Experiences on Buffer Size

Hongqiang "Harry" Liu, *Staff Engineer*
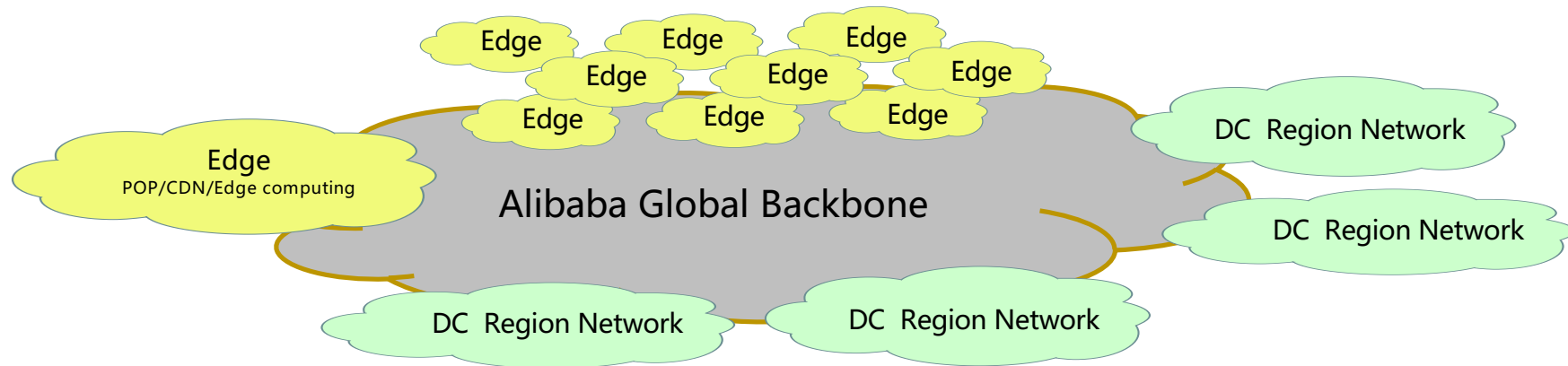**Yiqun Cai,** *VP of Networking*

