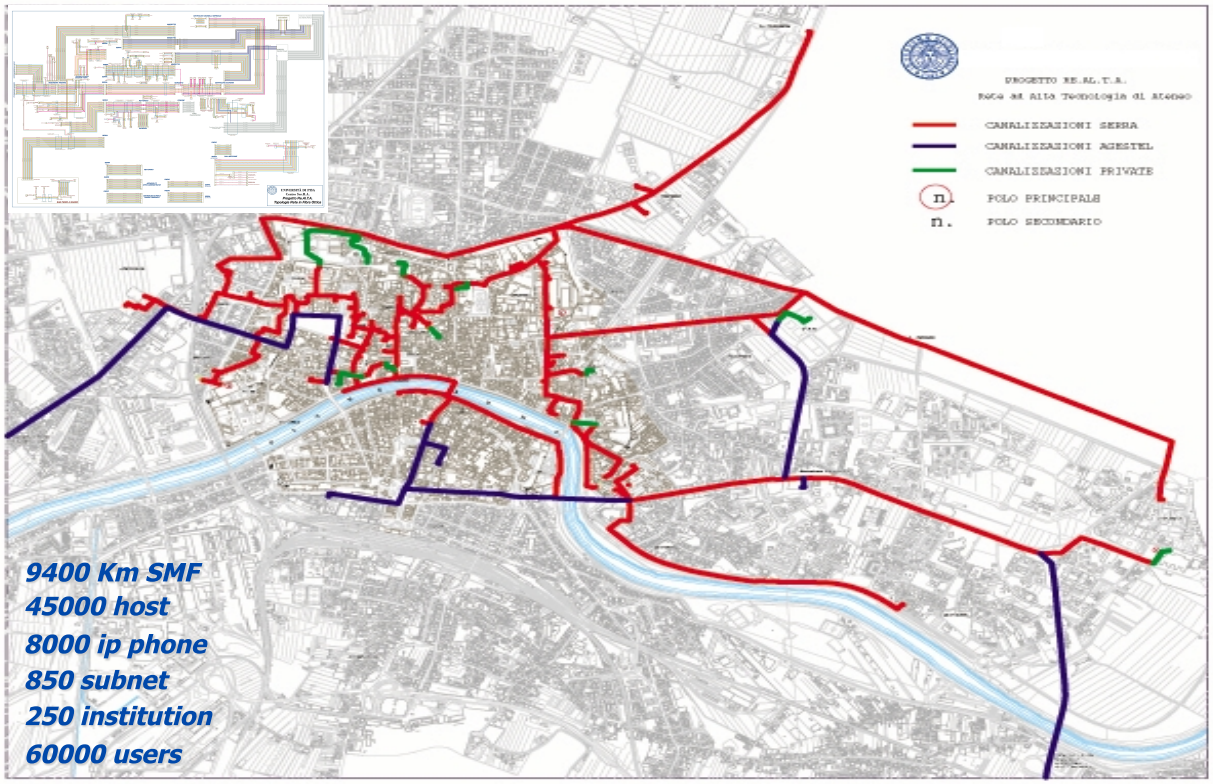
Integrazione di DataCenter nelle reti metropolitane IP/MPLS l'esperienza dell'Ateneo di Pisa

