pre-GDB workshop: data-centre network architectures

Introduction

E.M. & S.Z.

7-8 June 2021

Introduction

The idea for a workshop dedicated to Data Center Networking matured during last LHC-OPN,LHCONE meeting to have an opportunity to discuss the most common challenges and to see which solutions are being adopted to satisfy experiments requirements on different sites.

Network requirements for HEP data centers are changing rapidly in terms of bandwidth and functionalities.

- Network data challenges in preparation for HL-LHC are ”Rising the bar” from the throughput point of view.
- The massive adoption of “Cloud technologies” like orchestrated virtual systems (VMs and Containers) and multitenancy are pushing for new functionalities
- Integration between Virtual Machines / Containers networking and the physical network infrastructure is an important point to investigate further to make our datacenters more effective on supporting the computing models that are evolving in WLCG.

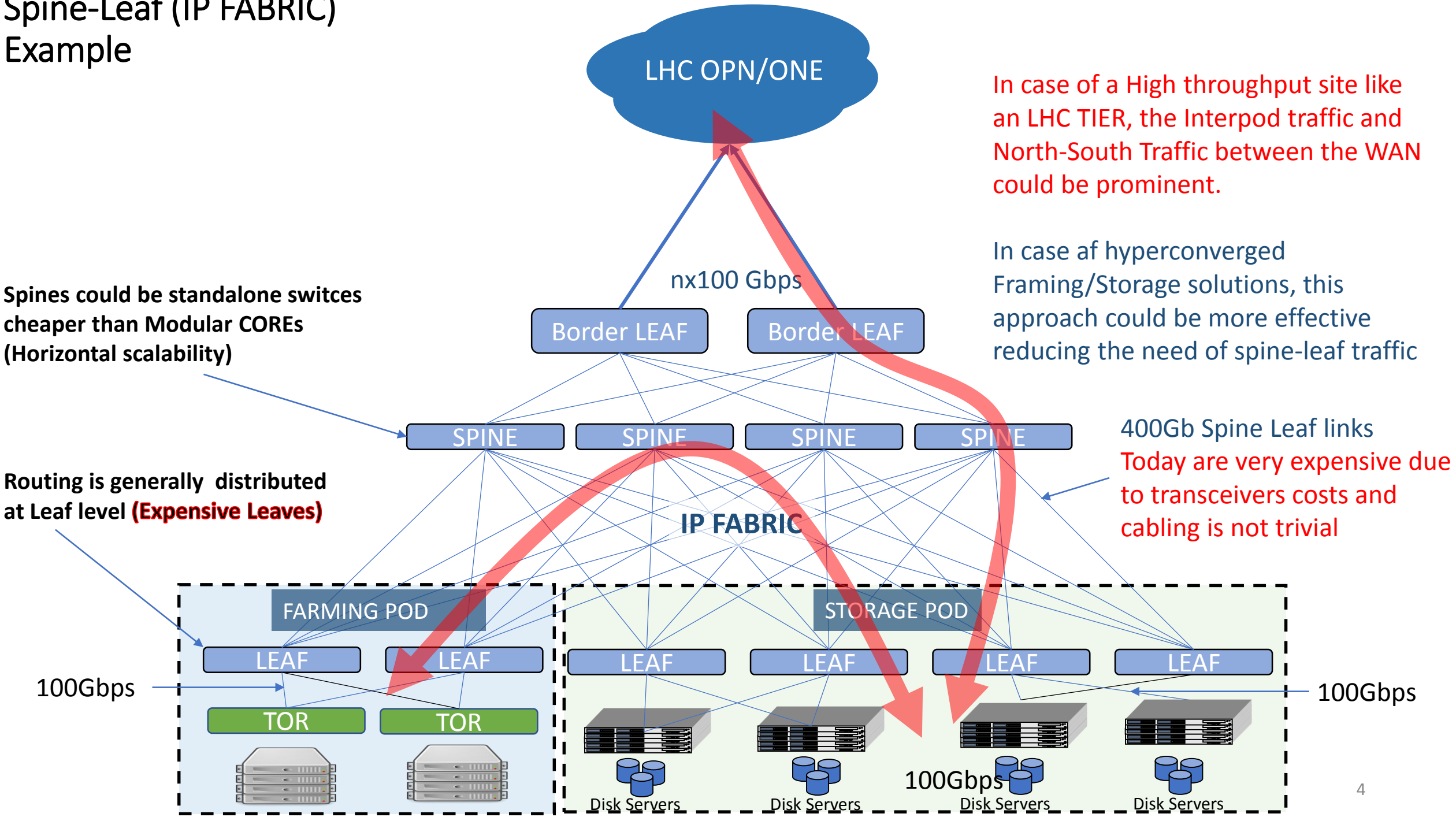
Physical topologies and protocols

There is a wide literature on cloud native data center networking and there are trends (also promoted by the vendors) on the adoption of a Spine-Leaf clos topology with a Layer 3 underlay (IP FABRIC) network and a L2/L3 overlay.

