Testo opzionale

Consortium GARR

RETE GARR

WHOLESALE WINERY FT Tour

Marco Marletta

La nuova rete GARR-T: progetto e processo di migrazione

Chi è GARR

Consortium GARR

La comunità degli utenti

Oltre 1000 siti connessi appartenenti a varie organizzazioni

- 100 Università
- 350 Istituti e Laboratori di Ricerca
- 60 Istituti Biomedici di Ricerca
- 65 Conservatori, Biblioteche, Musei ed Istituzioni Culturali
- oltre 1000 scuole di cui più di 150 direttamente connesse

Enti ed Istituzioni vigilate da vari ministeri:

- Ministero dell'Università e della Ricerca
- Ministero dello Sviluppo Economico
- Ministero della Salute
- Ministero della cultura
- Ministero delle Politiche Agricole e Forestali

Ministero dell'Università e della Ricerca

MINISTERO DELLA CULTURA

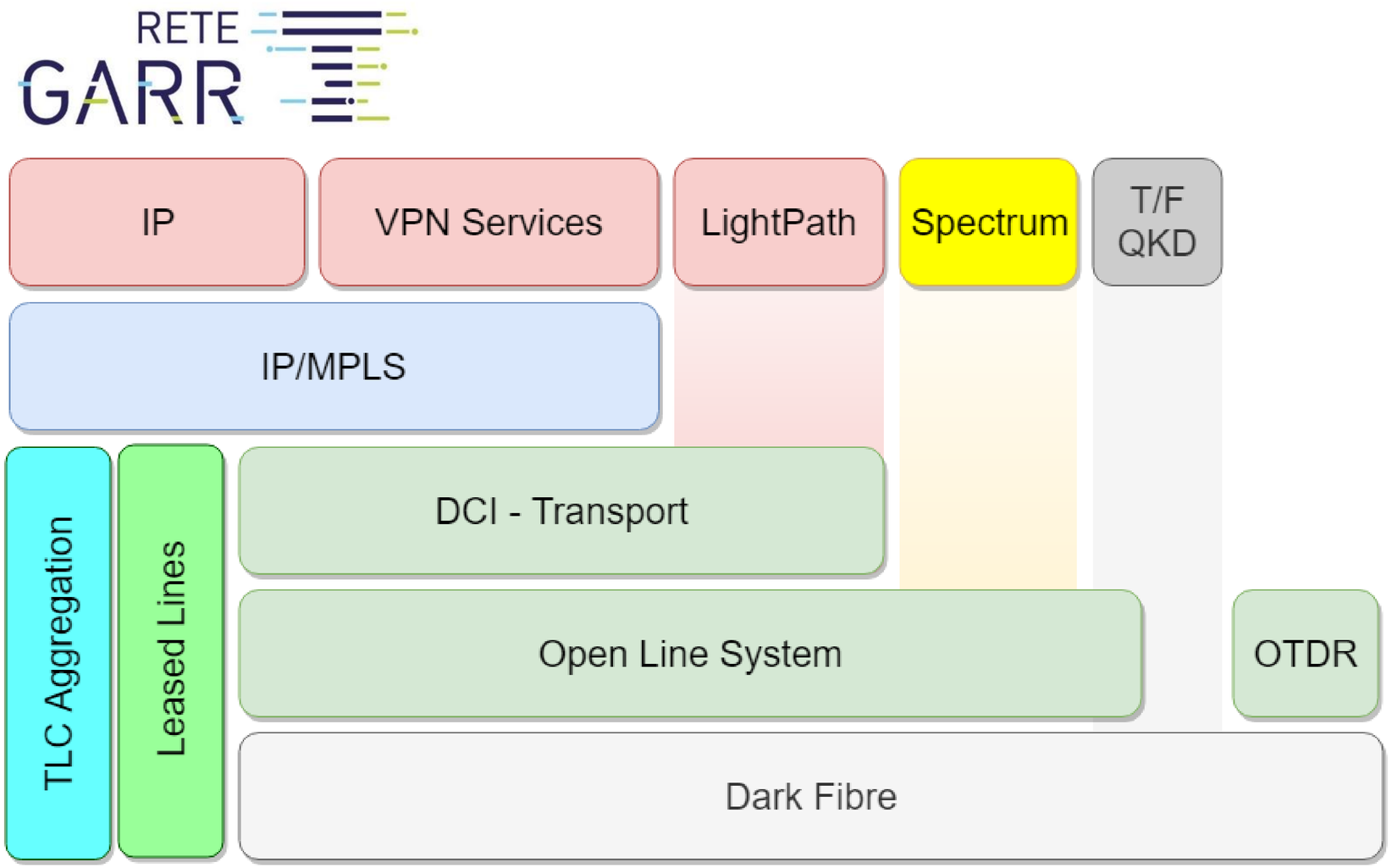
MINISTERO DELLE POLITICHE AGRICOLE ALIMENTARI E FORESTALI

Ministero della Salute

Ministero delle Imprese e del Made in Italy

L’idea e i requisiti

Architettura rete GARR-T

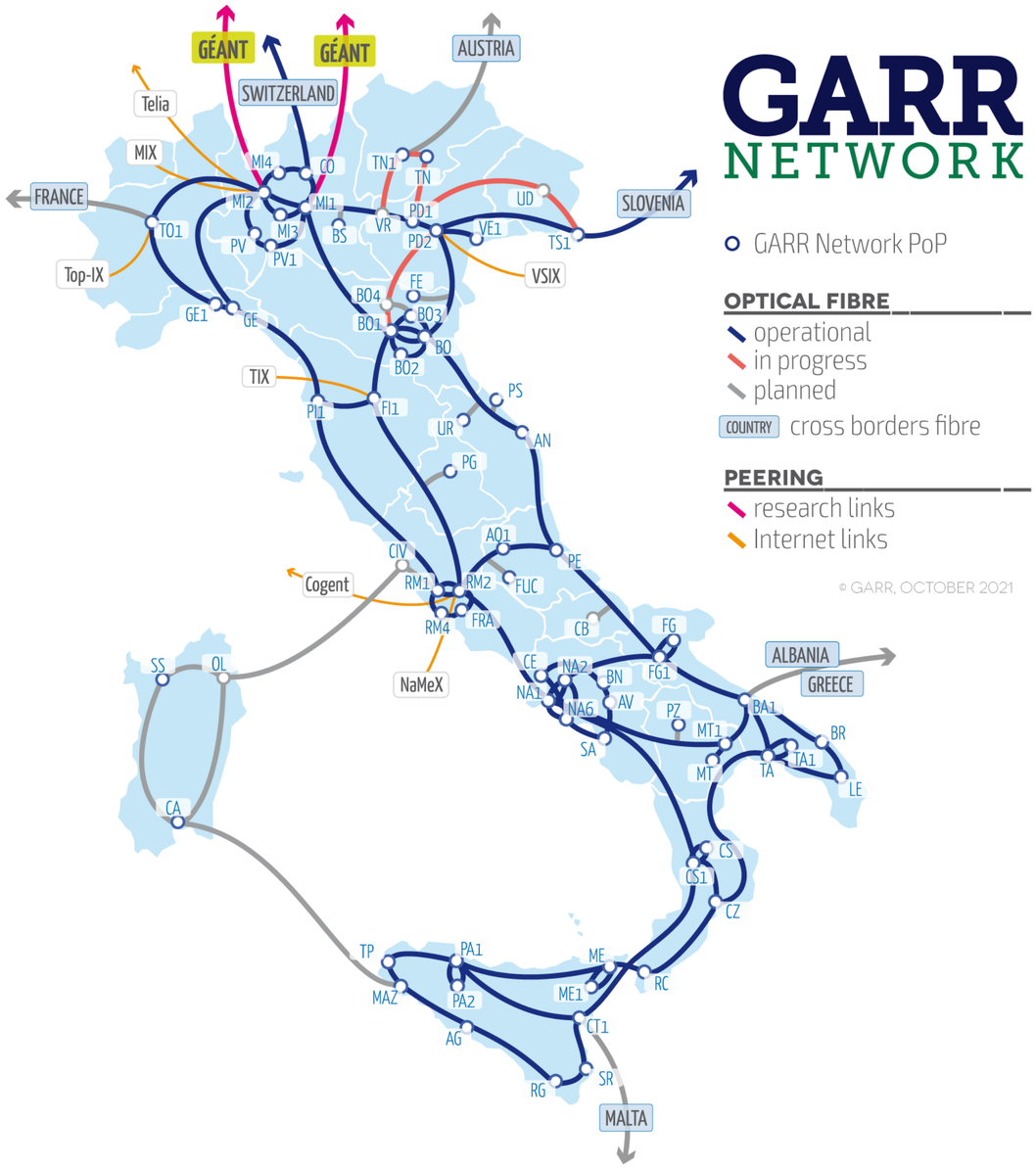


Modifiche infrastrutturali incluse nel progetto GARR-T

- **Infrastruttura fibra ottica:**
  - Richiusura magliatura nord-est
  - Nuovi anelli metropolitani
  - Ottimizzazione dorsali lunga distanza
  - Infrastruttura temporanea per migrazione