AliCloud & Alibaba Infrastructure Service

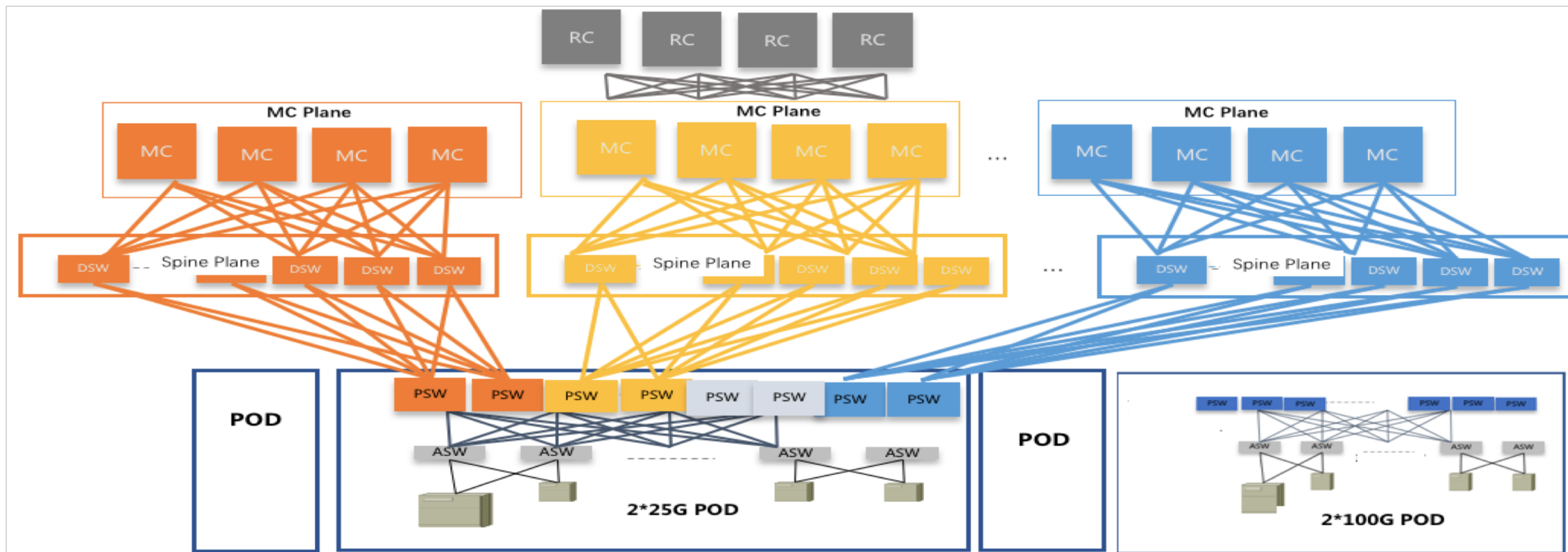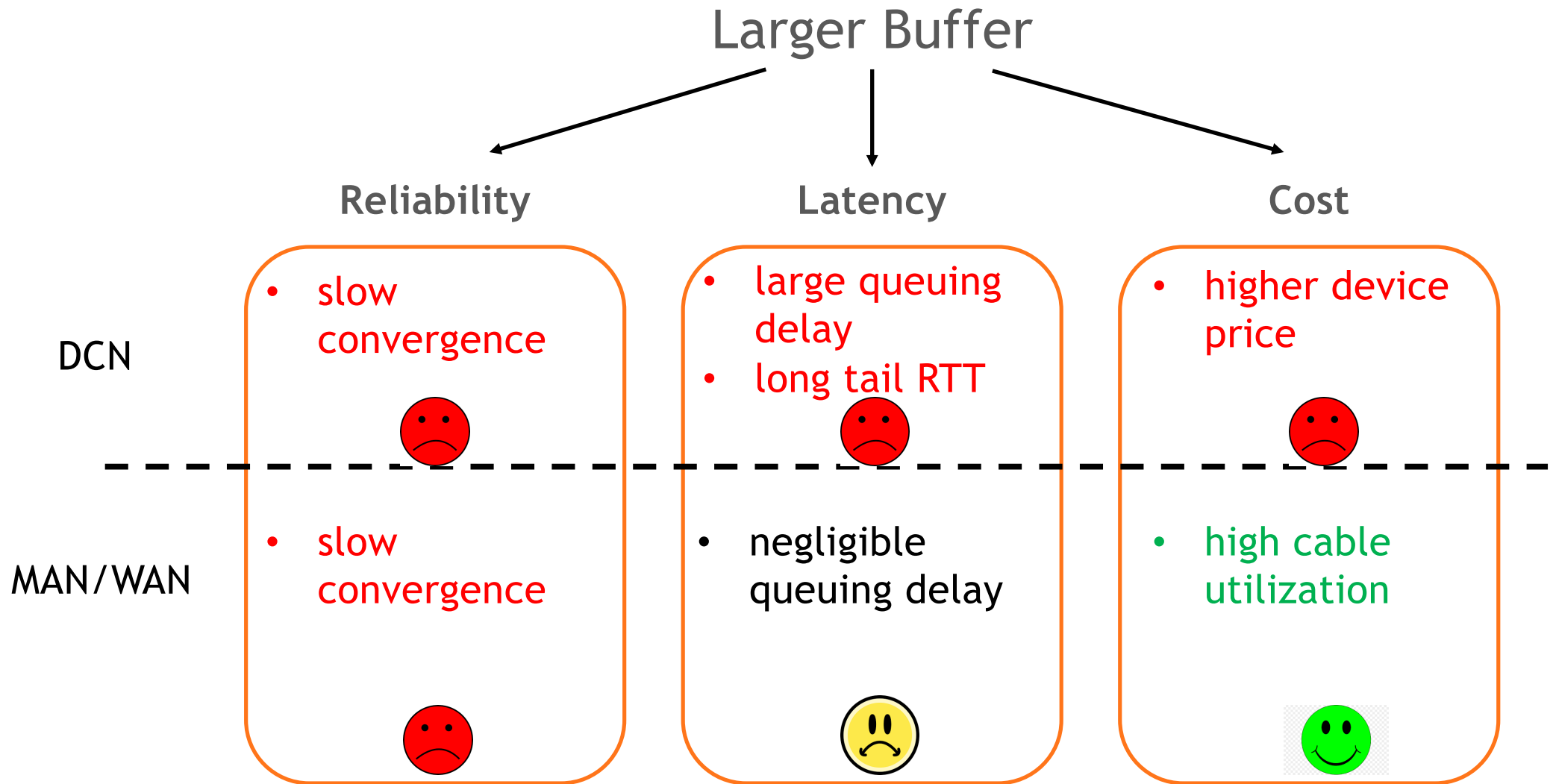**March 5th, 2019**

The Alibaba Ecosystem

Global Network Infrastructure & Hyper-scale Edge Network

Edge

Edge

Edge

Edge

Edge

Edge

Edge

Edge

Edge

Edge
POP/CDN/Edge computing

Alibaba Global Backbone

DC Region Network

DC Region Network

DC Region Network

DC Region Network

Hyper-scale Data Center Network

Larger Buffer

	Reliability	Latency	Cost
DCN	• slow convergence	• large queuing delay • long tail RTT	• higher device price
MAN/WAN	• slow convergence	• negligible queuing delay	• high cable utilization