This approach is aiming to:

- Get rid of spanning tree protocol (or proprietary solutions like Mlag and vPC)
- Add flexibility to the network, facilitating ”Vmotion” inside the datacenter and between different datacenters when needed.
- Improve scalability and resiliency to the network

Spine-Leaf (IP FABRIC) Example



Spines could be standalone switces cheaper than Modular CORES (Horizontal scalability)

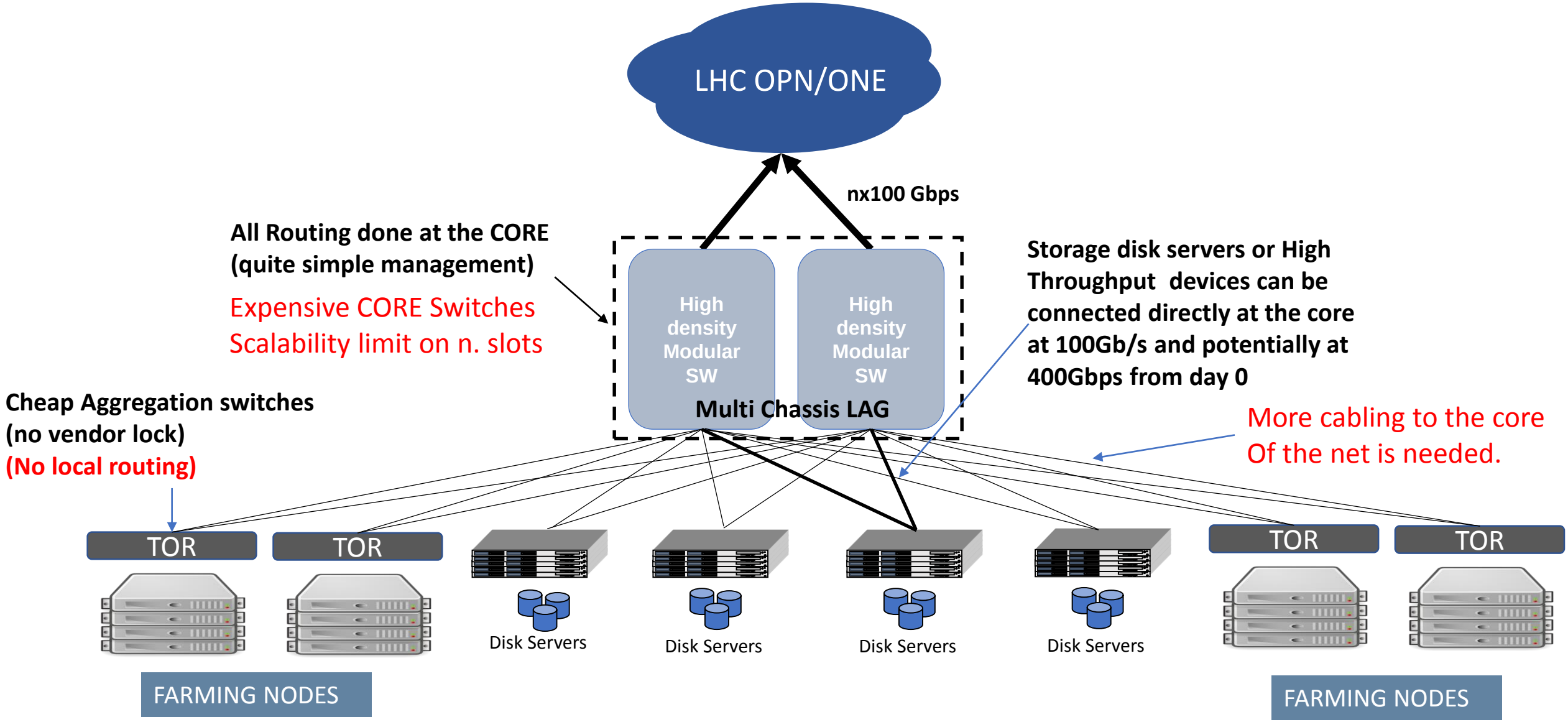
Routing is generally distributed at Leaf level (Expensive Leaves)

In case of a High throughput site like an LHC TIER, the Interpod traffic and North-South Traffic between the WAN could be prominent.

In case af hyperconverged Framing/Storage solutions, this approach could be more effective reducing the need of spine-leaf traffic

400Gb Spine Leaf links Today are very expensive due to transceivers costs and cabling is not trivial

(Traditional CORE-EDGE) Example



Some aspects to take in care

The advantages of an IP fabric are relevant but there are factors to consider:

- Real data flows inside and between HEP datacenters
- Costs of the network components like
  - Switches (routers) with Vxlan Evpn Routing and BGP capabilities
  - Cabling system and Transceiver optics
- Complexity in configurations, maintenance and troubleshooting
  - Is it necessary management software? In most cases management software and SDN controllers are proprietary →Risk of Vendor Lock-In. In principle it is possible to build a multivendor Spine-Leaf IP Fabric but in reality is not easy to implement and maintain.
  - Improvement on network automation tools is certainly necessary.
- Performance aspects related to the overlay implementation (ASIC on Net Devices? Encapsulation/Decapsulation offload implemented directly on NIC?)
- Security aspects of virtual overlay network have to be considered: Difficult to implement security on demarcation points → Distributed virtual firewalls or security agents?

Let’s start!

This workshop should be a good opportunity for exchanging ideas and experiences between the main HEP datacenters and Research Networks people.

Contributions are coming from CERN, the sites and the NRENs

Let’s Start!