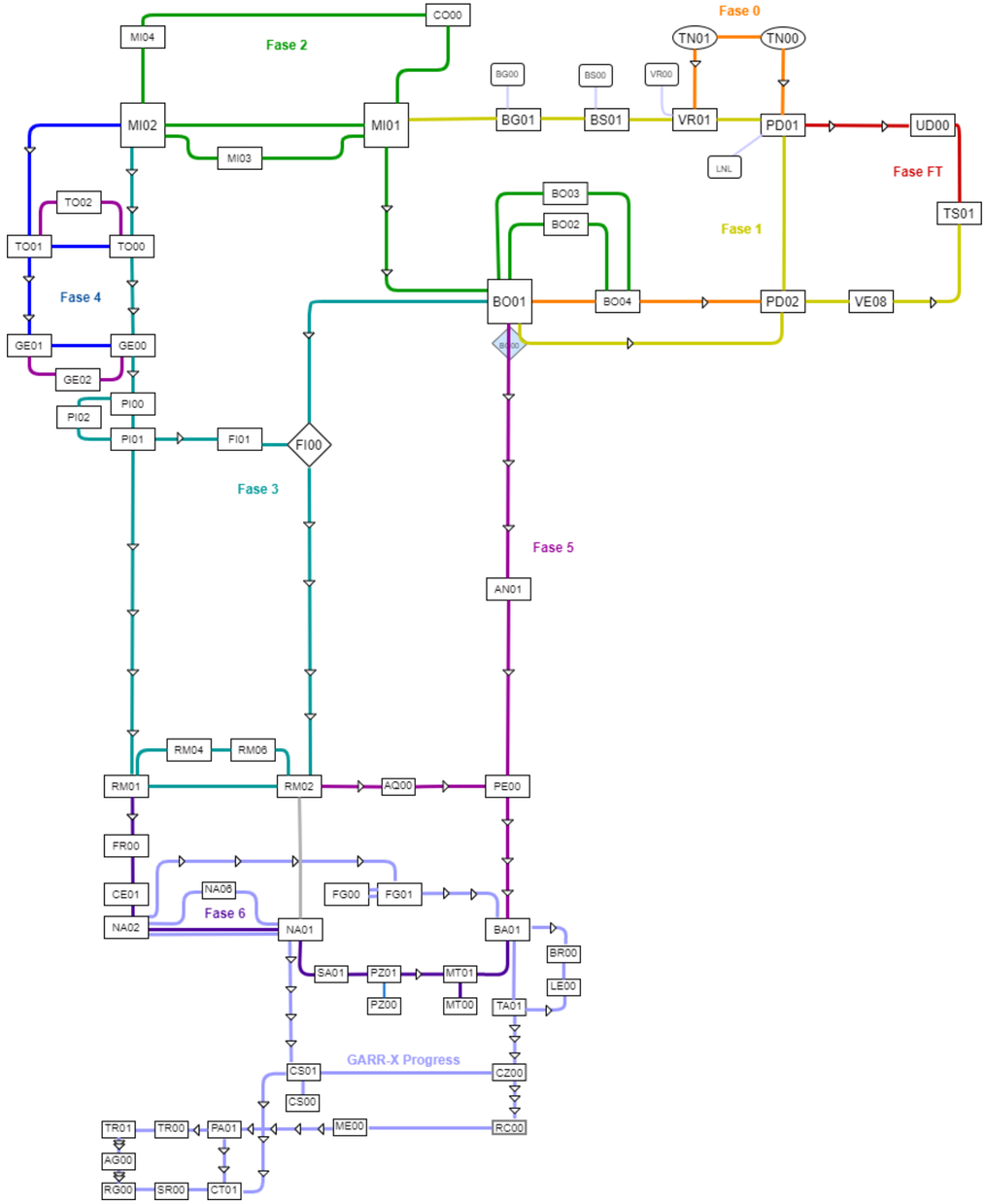
- **Rete trasporto ottica:**
  - Sostituzione piattaforma GARR-X (2011) con nuova piattaforma trasmissiva
  - Rete Trasmissiva GARR-X Progress già basata su piattaforme di nuova generazione

- **Rete a pacchetto:**
  - Completa sostituzione apparati a pacchetto su tutti i Punti di presenza GARR (78)



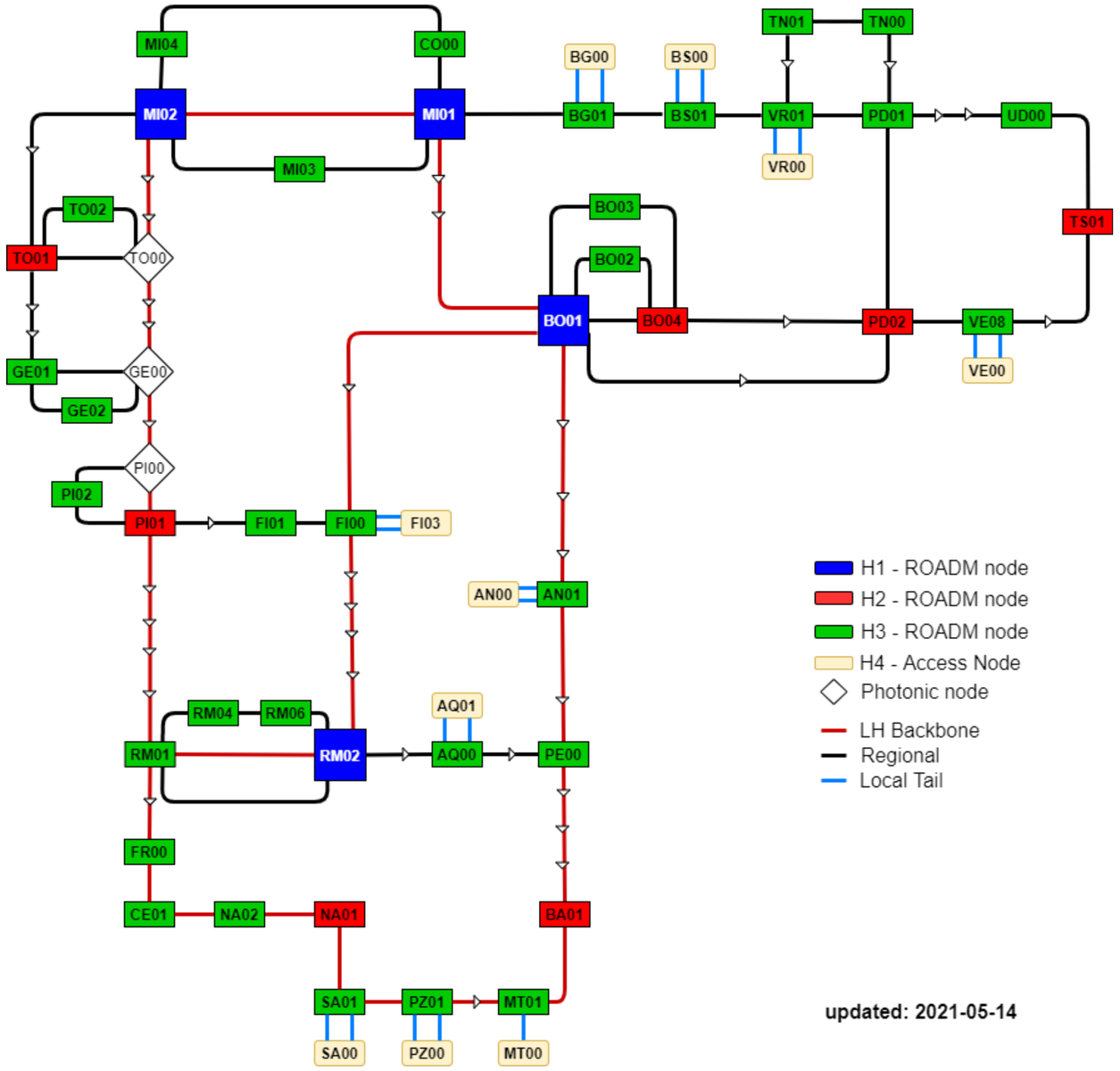
Infrastruttura fisica parzialmente nuova

rete in fibra ottica ri-equipaggiata	6155 km
nuova infrastruttura in fibra ottica	740 km
Rete in fibra ottica garr-x progress	2978 km
Totale rete backbone	9873 km
Fibra temporanea per migrazione	3650 km
Fibra migrazione «hot-swap»	750 km
Nuovi POP metropolitani	9
Raddoppio dei POP in città	6
Nuovi siti amplificazione	5



