CERN Computer Centre(s) Network evolution
(Part I)

Vincent DUCRET – vincent.ducret@cern.ch

HEPiX Spring 2023 – Taipei

Agenda

Part I:

- Reminder about 2019 status
- Datacentre migration (during COVID19 lockdown - 2020)
- Overview of current Datacentre Network
- Evolution of links between Main datacentre and other CERN sites
- Plans for 2023 and new Prévessin Data Centre (PDC)

Part II:

- New tools and/or features we started to deploy
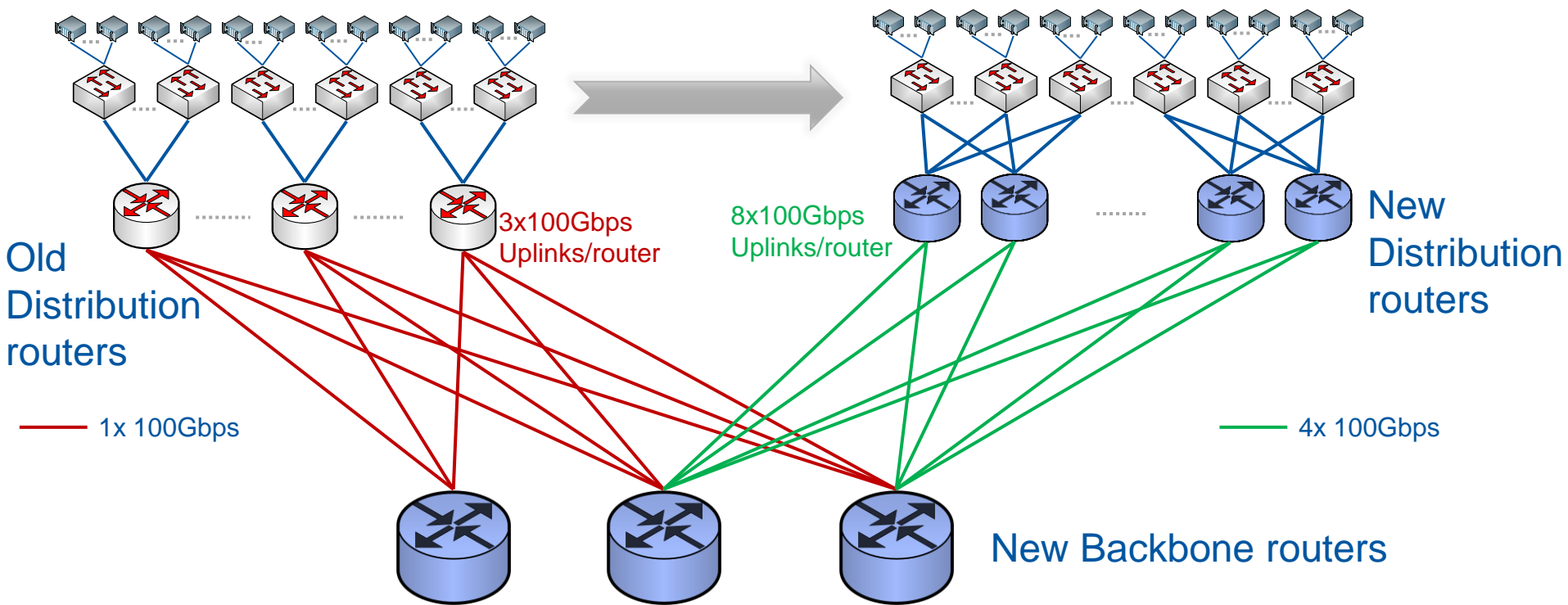- Main issues faced with new Datacentre Network setup