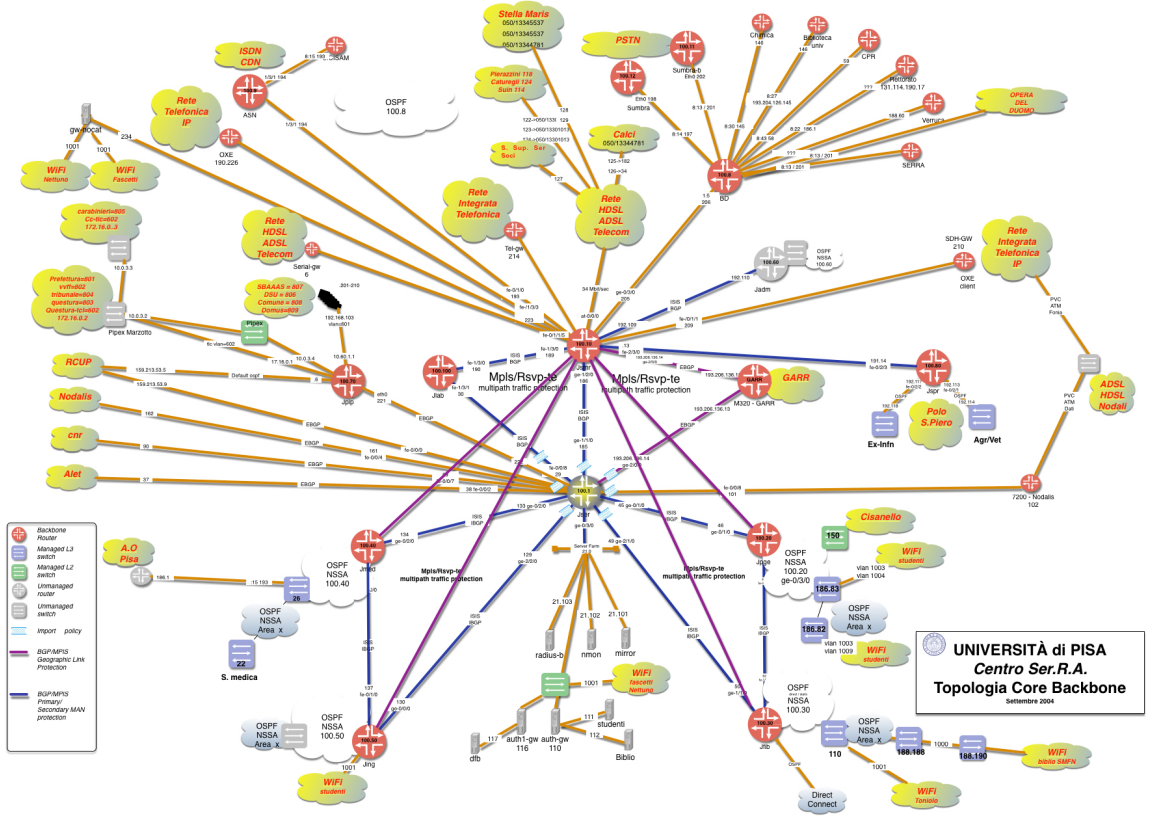
Paolo Caturegli
Rossella Caputo
Antonio Cisternino
Stefano Suin

UNIVERSITÀ DI PISA

Unipisa Backbone - 2000



2004:Il primo modello di trasporto per network providing centralizzazione del routing, trasporto MPLS



2006 - L'adattamento alle esigenze con la virtualizzazione compulsiva

- La necessità di aumentare il grado si scalabilità della rete per la proliferazione delle diverse comunità e dei diversi servizi si è declinata i una accelerazione del processo di virtualizzazione VLL, L2-vpn (VPLS) e L3-vrf,
  - Necessità di isolare risorse e comunità (reti **identity-driven**)
  - Necessità di **provisioning** Zero Touch e manutenzione più semplice in un ambiente complesso
  - Necessità di agevolare lo stretching dei broadcast domain per alcuni protocolli e servizi
  - Reti **multiservizio x comunità** (dati, voce, video, mission critical appl, intranet/extranet)
  - Performance, **traffic enginnering e QoS** (CoS) controllabile, predicibile e verificabile per il controllo degli SLA x comunità (policing e filtering x vpls)
  - **Sicurezza**: complessità = vulnerabilità