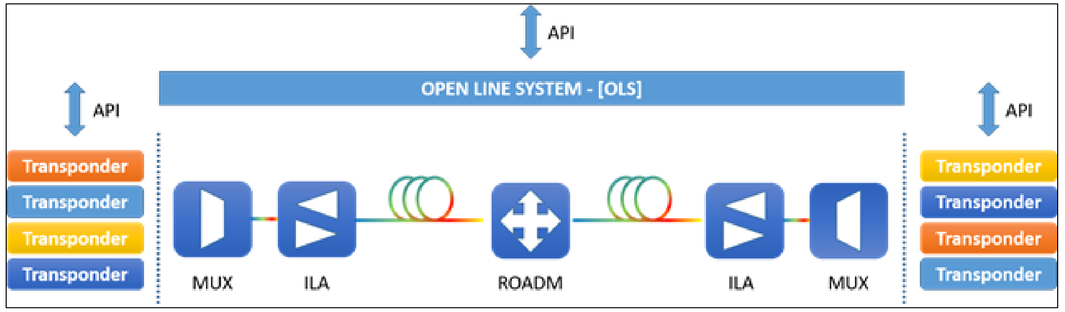
Nuova rete trasmissiva

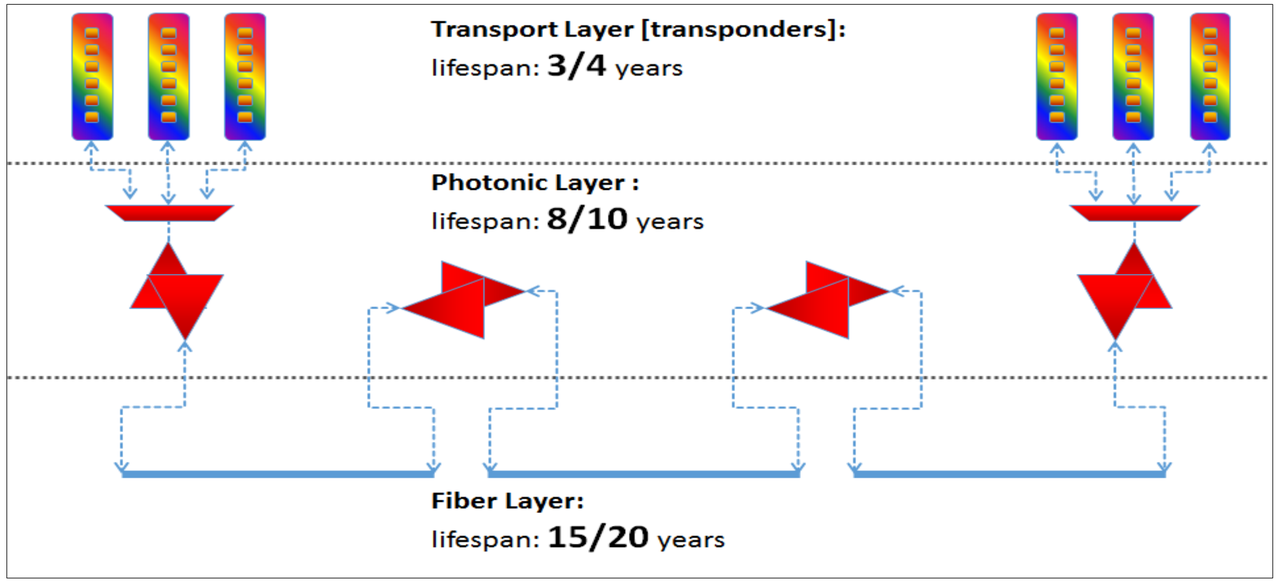
rete in fibra ottica	6155 km
nuova infrastruttura in fibra ottica	740 km
direzioni line system	128
PoP ROADM	42
In Line Amplifier (ILA)	35
Nuovi POP metropolitani	9
Raddoppio dei POP in citta’	6
Servizi 100GEth	130
Servizi 400GEth	11
capacità dorsale day1	17.4 Tbps



Open Line System: rete parzialmente disaggregata

- **Open Line System:** elementi fotonici di linea come amplificatori, mux/demux, ROADMs, OTDR, WSS.
- **Elementi di Rice/Trasmissione (DCI/Transponder):** interfacce di Rice/Trasmissione del segnale ottico
- **Interfacce Programmabili**
- **Elementi di Controllo, Gestione e Monitoraggio**





- Disaccoppiamento tra line system e transponder consente di indirizzare correttamente il ciclo di vita delle soluzioni
- Supporto di soluzioni eterogenee e multivendor

GARR-T apparati rete trasporto ottica

Open Line System: FlexILS Infinera

- Extended C-band (4.8 THz)
- FlexGrid (6.25 GHz slices)
- Alien e Spectrum Sharing by design
- Trasparente a tipologia di modulazione e baudrate
- Nodi adattabili a installazioni DataCenter
- Embedded OTDR

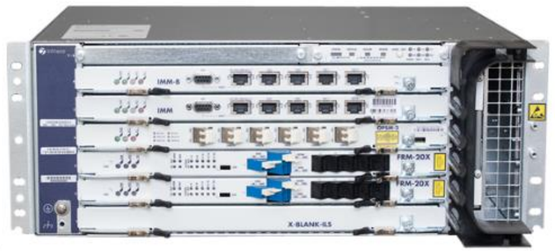
DCI: Groove G30

- 1 RU pizzabox fino a 2.4Tbps per chassis
- Interfacce network tunable
- Modulazione flessibile 100Gbps fino a 600Gbps
- Interfacce client **100GEth e 400GEth**
- Supporto multi-chassis