Reminder

- HEPiX Autumn 2019 presentation:
  - Datacentre backbone replaced by new Juniper devices
  - Pending migration to new Juniper distribution routers with dual router setup based on VxLAN ESI
  - WIGNER decommissioning and CPU nodes move to LHCb containers
  - Start moving connection between DC and Experiment directly to the DC Backbone routers

https://indico.cern.ch/event/810635/contributions/3592876/

Datacentre Migration

Old Distribution routers

3x100Gbps Uplinks/router

8x100Gbps Uplinks/router

New Distribution routers

1x 100Gbps
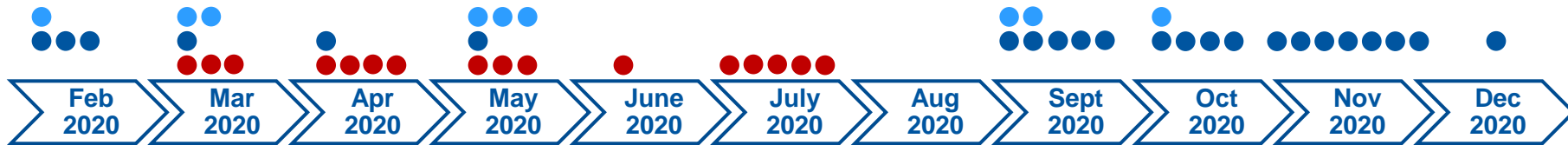
4x 100Gbps

New Backbone routers

Datacentre Migration

- Migration had to be done during 2020 (end of support of old routers)

