.cn/introduction-en.html

THANKS!

http://www.odcc.org.cn/introduction-en.html