Infinera

Open Line System

FlexILS



Data Center Interconnect

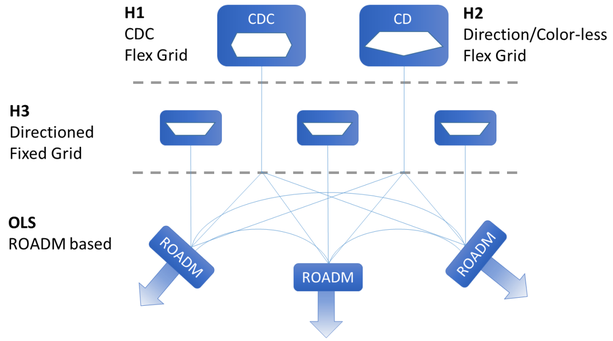
Groove G30

CHM1T

CHM2T

400G

1.2T



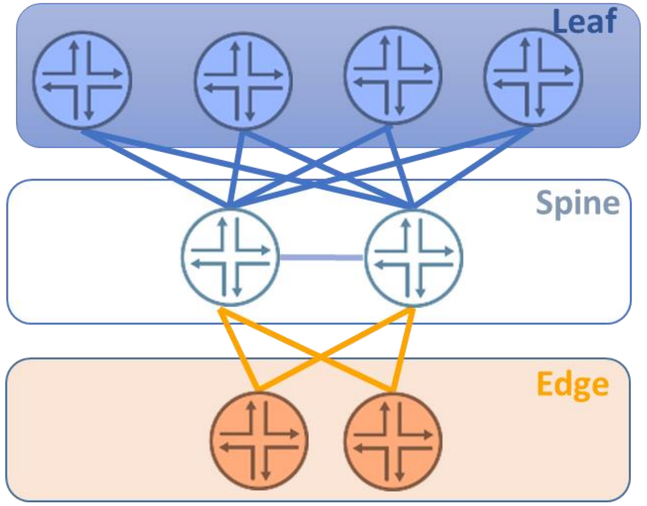
Requisiti della nuova rete a pacchetto

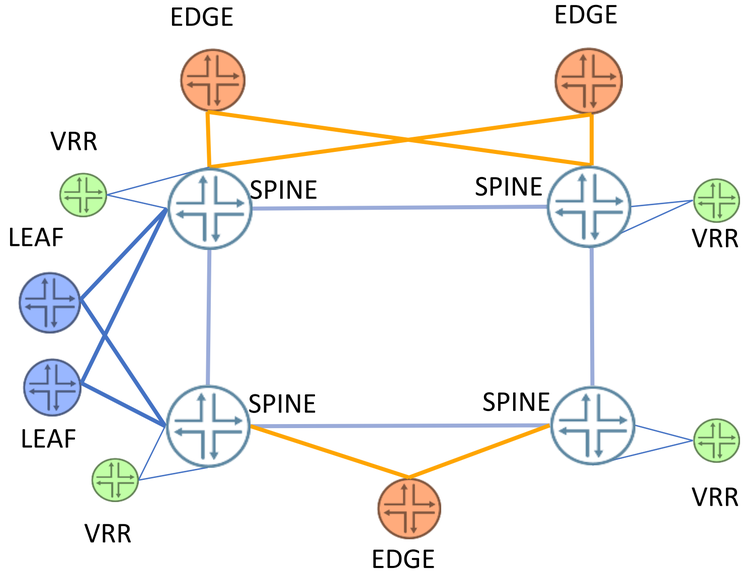
- Requisiti di partenza:
  - riuscire a sfruttare al massimo la capacità disponibile;
  - aumentare la ridondanza, topologia;
  - supportare tutti i servizi L2/L3 attuali e anche nuovi servizi;
  - incrementare la capillarità → MAN/RAN/DC;
  - aggregare da 1G a 100G e anche 400G sul backbone;
  - cercare di ridurre consumi e spazi rispetto alla situazione odierna.
- Necessità di strumenti per:
  - nuove capacità di monitoring (Telemetria);
  - strumenti per la gestione «dinamica» della rete;
  - funzionalità di traffic engineering più evolute.
  - automatizzare:
    - operazioni quotidiane;
    - provisioning dei servizi;
    - upgrade di release.

Rete gerarchica a pacchetto: Spine, Leaf, Edge

- Gli apparati con funzione **SPINE** realizzano lo stadio di trasporto della rete:
  - link a elevata capacità di banda, 400/100 Gbps e multipli;
  - non necessarie funzionalità per raccolta utenti e servizi di edge, per routing internet e transiti.
- Gli apparati con funzione **LEAF** realizzano lo stadio di accesso della rete:
  - accessi utente 100/10/1 Gbps;
  - supportano feature avanzate per i servizi
    - accounting traffico;
    - monitoring (netflow, telemetry);
    - filtering, policing;
    - QoS, VPN, ecc;
- Gli apparati con funzione **EDGE** realizzano lo stadio di interconnessione con le reti esterne:
  - interconnessione con reti esterne, Geant, upstream provider e IXP, capacità 100/10 Gbps, collegati a SPINE differenti;
  - funzionalmente sono identici alle LEAF.
- **vRR**, funzione di Route Reflector trasferita su apparati dedicati.

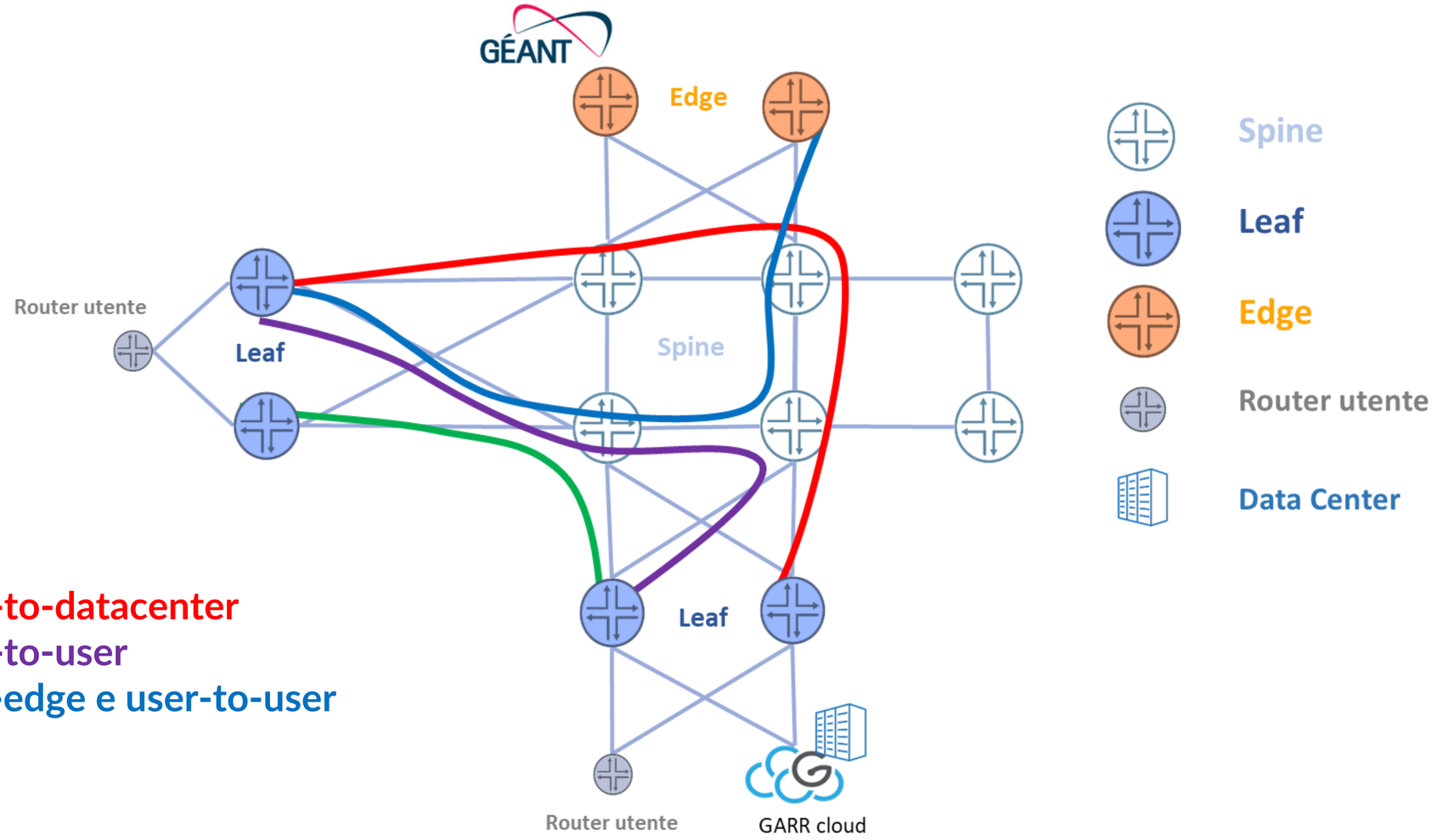
CLOS





Architettura dei servizi

IP/MPLS based



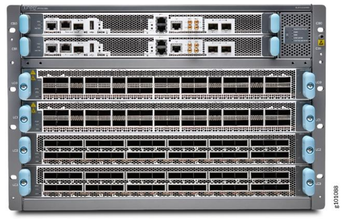
Servizi:

- P2P L2 services user-to-datacenter
- P2P L2 services user-to-user
- L3 IP service user-to-edge e user-to-user
- L3VPN user-to-user

GARR-T Apparati previsti rete a pacchetto

Apparati	Tipo	Spazio (RU)	Consumo (kw)	Interfacce	Performance
MX204	fixed	1	0,3	100GE/40GE/10GE/1GE	400 Gbps
MX10003	fixed	3	2,0	100GE/40GE/10GE	2.4 Tbps
MX480	modulare	8	4,1	100GE/10GE-1GE dual rate	9,0 Tbps (espansione)
PTX10001-36MR	fixed	1	1,8	400GE/100GE/10GE	9.6 Tbps
PTX10004	modulare	7	7,3	400GE/100GE/10GE	9,6 Tbps (espansione)
JRR200	fixed	1	0,5	10GE/1GE	