- Reorganisation due to COVID-19 lockdown:
  - Minimise number of people in the room
  - Schedule intervention (early morning) to lower impact for teleworkers
  - Re-schedule some interventions for critical system

Datacentre Migration

- Total of 48 interventions done between February and December 2020
  - Fibre preparation done in advance by technicians
  - Once fibre ready, on-site intervention for migration (max two persons in the room)

● Afternoon
● Morning (working hours)
● Early morning (6:30-8:00am)

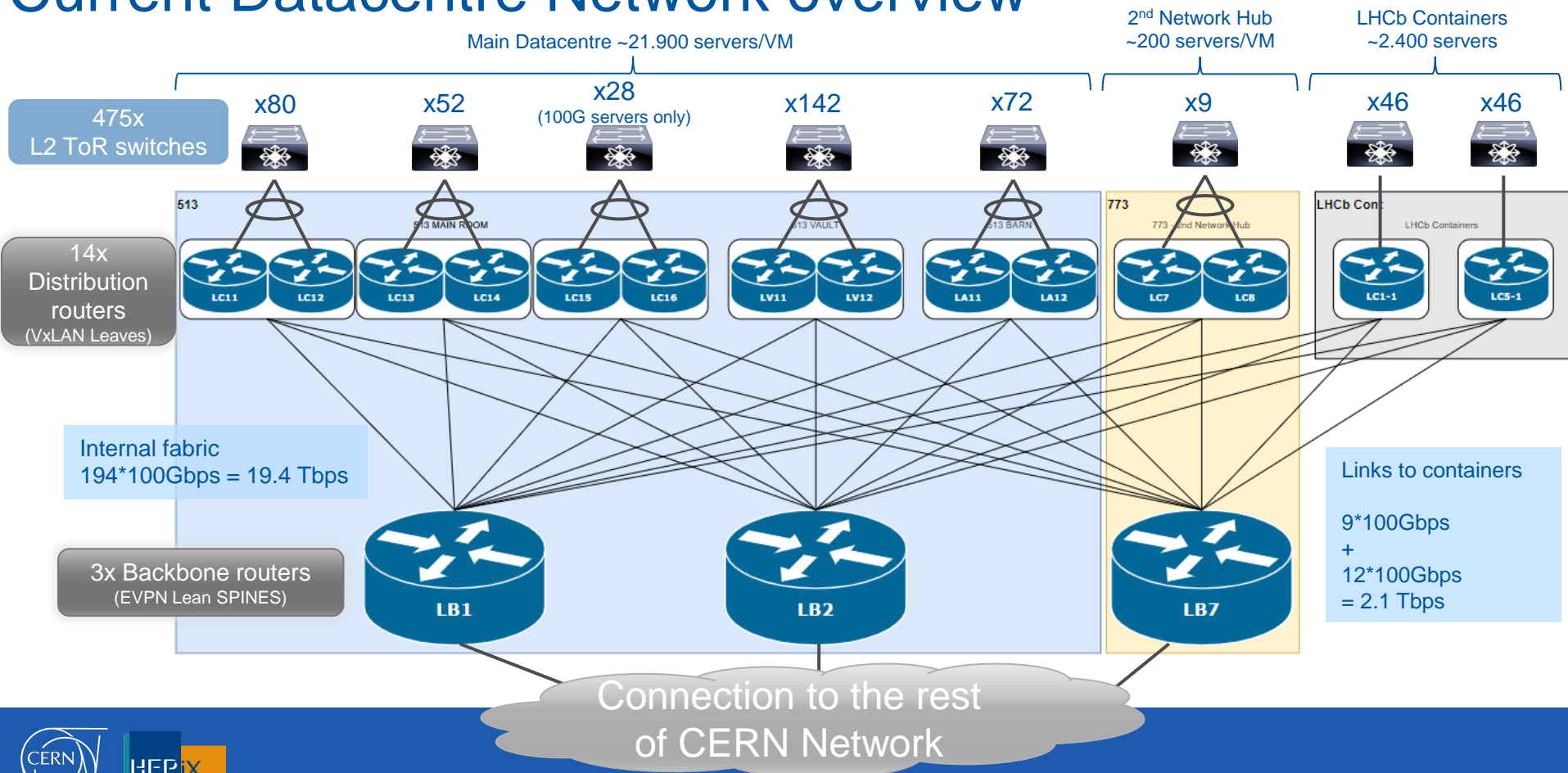| Feb 2020	Mar 2020	Apr 2020	May 2020	June 2020	July 2020	Aug 2020	Sept 2020	Oct 2020	Nov 2020	Dec 2020