Some special cases

- **Private clouds for traditional enterprise networks**
  - demanding larger buffer for large bandwidth-delay product (BDP)

- **Live videos**
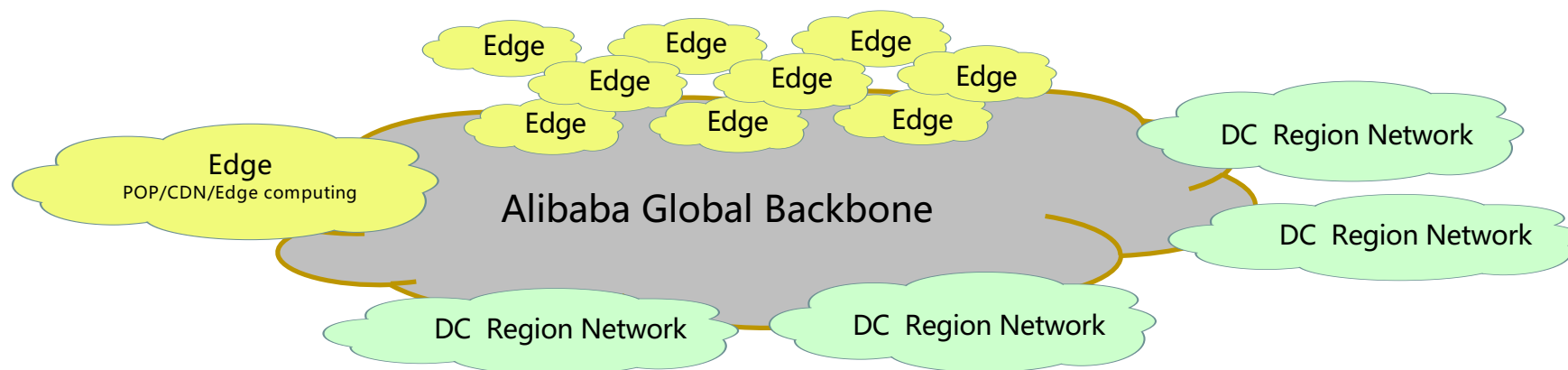 - demanding larger buffer for super-spiky traffic patterns

- ## Better congestion control
  - reducing the needs of in-network buffering, e.g. HULL

- ## Better topology design
  - making a more reasonable over-subscription ratio of switch links

- ## Better buffer monitoring
  - micro-second granularity buffer monitoring is necessary

Q&A

Thanks

Global Network Infrastructure & Hyper-scale Edge Network

Edge (×multiple)

Edge
POP/CDN/Edge computing

Alibaba Global Backbone

DC Region Network

DC Region Network

DC Region Network

DC Region Network

Hyper-scale Edge Network
Millions of servers over 10-100s of 1000s of edge computing sites
Operation simplicity
Cost optimized

Intelligent Global Network Orchestration System (Alibaba NetO System)
Best user experience: path selection, failure mitigation
Optimized link utilization
Overlay and Underlay integration

In-house Built Network Security System
High Performance Platform
Programmable HW chip
Anti-DDoS
IDS/IPS, WAF, DLP