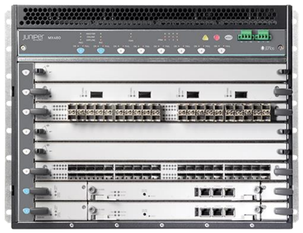
SPINE



PTX10004

PTX10001-36MR

LEAF



MX480

MX10003

MX204

VRR

JRR200

Totale 130 nuovi apparati

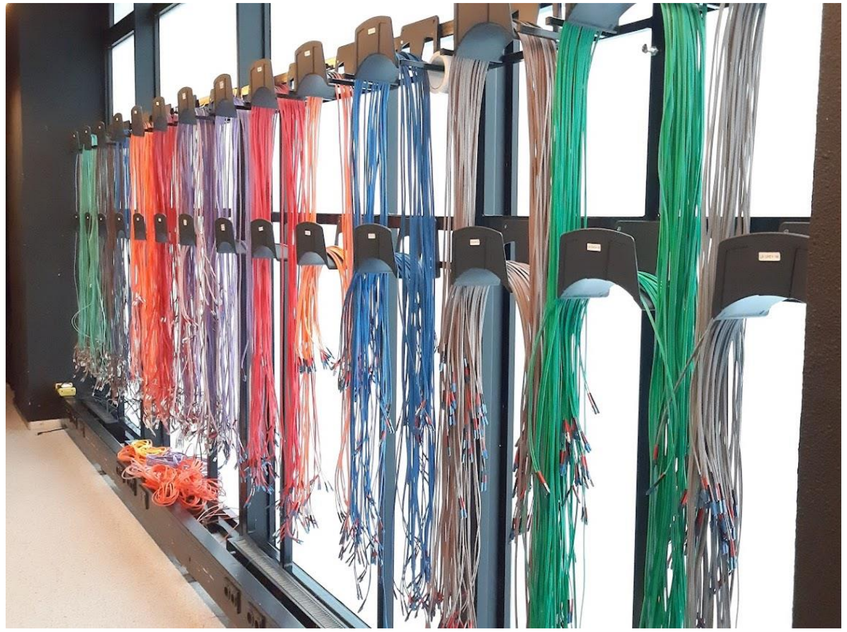
La progettazione di dettaglio della rete IP

Tempi del progetto

- Giugno 2019: avvio dello scouting tecnologico
- Settembre 2019: PoC presso Juniper – Amsterdam
- Gennaio 2020: redazione capitolato di gara
- Dicembre 2020: bando pubblico di gara
- Giugno 2021: contratto di fornitura
- Settembre 2021: emissione ordine
- …
- Chip shortage
- …
- 6 mesi di ritardo consegna apparati (per alcuni 13)

- Necessità di cambio in corsa, per poter iniziare abbiamo temporaneamente usato macchine destinate ad altri pop

- Settembre 2022: avvio migrazione
- Settembre 2023 (?): fine migrazione



Ridondanza e tolleranza al guasto

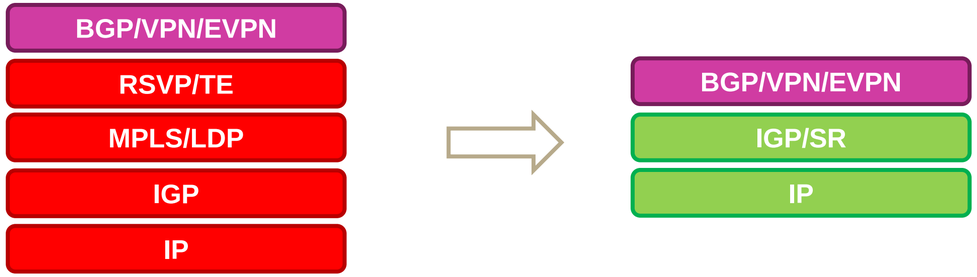
- Da progetto, nessun reinstradamento avviene sul livello trasmissivo
  - Tutte le capacità di reinstradamento sono lasciate al livello IP
  - Ogni PoP foglia è collegato a 2 spine
  - I circuiti uscenti dai pop foglia usano almeno 2 vie fisiche distinte
  - I circuiti sono attestati su leaf distinte, o dove non possibile, su card distinte.

- In totale l'accesso ammonta a **3.9 TB**, backbone oltre **17 TB**
  - La capacità installata sulla rete IP è molto maggiore del necessario, proprio per permettere che il traffico venga reinstradato senza congestione
  - Ad ogni taglio fibra vanno down parecchi circuiti, ma ce ne sono tanti altri che rimangono su
  - I PoP piccoli hanno 2x100G perché non ha senso fare altrimenti
  - MX204 in questi casi ha fatto la differenza: poco spazio, ottime prestazioni, porte 100G

- Funzionalità usate sulla rete IP per aumentare la tolleranza al guasto:
  - LACP (Link Aggregation Control Protocol)
  - Bidirectional Forwarding Detection (BFD) for IS-IS adjacencies
  - Micro-BFD for each child interface for LAGs
  - Topology Indipendent-Loop Free Alternate (TI-LFA)
  - GRES (Graceful Routing Engine Switchover)
  - Non-stop active routing (NSR) for all devices with two Routing Engines (REs)
  - Seamless-BFD (RFC7880) for SR-TE protection

Il nuovo stack protocollare

- Migrazione IGP da OSPFv2/OSPFv3 ad IS-IS
  - ISIS estensibile, intrinsecamente multiprotocollo (ok per dual-stack)
  - Metriche automatiche calcolate in funzione della banda dei link di backbone
- Migrazione label distribution da RSVP/LDP a SR
- Local repair e reinstradamento via TI-LFA
  - Risolve parecchi problemi di LFA e topologie non coperte
  - Reinstradamento fulmineo in ms
- MPLS mantenuto sul forwarding plane per continuare a erogare i servizi esistenti
- Aggiunta control plane EVPN per introdurre nuovi servizi
  - Traffic-engineering realizzato via SR, rimpiazzato RSVP (stateful)
  - Il path computation verrà realizzato tramite un controller esterno e inviato via PCEP ai router



Nuova configurazione del routing

- BGP utilizza 2 route reflector collocati in corrispondenza dei pop centrali della rete.
- Tutti i router ricevono 2 copie della full routing table, per convergenza immediata in caso di variazioni esterne.
- La funzionalità dei route reflector è stata spostata su appliance dedicate, fuori dal forwarding path.
- Questo si è rivelato un grosso vantaggio dal punto di vista operativo: ogni modifica al BGP non si ripercuote sul forwarding, (e ovviamente come prima grazie alla ridondanza, ogni modifica al forwarding si ripercuote solo sul rr adiacente)
- Routing multicast attivo su tutta la rete, SSM intra ed interdominio, ASM solo intradominio.
- Forwarding multicast tutto via NG-MVPN
- BGP VPN route target non propaga route VPN routes verso I PE non membri (RFC 4684)

Lista completa dei BGP NLRI supportati:

1/1:	family inet (with add-path)
1/2:	family inet multicast
1/4:	family inet label unicast
1/5:	family inet-mvpn
1/128:	family inet-vpn
1/132:	family route-target
1/133:	family inet flow
2/1:	family inet6 (with add-path)
2/2:	family inet6 multicast
2/5:	family inet6-mvpn
2/128:	family inet6-vpn
2/133	family inet6 flow
25/70:	family evpn signaling

Nuove modalità di gestione

Sensori telemetrici nativi o attivati dinamicamente via GNMI

- Juniper ha un certo numero di sensori che può inviare dati in telemetria direttamente dalle card, senza impegnare cicli CPU
- Altri sensori che impegnano la CPU possono essere attivati e disattivati dinamicamente

Real-time performance monitoring (RPM) – delay, jitter, loss, dati raccolti via telemetria con risoluzione 30 secondi