Datacentre Migration

Protections evolutions

Supermoon above CERN DC
April 8th 2020 6:08 am →

Current Datacentre Network overview

2nd Network Hub
~200 servers/VM

LHCb Containers
~2.400 servers

Main Datacentre ~21.900 servers/VM

475x
L2 ToR switches

x80   x52   x28
(100G servers only)   x142   x72   x9   x46   x46

14x
Distribution
routers
(VxLAN Leaves)

513   LC11   LC12   513 MAIN ROOM   LC13   LC14   LC15   LC16   513 VAULT   LV11   LV12   513 BARN   LA11   LA12   773   2nd Network Hub   LC7   LC8   LHCb Cont   LHCb Containers   LC1-1   LC5-1

Internal fabric
194*100Gbps = 19.4 Tbps

Links to containers

9*100Gbps
+
12*100Gbps
= 2.1 Tbps

3x Backbone routers
(EVPN Lean SPINES)

LB1   LB2   LB7

Connection to the rest
of CERN Network

Evolution of links between Main datacentre and other CERN sites

- ## LHCb containers (used by CERN IT)
  - 2 containers dedicated for CPU/Batch nodes (total of ~2.400 servers)
  - Link to 513 (main DC) evolved
    - Initial setup: 4x100Gbps (connected to two DC Backbone routers)
    - Jan 2021: 6x100Gbps (connected to three DC Backbone routers)
    - Apr 2022: 9x100Gbps or 12x100Gbps (connected to three DC Backbone routers)
  - Still some peaks due to the nature of traffic – not considered as an issue.
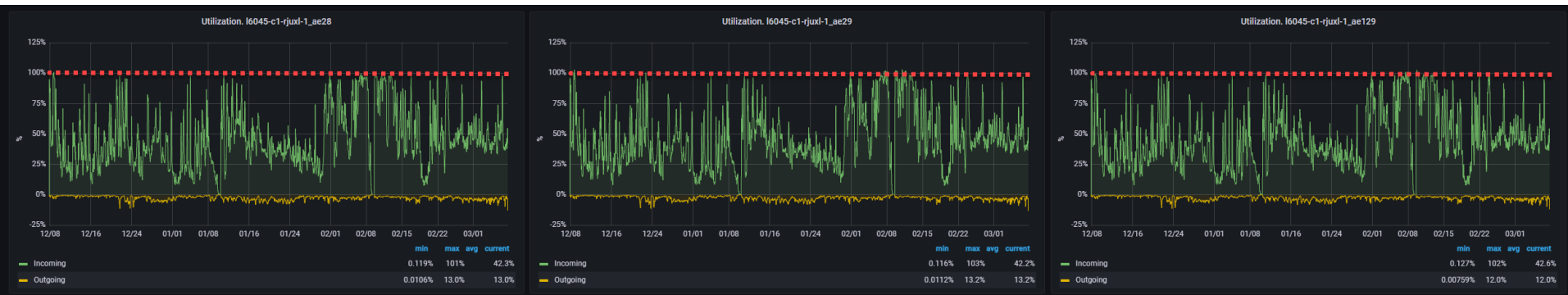
Evolution of links between Main datacentre and other CERN sites

- ## Central Data Recording (CDR) links (between DC and experiments DAQ)
  - ### All moved to DC Backbone routers and attached to two routers

| Site	Former connection	Current Connection
ALICE	4x 40Gbps	2x 100Gbps
ALICE O2	-	24/32x 100Gbps *(based on DWDM)*
ATLAS	8x 10Gbps	2x 100Gbps
CMS	4x 40Gbps	4x 100Gbps
LHCb	2x 10Gbps	4x 100Gbps
NA61	-	1x 100Gbps (+1x 100Gbps standby)
Protodune NP02+NP04	2x 40Gbps (no redundancy between sites)	2x 40Gbps (redundancy between sites)

Evolution of links between Main datacentre and other CERN sites

- ### ALICE O2 container

  - Host an Infiniband setup managed by ALICE team

  - DWDM system to build a 2.4 Tbps connection to main datacentre

  - Network design evolved:

    - Initial need = 1Tbps ➔ we build 2x 1.2Tbps to provide full redundancy

    - Local 100G servers required ➔ replace the small routers with a QFX10008 chassis

    - Current need: 200GBps = 1.6Tbps redundant ➔ we will increase to 2x 1.6Tbps (add WDM optics)

  More details on the dedicated presentation from HEPiX Autumn 2022:
  https://indico.cern.ch/event/1200682/contributions/5087580/