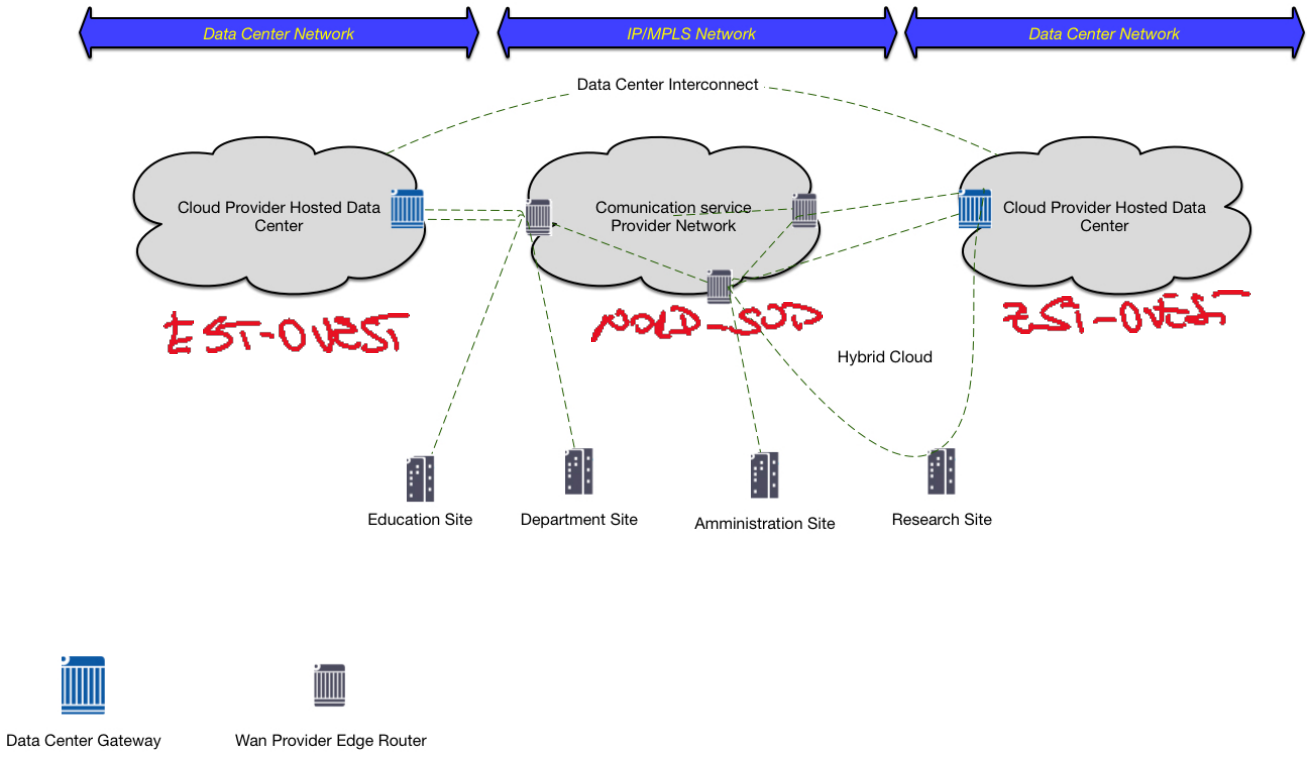
I Datacenter come punti di erogazione dei servizi IT

- **L'adozione delle soluzioni di cloud e di virtualizzazione sulle quali si basa lo sviluppo e l'erogazione dei servizi It è un processo irreversibile**, in una progressione che non è soggetta ad alcune inversione percepibile di tendenza ha comportato la realizzazione di datacenter per l'erogazione di servizi IT
    - 174 dc di varia taglia nella rete di ateneo
- A un certo punto è stato necessario razionalizzare a pochi DC con caratteristiche adeguate alla criticità e ai nuovi ruoli assunti dalle infrastrutture IT e dal cooperative working
- ..in ogni caso ….la normativa non lascia spazi: piano triennale per l'informatica nella PA 2017-2019, investimenti nel settore ICT-PA, Circolare n. 2 del 24 giugno 2016 -> razionalizzazione dei DC da 174 a 3.

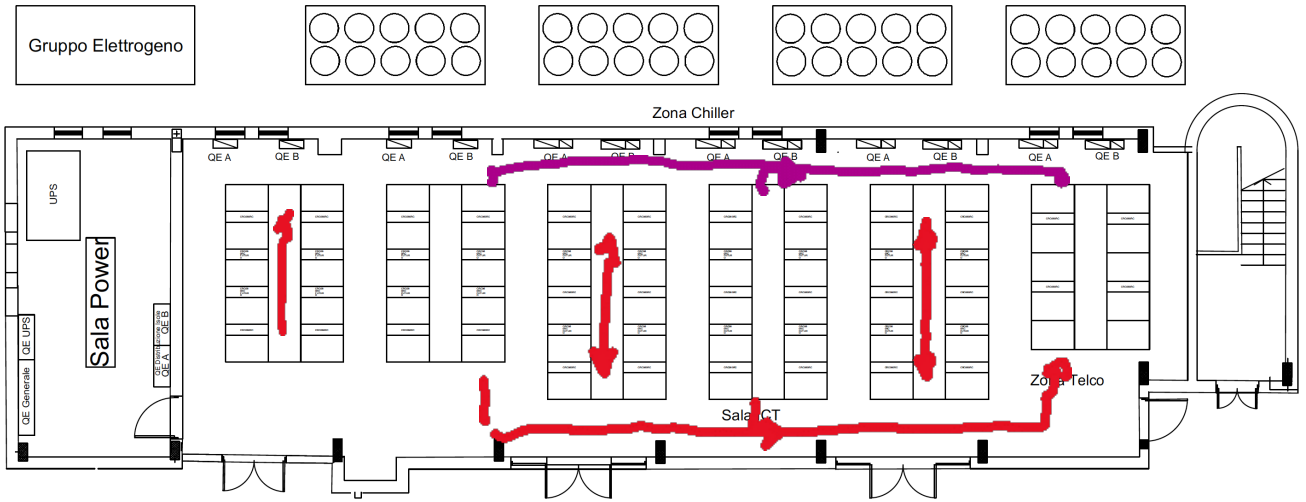
Come cambiano i Datacenter
Le sfide di integrazione
Cloud Providing vs Network providing