Paragon Insights - piattaforma di gestione per allarmi e performance

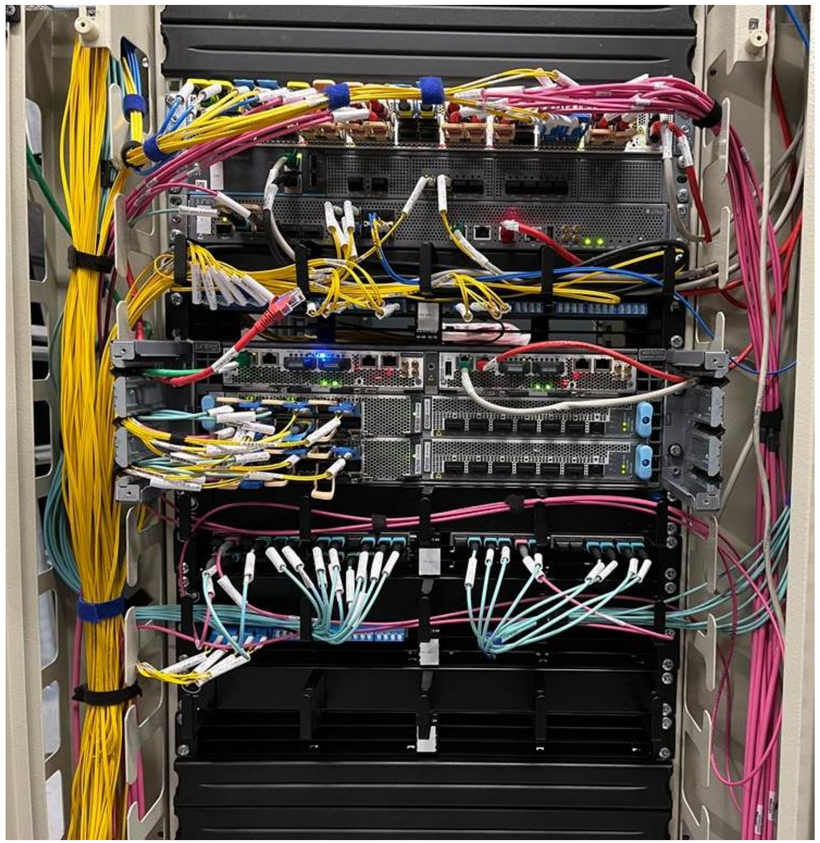
- raccoglie informazioni dagli apparati via telemetria, Openconfig, netconf/SSH, NetFlow, SNMP and syslog
- Analizza su base soglie anche adattive (anomaly detection via machine learning)
- Oltre alla GUI esporta via API/webhook/slack

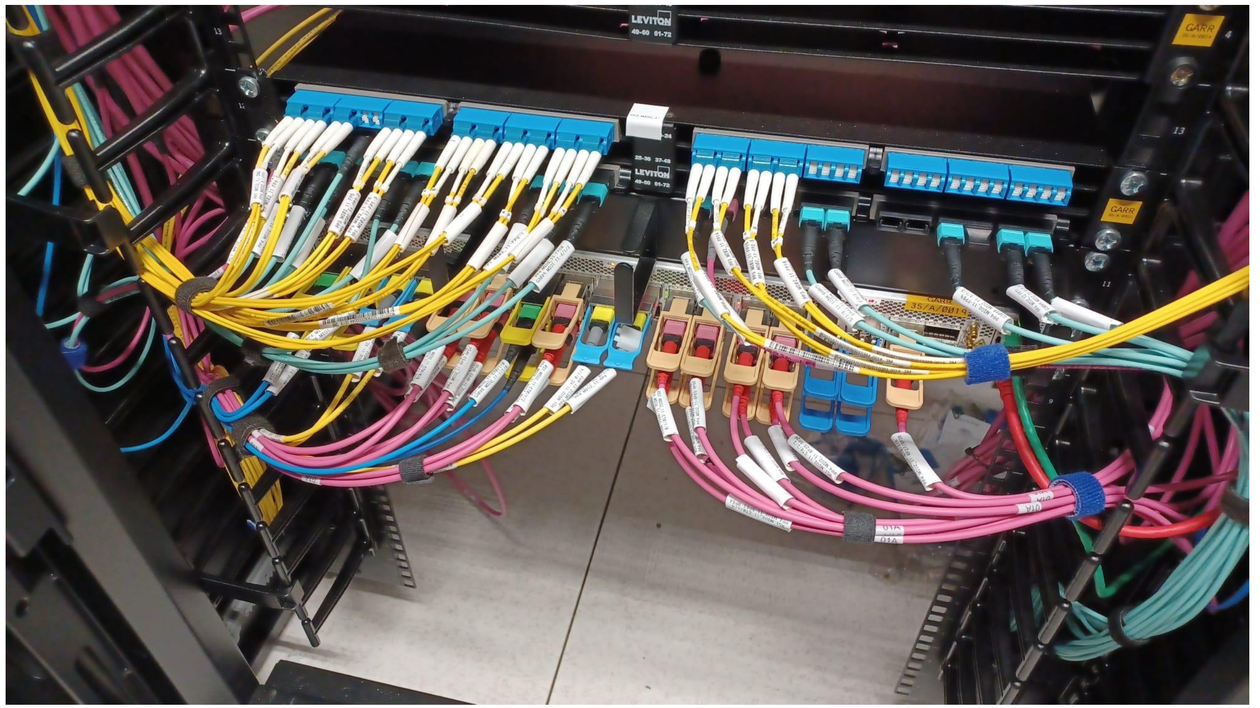
Paragon Pathfinder – piattaforma per il traffic engineering

- Riceve aggiornamenti sulla topologia in tempo reale via BGP-LS
- Collegato a tutti i PE della rete via PCEP, può configurare path di traffic engineering o prendere in carico quelli già configurati in rete
- Stiamo studiando come fare ottimizzazioni in base al traffico, delay e anche riparare packet loss

La realizzazione

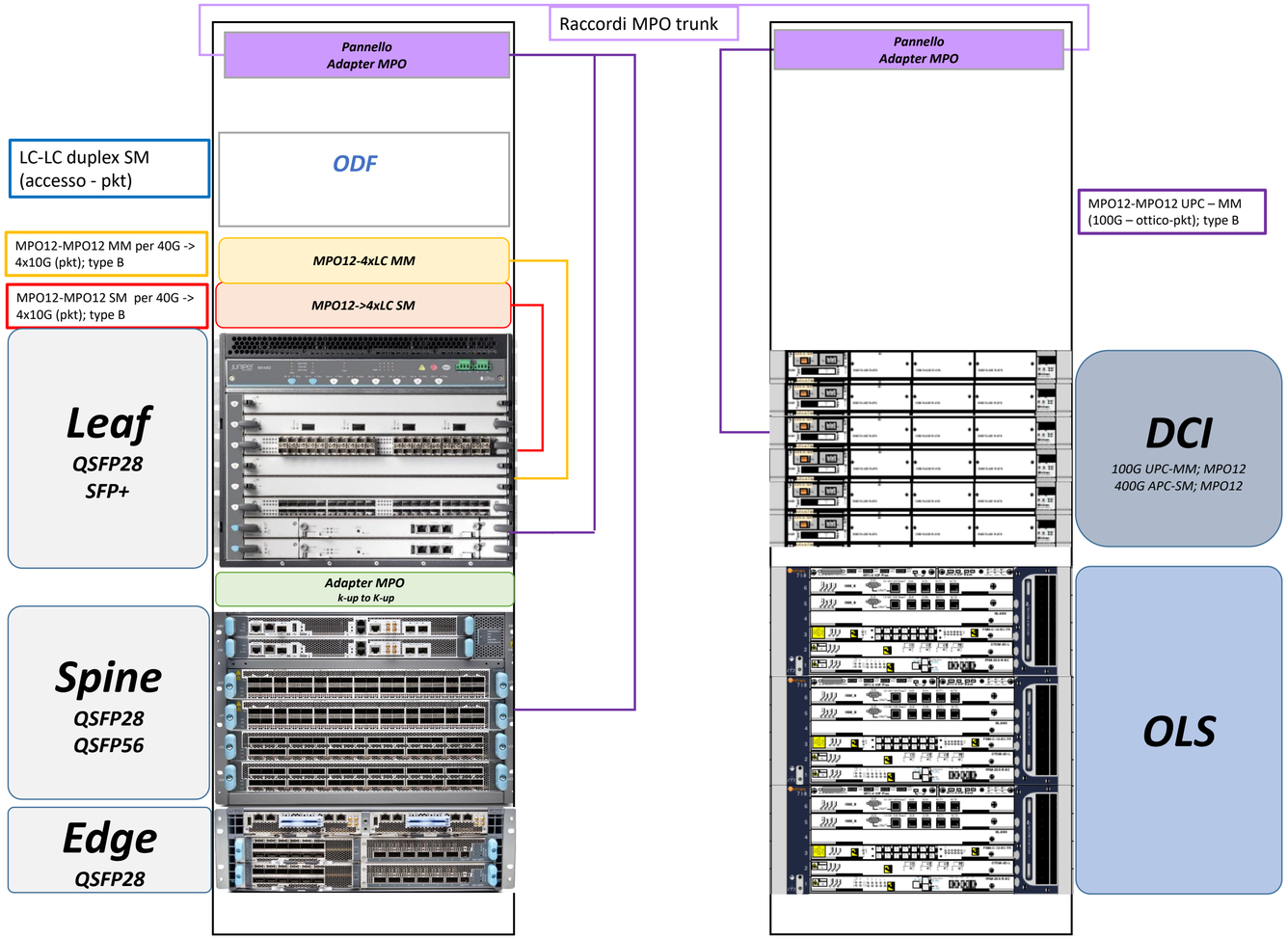
Nuovo cabling, nuovi connettori





Cablaggi 100G intranodo e verso trasmissivo tutti SR4
Cablaggi 400G intranodo e verso trasmissivo tutti DR4
Porte utente, PNI, IXP 100G tutte LR4
Porte utente 10G/1G a seconda delle esigenze (4x10G sia LR che SR)