Evolution of links between Main datacentre and other CERN sites
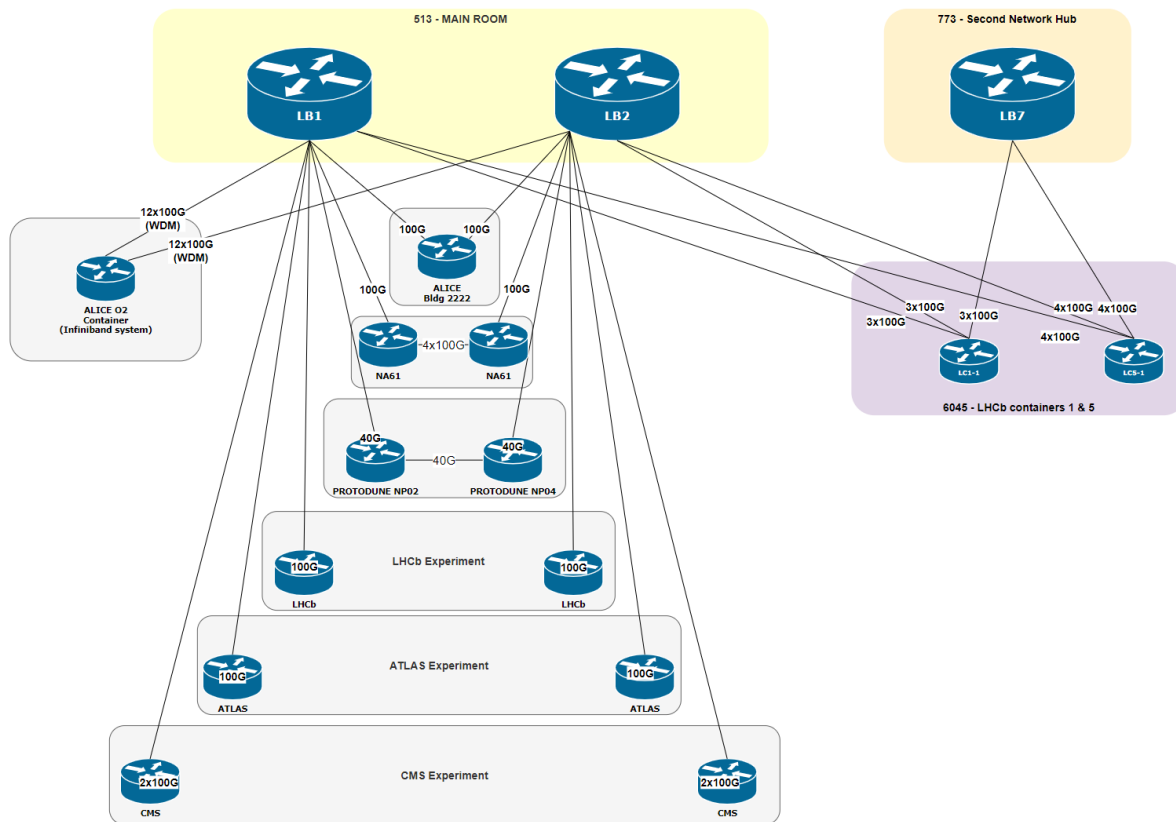
What's Next

- Prévessin Data Centre (PDC)
- Replace current Backbone routers:
  - Migrate to Juniper PTX10008
  - Support 400G (required for links to PDC)
- Enhance ZTP process: *(see details in presentation Part II)*
  - merge "campus" and "DC" ZTP process
  - ease "switch replacement"
  - integrate ZTP logs with our configuration tool
- Test "SDN" with Open Virtual Network *(driven by IT-CD group)*