- Le tecnologie di rete utilizzate nelle reti metropolitane IP/MPLS e nei data Center sono tipicamente differenti –
  - i Dc rappresentano un elemento di discontinuità nella gestione delle reti (M.Carboni, 2018)
  - 1 soluzione inserire i dc come semplici foglie della rete: La centralizzazione del routing può portare ad effetti molto negativi sulla rete (es. Tromboning :first hop router distante dall'origine, stress dei link DCI), latenze impredicibili, congestione di link):
  - 2 soluzione, rivedere i modelli di trasporto differenziando le esigenze
    - Incremento esponenziale della necessità di **gestire Traffico est-ovest vs traffico nord-sud**
    - Impatti sul cablaggio, che, nella fattispecie, non può essere solo in un modo o nell'altro
    - Le reti ip/mpls utilizzano pesantemente underlay network basate su tecnologie l2-vpn e l3-vpn, mentre le reti DC sono sostanzialmente basate su trasporti L2 nativi, vlan o **vxlan il data-plane e il control plane devono cooperare** con soluzioni di continuità
  - **La scalabilità e la flessibilità che gli ambienti virtuali richiedono sono ben maggiori di quanto le architetture di reti tradizionali possano offrire.**

Integrazione nel backbone

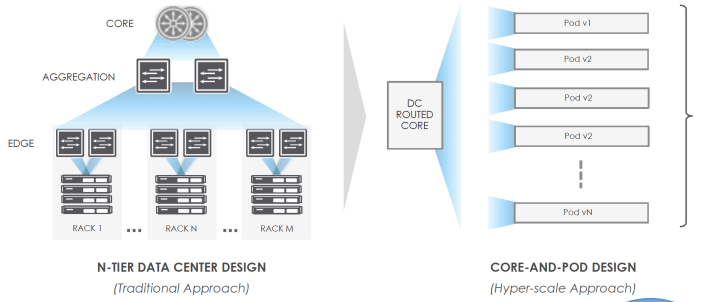


Impatti sul cablaggio.. Facile fare i modelli di rete sulla carta… poi il caturegli mi si imbufalisce.



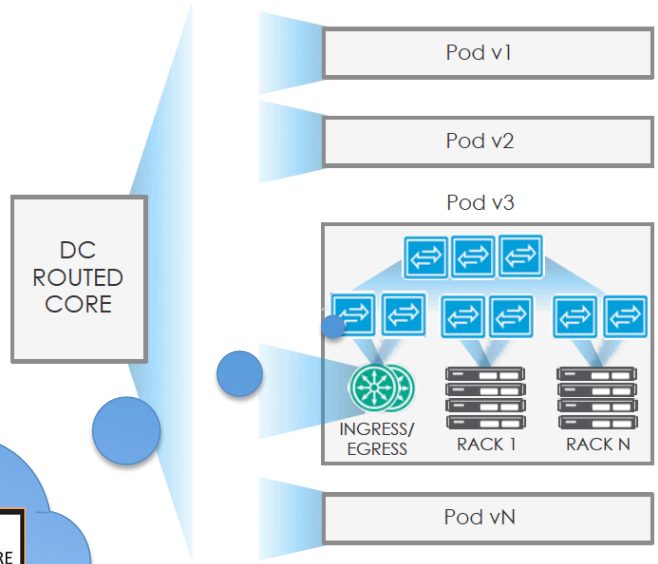
Impatto nel datacenter

Garantire scalabilità e modularità anche in una ottica multiDC, e affrontare ilproblema dell'oversubscription (tollerabile nelle reti network providing): il modello core and pod (computing, network and storage)