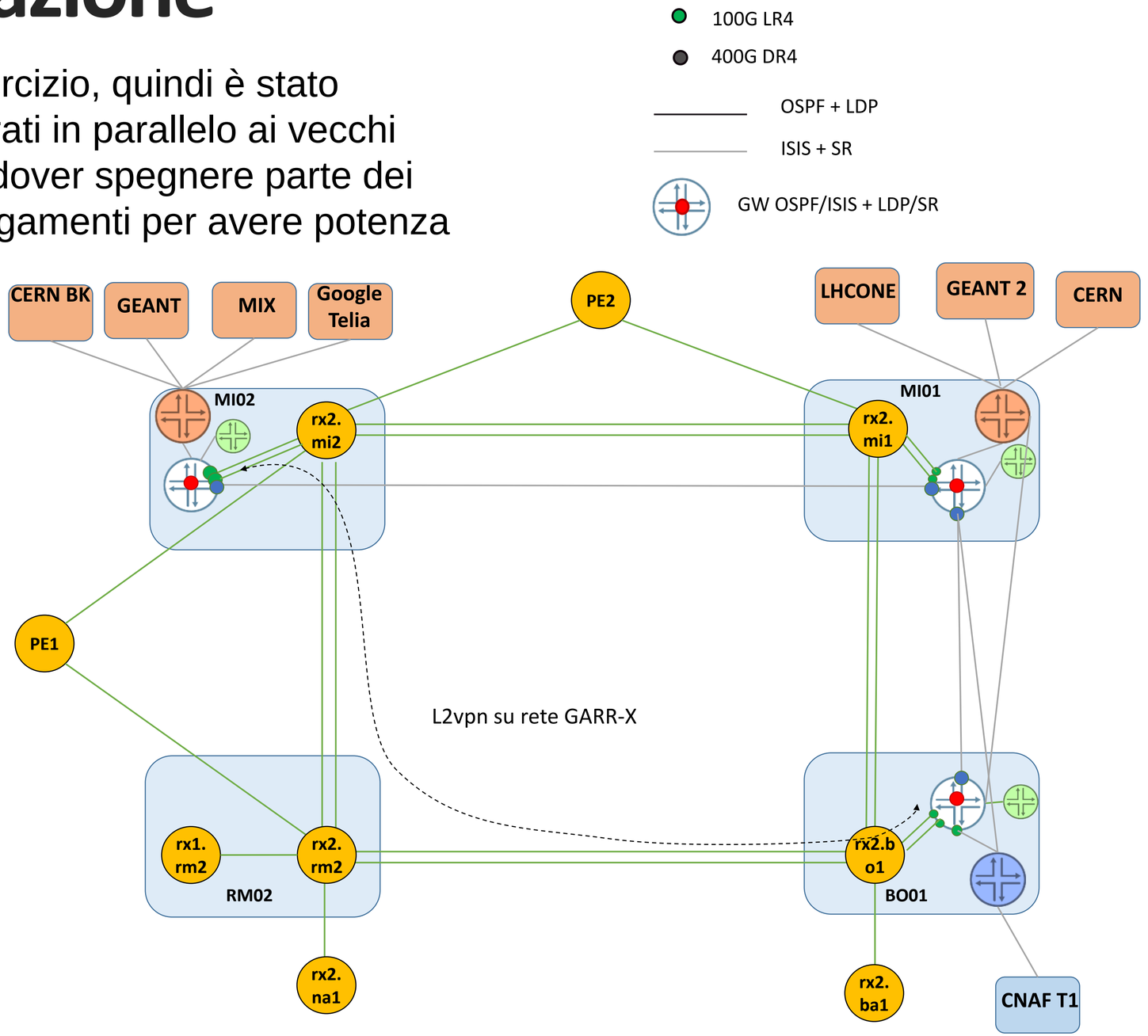
Rack layout (PoP medio-grande)



Il processo di migrazione

- La migrazione avviene a rete in esercizio, quindi è stato necessario accendere i nuovi apparati in parallelo ai vecchi
- In alcuni casi questo ha significato dover spegnere parte dei vecchi apparati compattando i collegamenti per avere potenza sufficiente

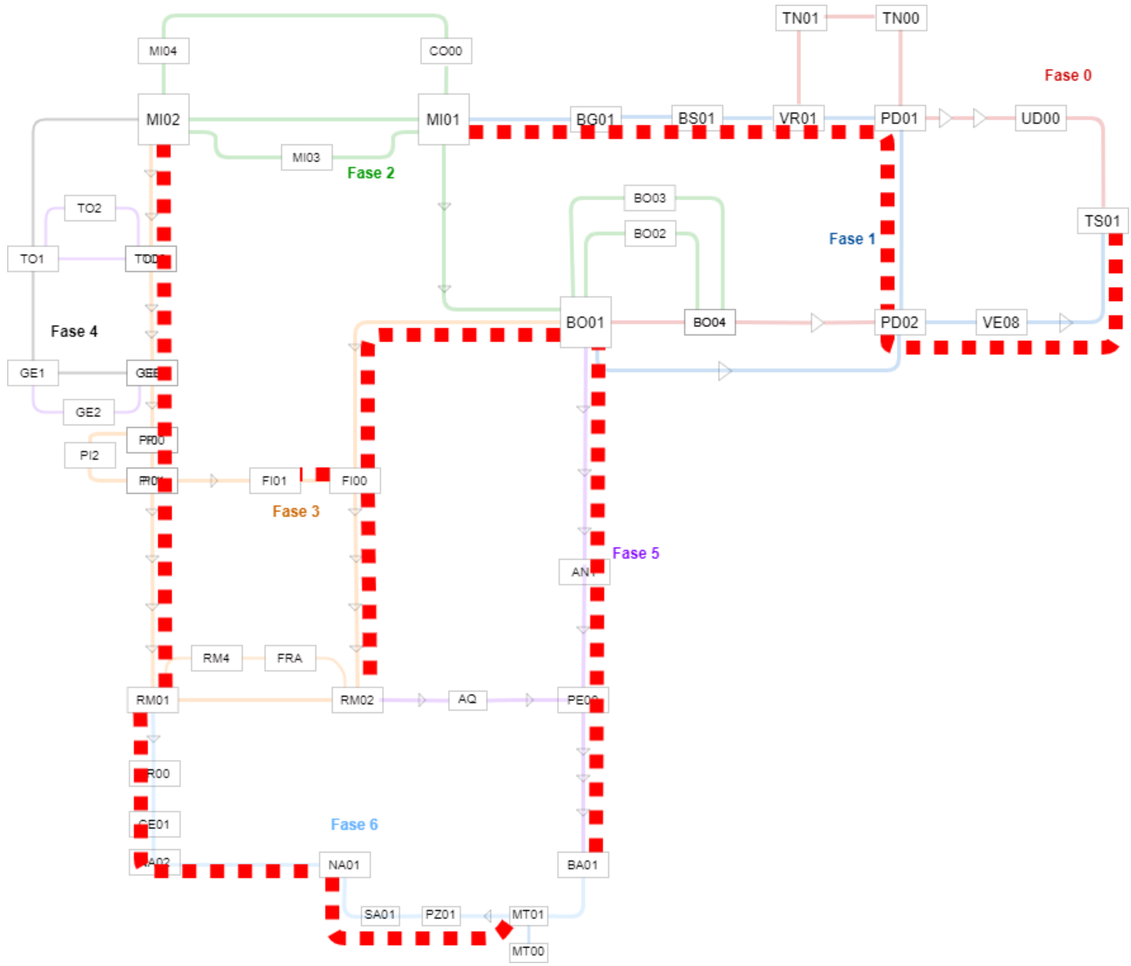
- Le regole imposte dalla sovrapposizione degli stack protocollari hanno richiesto la creazione di alcuni gateway di conversione e redistribuzione dei protocolli
- I gateway garantiscono anche la comunicazione fra vecchia e nuova rete
- Ove necessario sono stati anche realizzati collegamenti virtuali tramite L2VPN per richiudere topologie non ancora disponibili