Prévessin Data Centre

February 2023

Architecture view
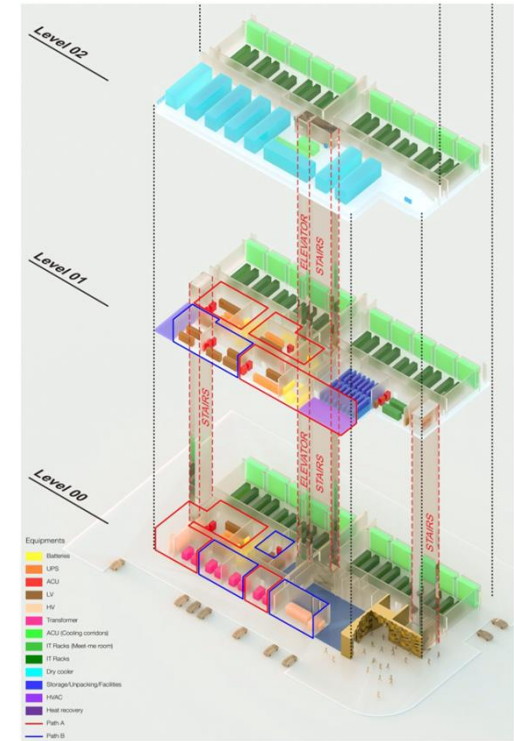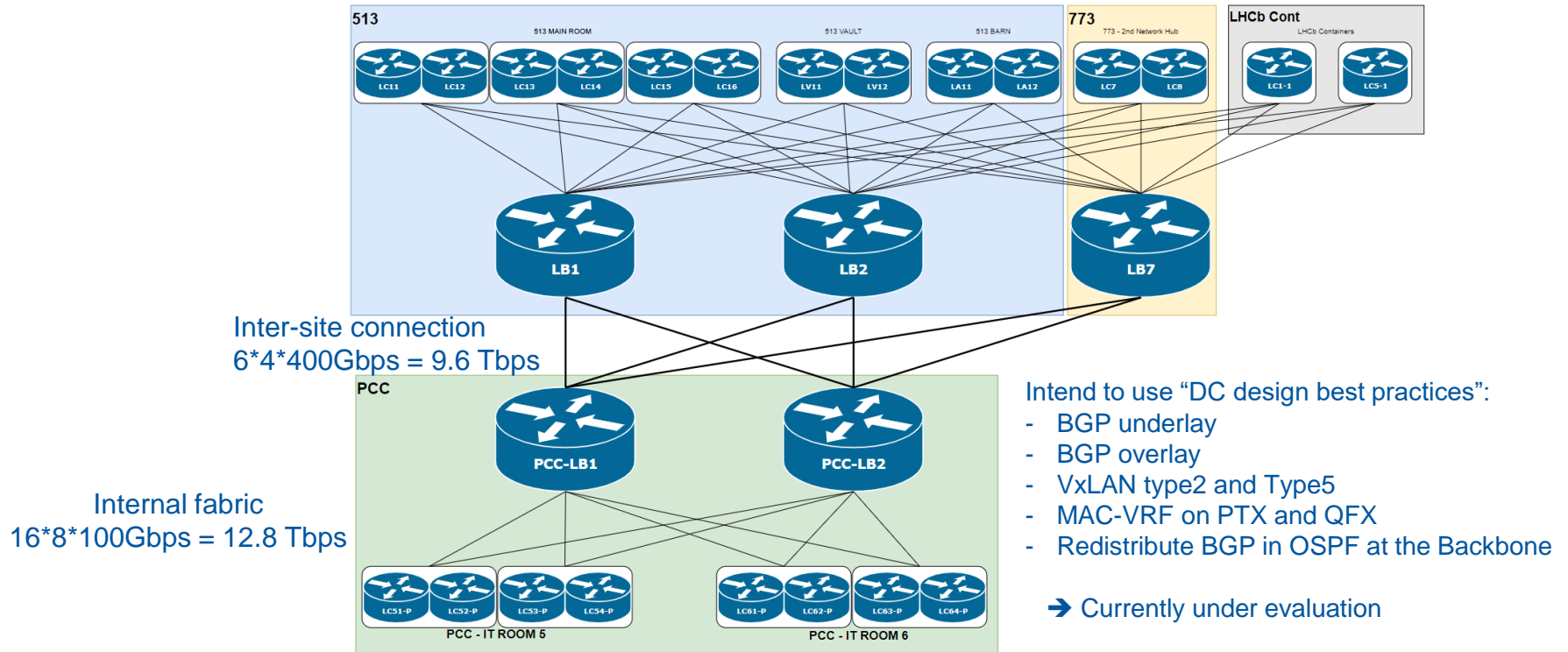
Prévessin Data Centre