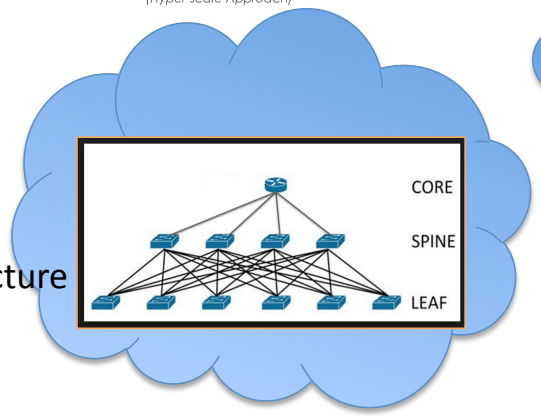
N-TIER DATA CENTER DESIGN
(Traditional Approach)

CORE-AND-POD DESIGN
(Hyper-scale Approach)



CORE-AND-POD DESIGN
(Hyper-scale Approach)

Spine and leaf architecture



- Il cuore dello spine and leaf: MC-LAG (Multi chassis Link aggregation group)

Nessun loop logico fra gli edge devices e i multi-chassis peering switch anche se un loop fisico

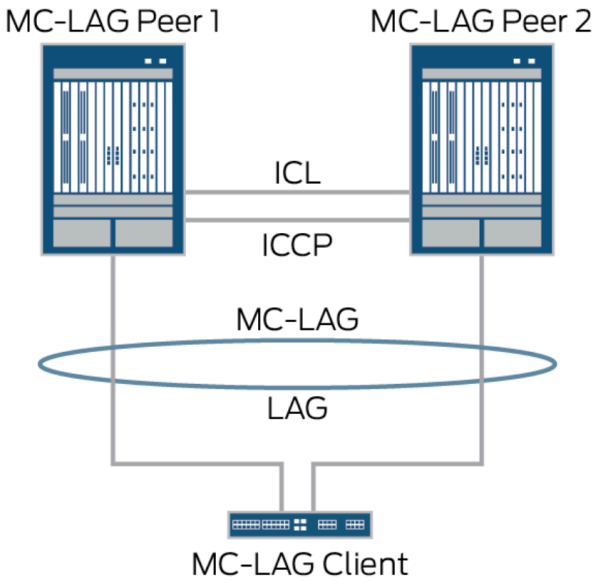
Ridondanza link **Active/Active** e load balancing – reti in non blocking state

Loop-free l2 network **senza STP, no blocking port**

Connessione con protocolli standard Static LAG o 802.3ad LACP

Single interface servers, storage e edge devices possono essere connessi in singola interfaccia ad ognuno degli MCLAG switch.

Se la risilienza di nodo non è fondamentale connessioni dual-homed possono essere stabiliti con solo un MC-LAG edge device



- L’implementazione a Pisa

Netowork POD : MC-LAG MESH

2 gruppi MCLAG connessi back-to-back in un mesh.

Il MESH è limitato a due gruppi .