2 migrazioni contemporanee

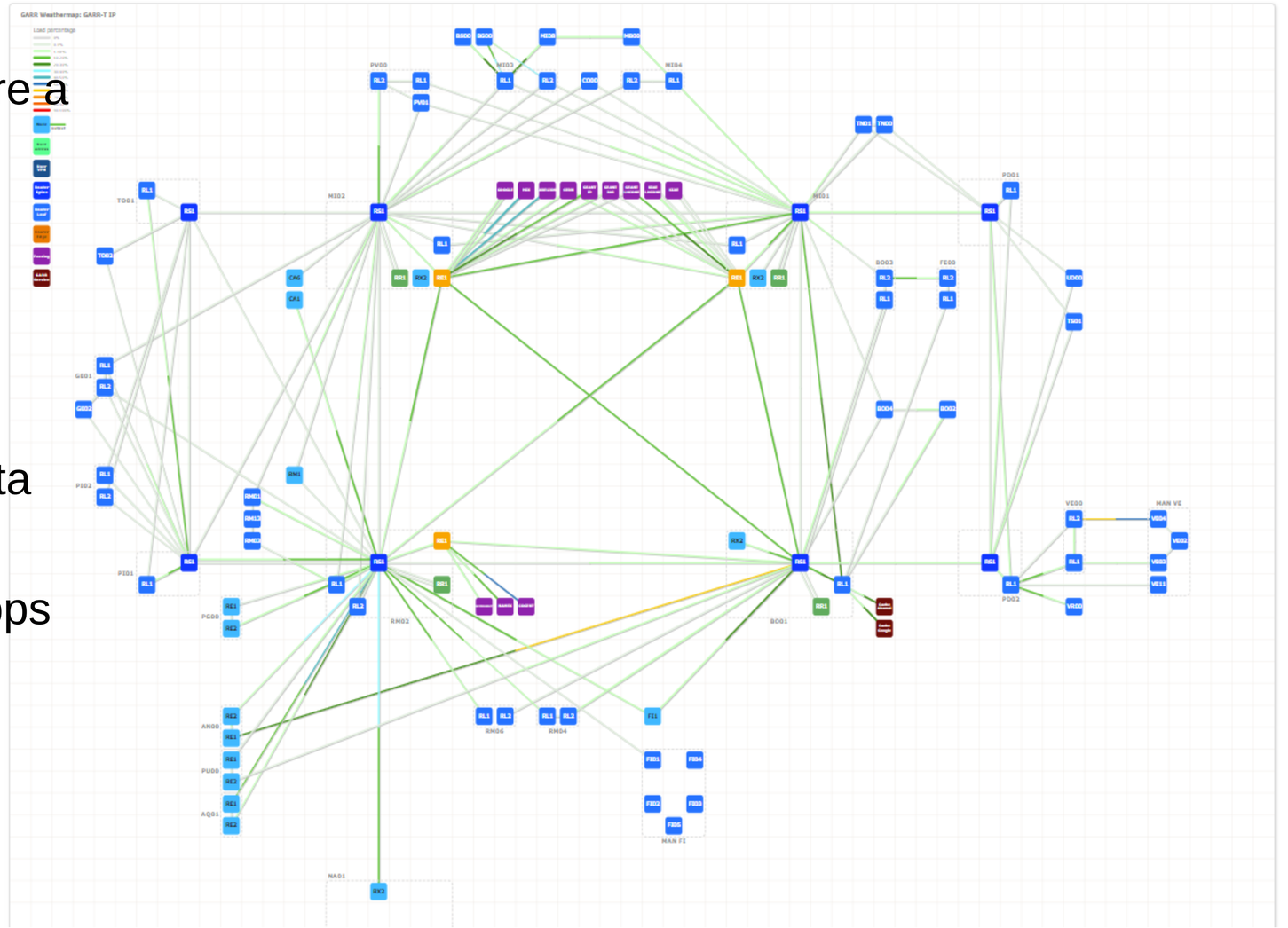
- Nel frattempo avviene anche la migrazione della rete trasmissiva
- La nuova rete trasmissiva prevede tre tipologie di tratte:
  - Nuova fibra (caso facile)
  - In alcuni casi una coppia di fibre temporanea e sovrapposta (a disposizione per circa 1 anno).
  - In altri casi dove non era prevista la seconda coppia di fibre la migrazione è avvenuta con un'attività di hot-swap
    - Tratte con 4 siti da migrare in contemporanea
    - Necessario anche tagliare e risaldare connettori

Fibra in sovrapposizione 3650 km
Migrazione hot-swap 750 km
Nuova infrastruttura 740 km



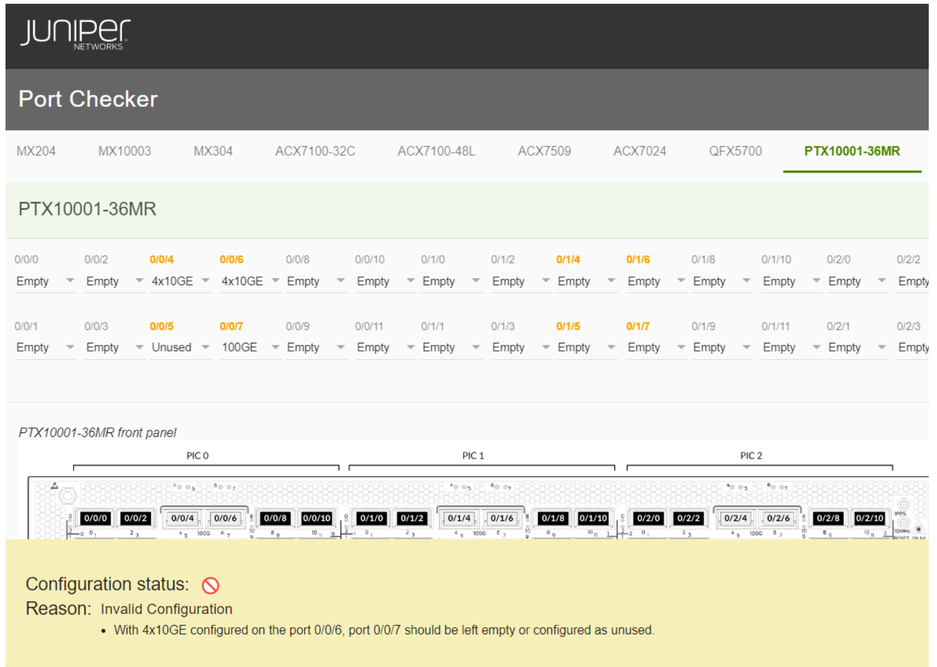
Stato della migrazione

- Inizio: settembre 2022
  - attivazione del nuovo super-core a **800G MI1-MI2-RM2-BO1**
  - spostamento transiti, peering e interconnessioni esterne
  - Attivazione gateway GARR-T/GARR-X **(4x200G)**
- Nei mesi successivi utenza migrata su 45 di 78 PoP
- Dorsale attivata 12 Tbps su 17 Tbps totali
- Nuova fibra attivata **4.500km su 5.200km**
- Attività ancora in corso, termine previsto autunno 2023



Lessons learnt

- Contrariamente al passato, i nuovi apparati Juniper richiedono una accurata pianificazione delle per l'utilizzo futuro – non tutte le porte sono utilizzabili in modo indifferente (port checker)
- Inizialmente pensavamo che gli apparati monolitici (PTX10003, MX10003, MX204) fossero più adatti e snelli, ma per le maggiori densità i modulari (PTX10004, MX480) sono ancora imbattibili
- Le tipologie di porte sono cresciute rispetto al passato, ma questo permette anche molta più flessibilità (es. porte dual rate)
- Le ottiche 4x10G possono essere usate anche in modalità mista (speed 1G/10G)… a patto che usi la release giusta..
- Avere un laboratorio su cui testare hardware e procedure è ancora più necessario che in passato



News aggiornate su:
https://www.garr.it/it/infrastrutture/rete-nazionale/rete-garr-t

marco.marletta@garr.it