- 2 IT rooms in 2023 (Phase I)
    - Total 4 MW
    - Total of 108 racks for CPU/Batch nodes
    - Total of 48 racks for Business Continuity and other IT Services nodes
- 2 more IT rooms at Phase II (~2026)
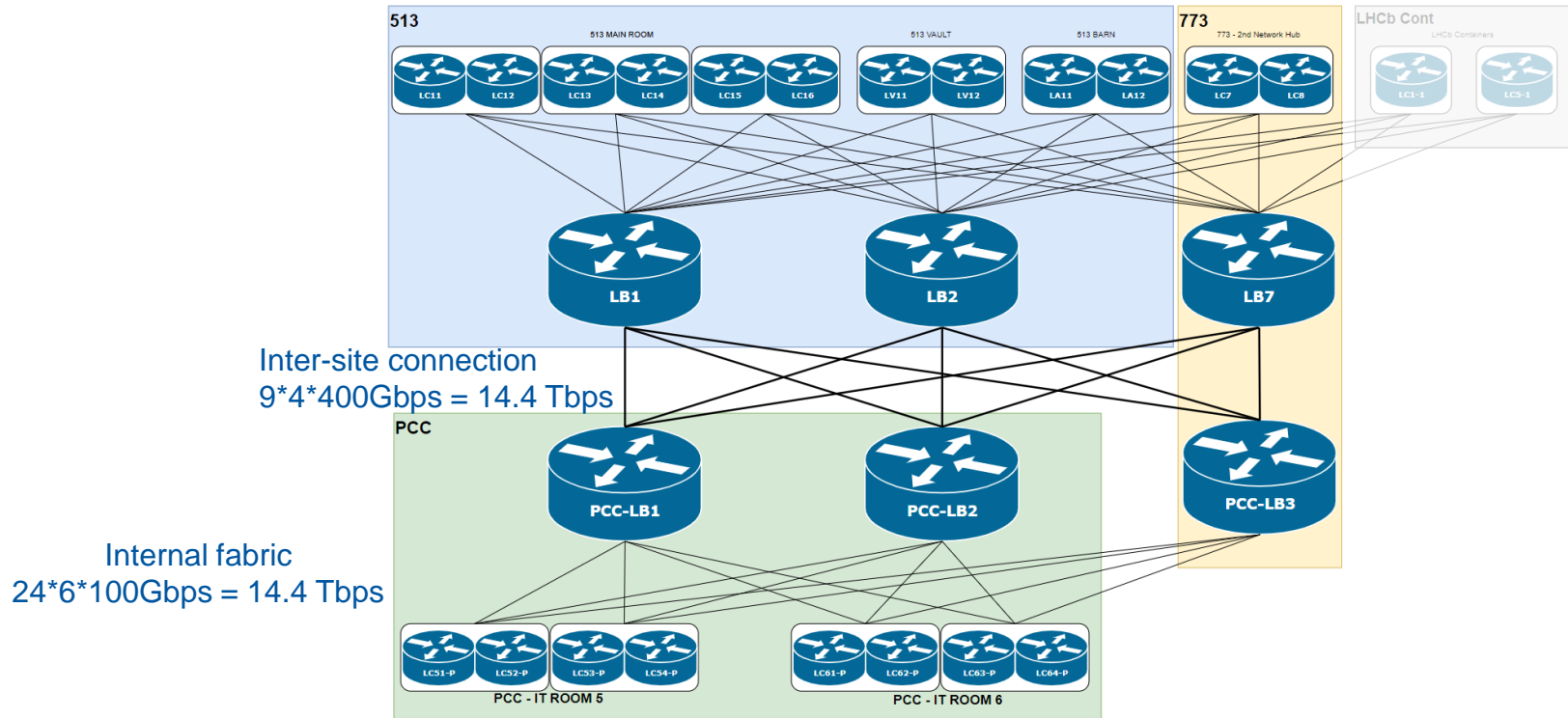- 2 more IT rooms at Phase III (~2029)

Prévessin Data Centre

Inter-site connection
6*4*400Gbps = 9.6 Tbps

Internal fabric
16*8*100Gbps = 12.8 Tbps

Intend to use "DC design best practices":
- BGP underlay
- BGP overlay
- VxLAN type2 and Type5
- MAC-VRF on PTX and QFX
- Redistribute BGP in OSPF at the Backbone

→ Currently under evaluation

Prévessin Data Centre

Inter-site connection
9*4*400Gbps = 14.4 Tbps

Internal fabric
24*6*100Gbps = 14.4 Tbps

Q&A