Link active/active o single active

Resilienza complete di link e di nodo

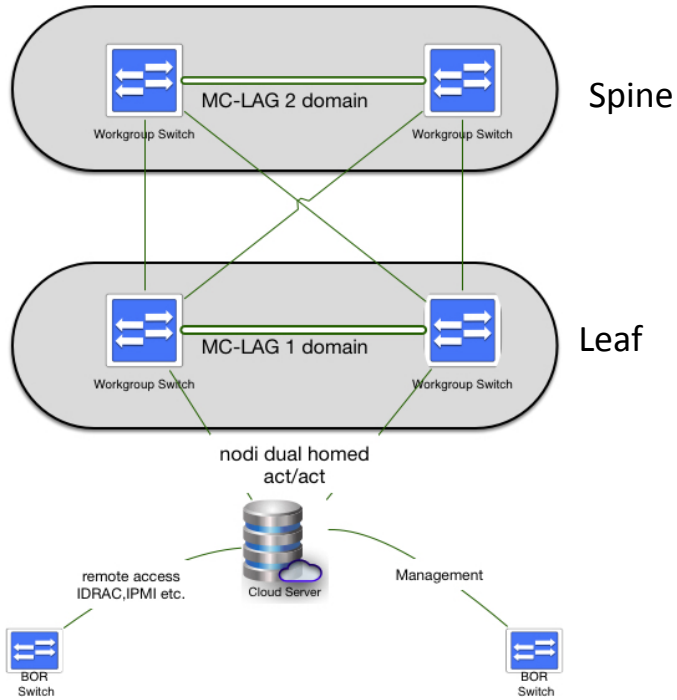
No SPT

Buona soluzione alll’Oversubscription (modulare… MC-LAG dedicati per funzione(Computing (1:4), Storage (1:1), Management 1:100), Application..etc)

Semplice inserimento di nodi dual-homed in LACP o in interface bonding

svantaggio

Il costo. Ancora abbordarabile il leaf a 10/25 e lo spine a 100



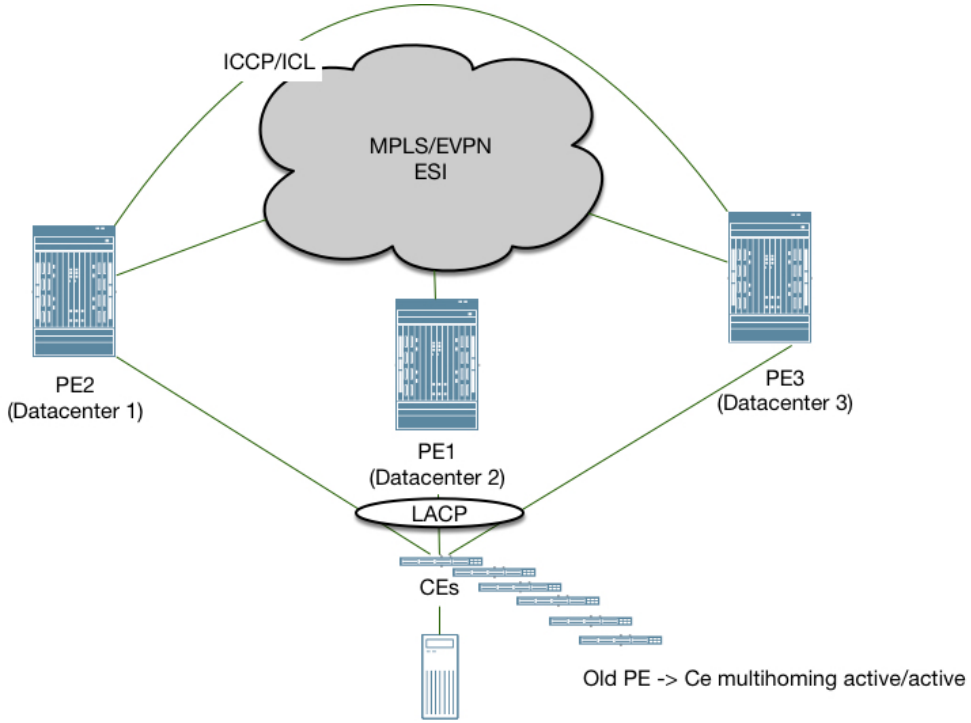
Le scelte di trasporto del control plane e del data plane

- Evpn – Control plane
  - main application of EVPN is Data Center Interconnect (DCI), which provides the ability to extend Layer 2 connectivity between different data centers
  - Tecnologicamente una rete virtuale multipunto che offrono connettività in bridged mode fra domini L2 differenti su un backbone IP or IP/MPLS
  - Control plane flood mac learning
    - » Active multihomed edge device
    - » Fast convergence
    - » Load balancing across active links
    - » Easy MAC address mobility
    - » Multitenancy
  - Ma quale data plane utilizzare?