Hyper-scale Data Center Network
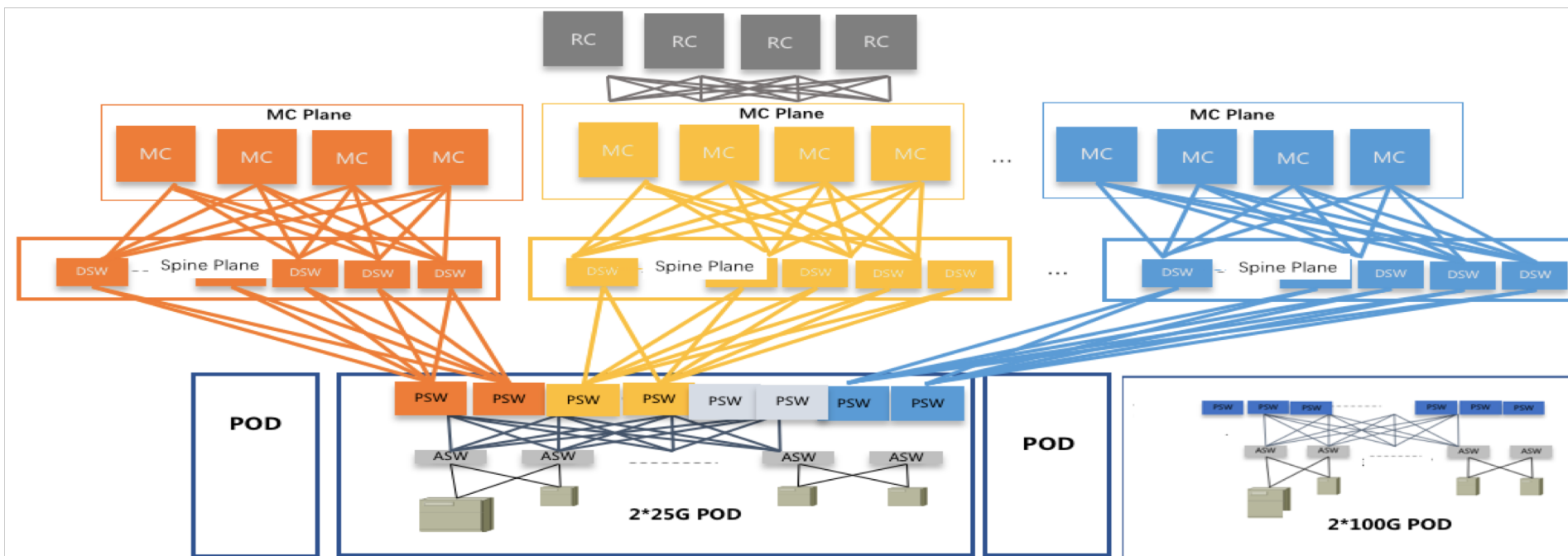
Hyper-scale DC Network
100s of 1000s of servers per DC region
Bandwidth Evolution:
25G→100G→400G
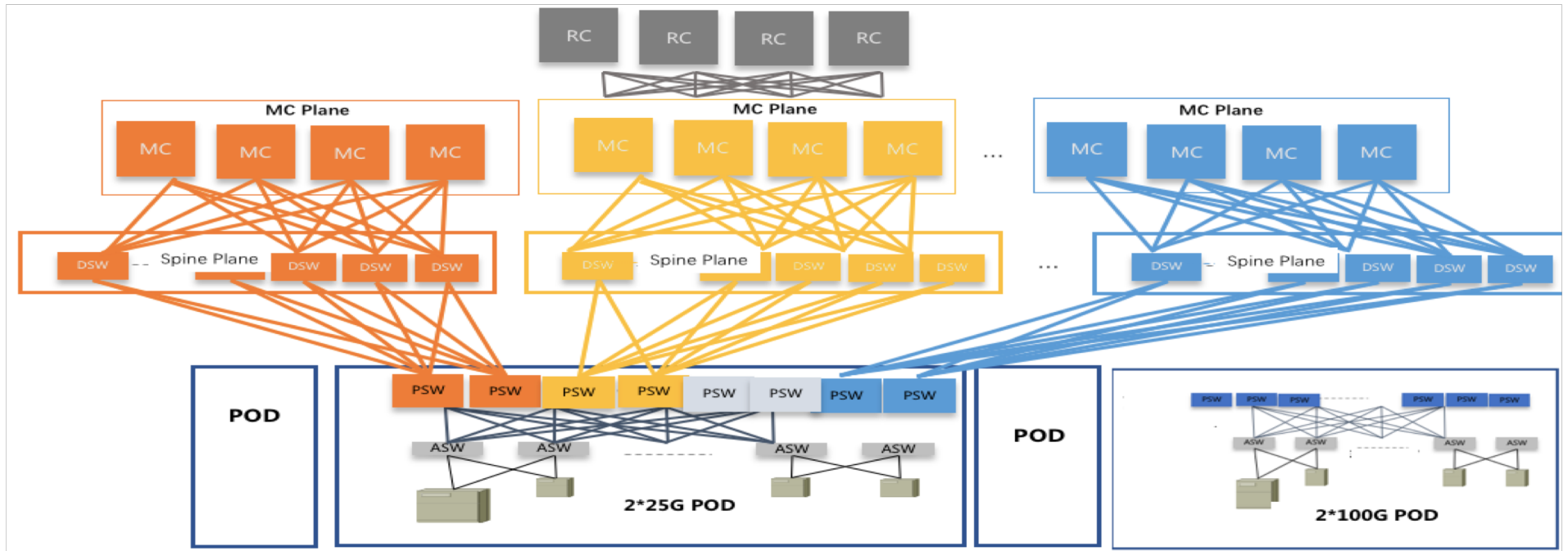Super low network latency
Cost optimized

SW&HW Development
DC Switch, High Performance Software Platform, L4-L7 Gateway, BGP SW Stack, 25G/40G/100G/400G Optics
Open Optical Transport System
Large-scale RDMA

Data-driven Network
Network Telemetry, flow-level granularity (INT)
Data-driven traffic engineering
Self-driving network