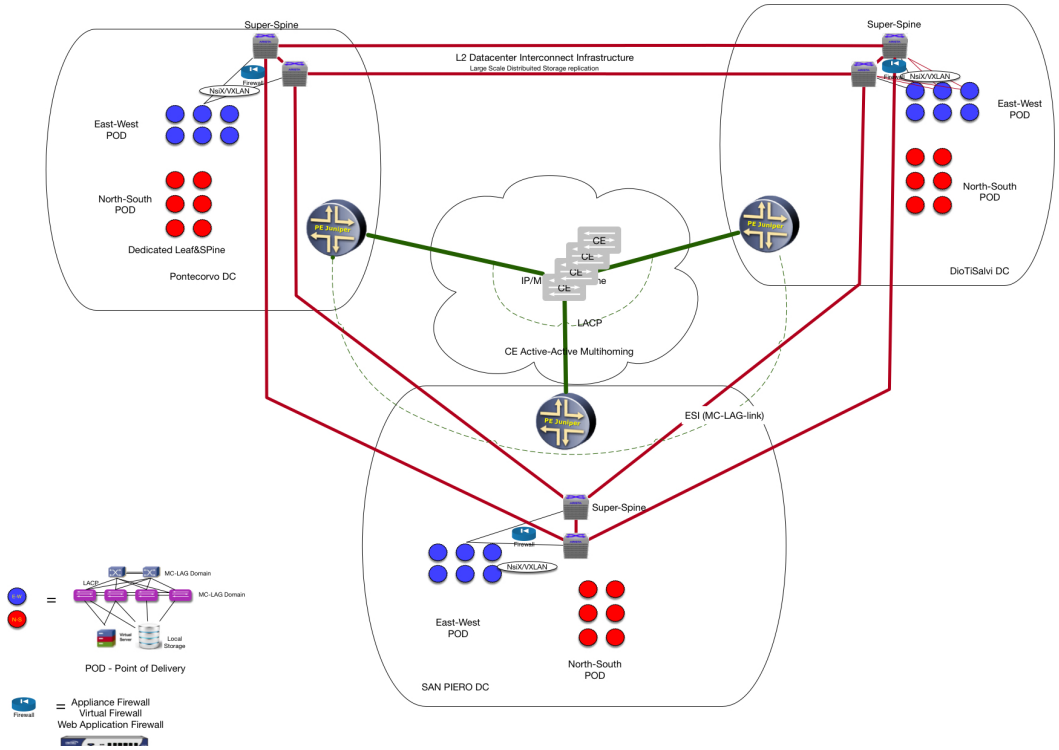
Le scelte di trasporto del control plane e del data plane

- VxLan – il data plane degli hypervisor
  - VXLAN tunneling protocol encapsulates Layer 2 Ethernet frames in Layer 3 UDP packets. This encapsulation enables you to create virtual Layer 2 subnets or segments that can span physical Layer 3 networks.
    - No stp
    - No blocking port
    - No brosdcast storm
    - Load balancing ip
    - 24 bit vs 12 bit

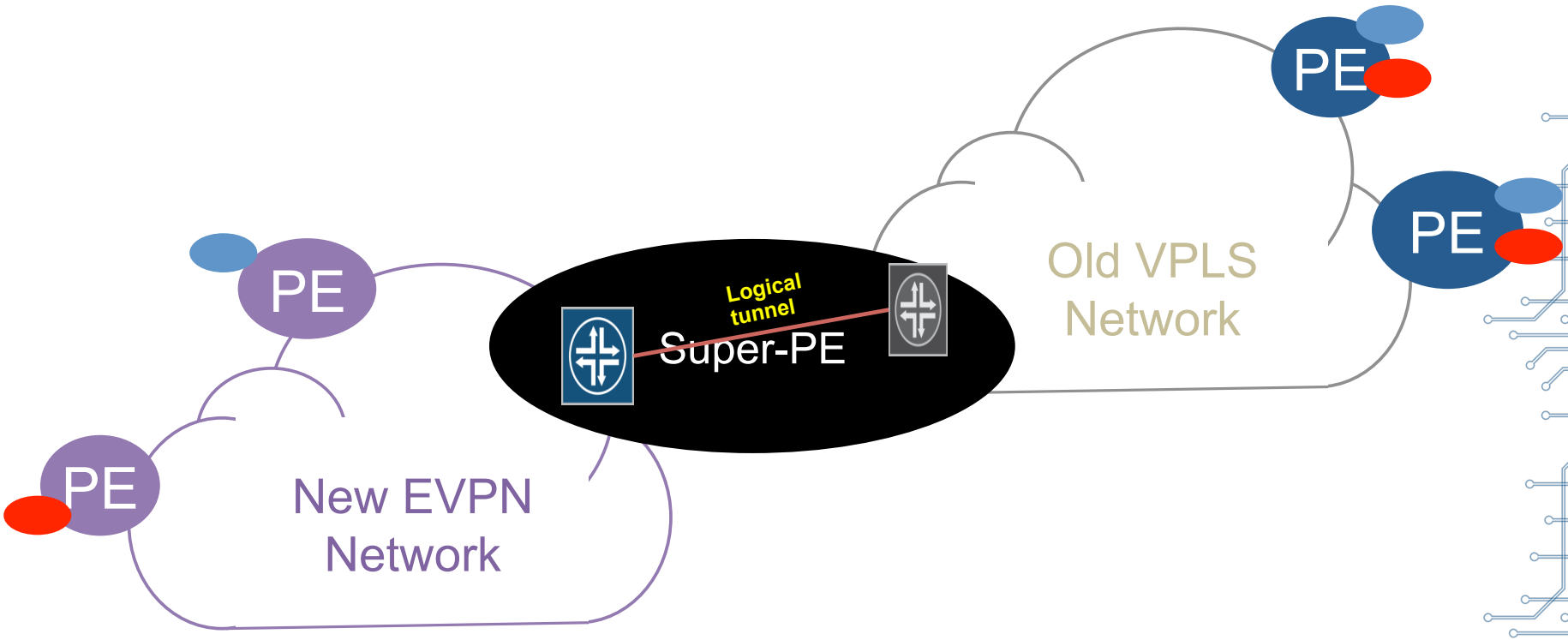
EVPN-MPLS Interworking with an MC-LAG Topology



Il nuovo modello



Migration path: VPLS -> EVPN



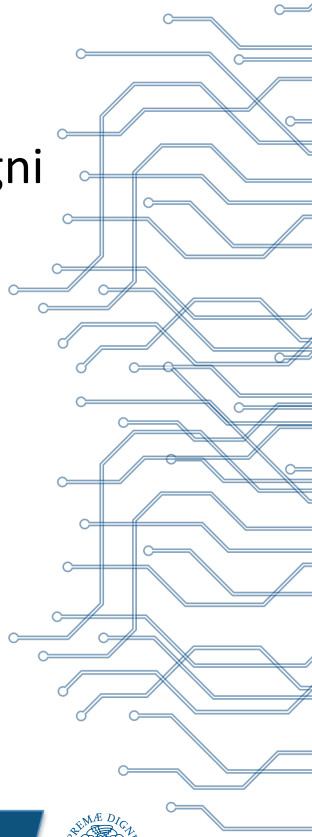
Migration Path

- VPLS -> EVPN in tecnologia vlan bundle
- VRF -> NLRI BGP type 5 – il bgp è troppo forte…
  - Multiprotocol extension per bgp 4
  - Route type-5 IP prefix Route—An IP prefix route provides encoding for inter-subnet forwarding. In the control plane, EVPN type-5 routes are used to advertise IP prefixes for inter-subnet connectivity across data centers. It is used to communicate between data centers (DC) when the Layer 2 connection does not extend across DCs and the IP subnet in a Layer 2 domain is confined within a single DC

Obiettivo: semplicità

- Separazione control plane – data plane
  - Presupposto per l'inserimento di tecnologie SW
- Provisioning nuove underlay network: banale. Una istanza vlan-bundle per ogni PE del datacenter

```
EVPN-Test {
    instance-type virtual-switch;
    interface ae1.1757;
    route-distinguisher 131.114.100.70:1757;
    vrf-target target:64861:1757;
    protocols {
        evpn {
            extended-vlan-list 1757, 1768, 2000;
        }
    }
    bridge-domains {
        vlan.1757 {
            vlan-id-list 1757, 1768, 2000;
        }
    }
}
```



Interrogativi aperti

- Routing nello spine o nel leaf?
  - Latenza vs semplicità di gestione
- Dove si aggancia il connettore vxlan: HW VTEP o SW VTEP?
- Dove il punto di incontro fra accesso (network providing) e erogazione dei servizi (cloud providing) – super-spine?
- Rimpiazzate VPLS e VRF, l'MPLS serve ancora nelle reti metropolitane?
  - Traffic engineering
  - CoS
  - Circuit cross connect
- La separazione assoluta control plane/data plane apre le porte a tecnologie SD-WAN. L'investimento è da fare?
- DCI con super spine o con strech ottico?