

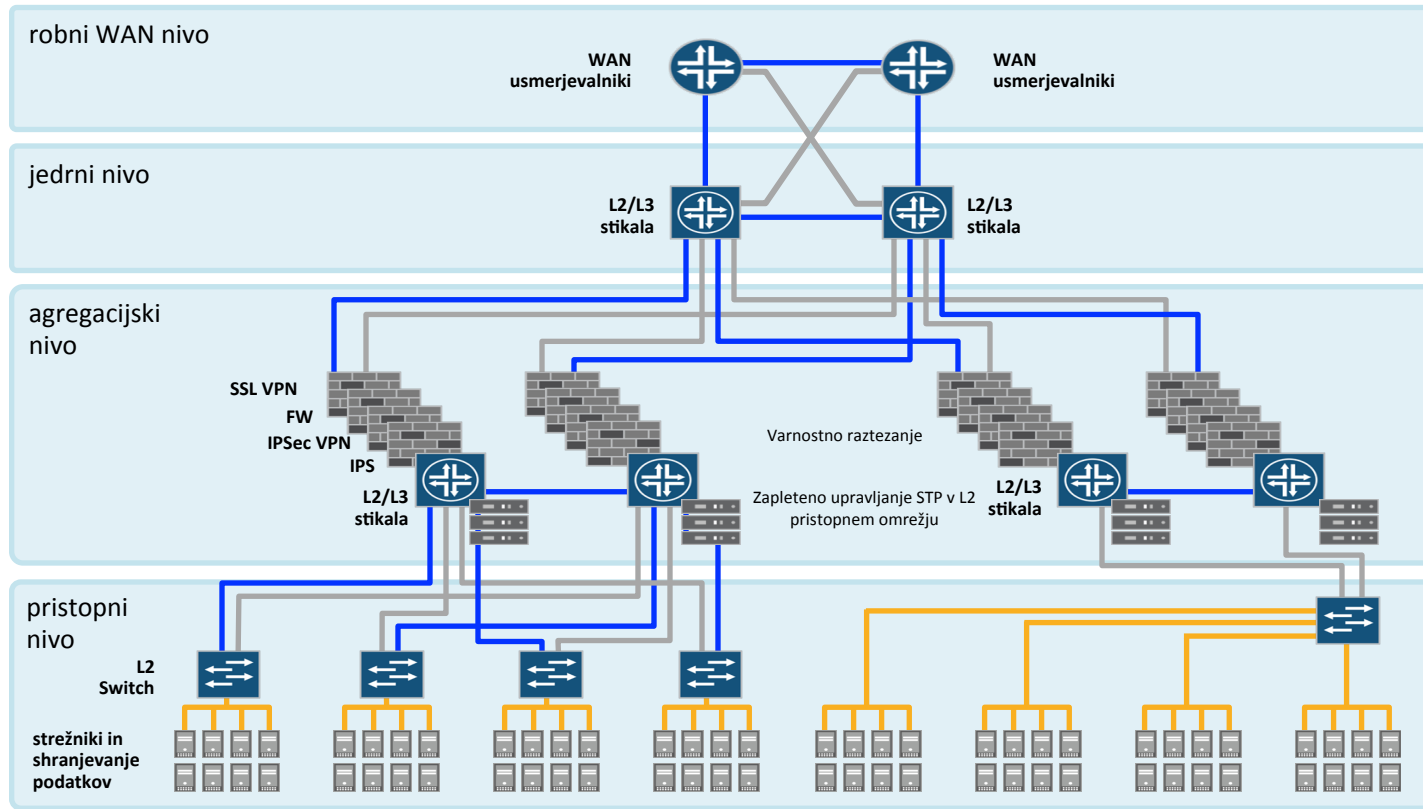


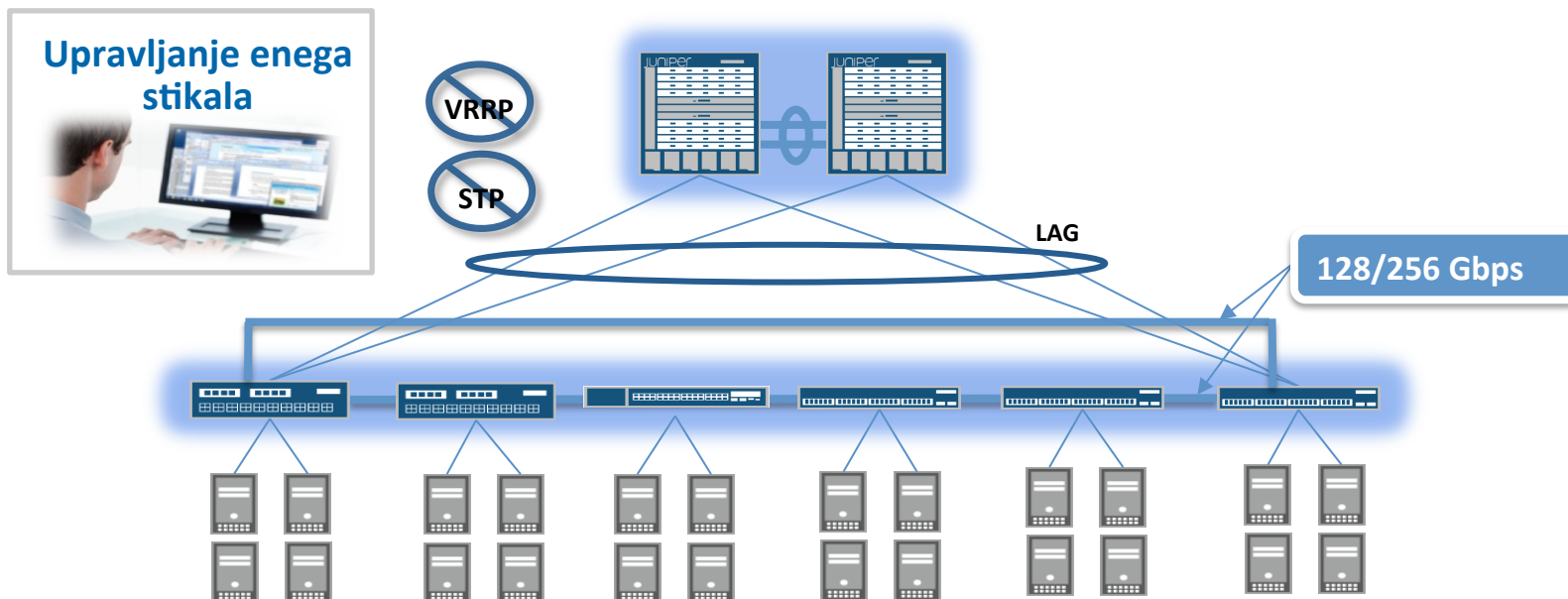
Arhitekture podatkovnih centrov SINOŠ

Janko Kersnik

janko.kersnik@smart-com.si

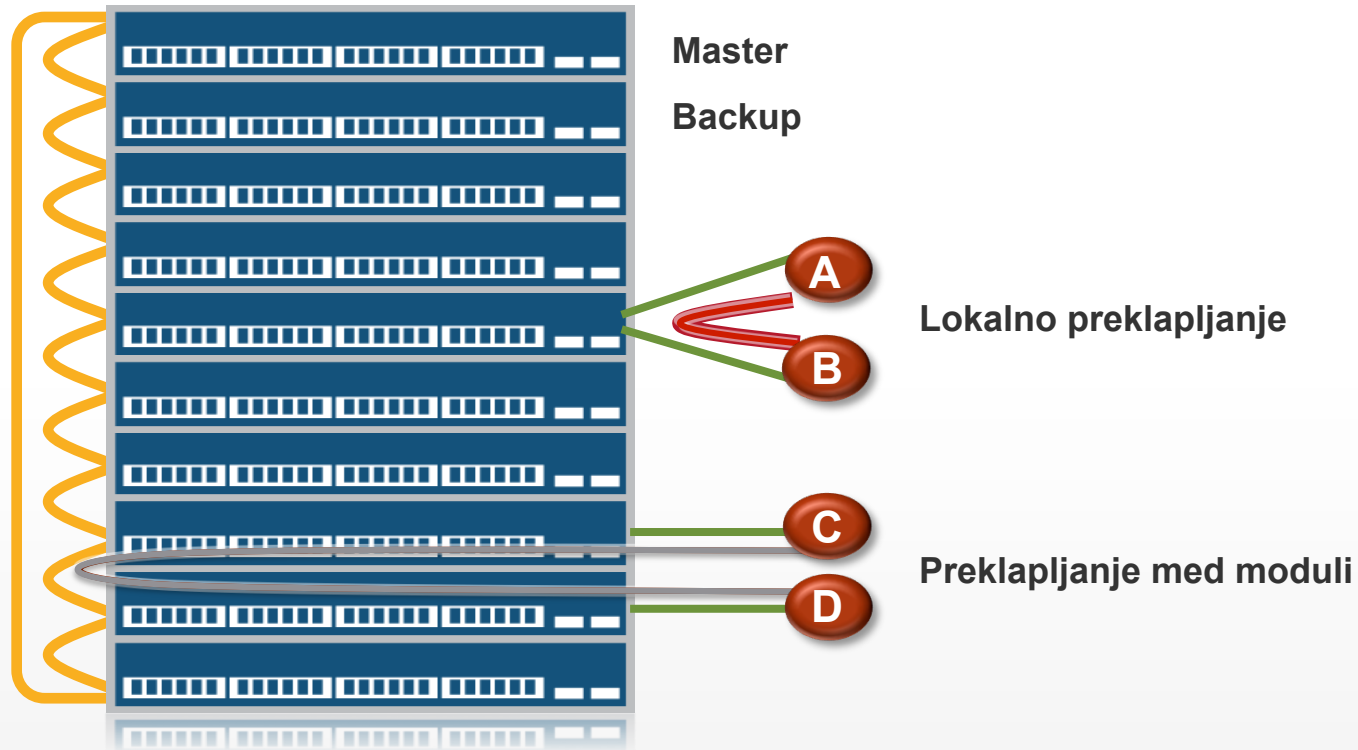
- 1 Preveč naprav in nivojev
- 2 Preveč operacijskih sistemov
- 3 Visoke zakasnitve
- 4 Izziv je zagotavljanje povezljivosti
- 5 Porabi se veliko časa za uvajanje novih aplikacij
- 6 Polovičen izkoristek povezav navzgor





- Konsolidacija pristopnega in agregacijskega nivoja
- prehod iz 3-nivojske v 2-nivojsko arhitekturo
- Ločene povezave za backplain (128 Gbps oziroma 256 Gbps)
- Prilagojeno povečanemu horizontalnemu prometu vzhod-zahod
- Odstrani STP in VRRP protokol
- Auto LAG omogoči polni izkoristek vseh povezav navzgor
- Upravljanje 10 stikal je v bistvu upravljanje enega stikala

Porazdeljeno preklapljanje

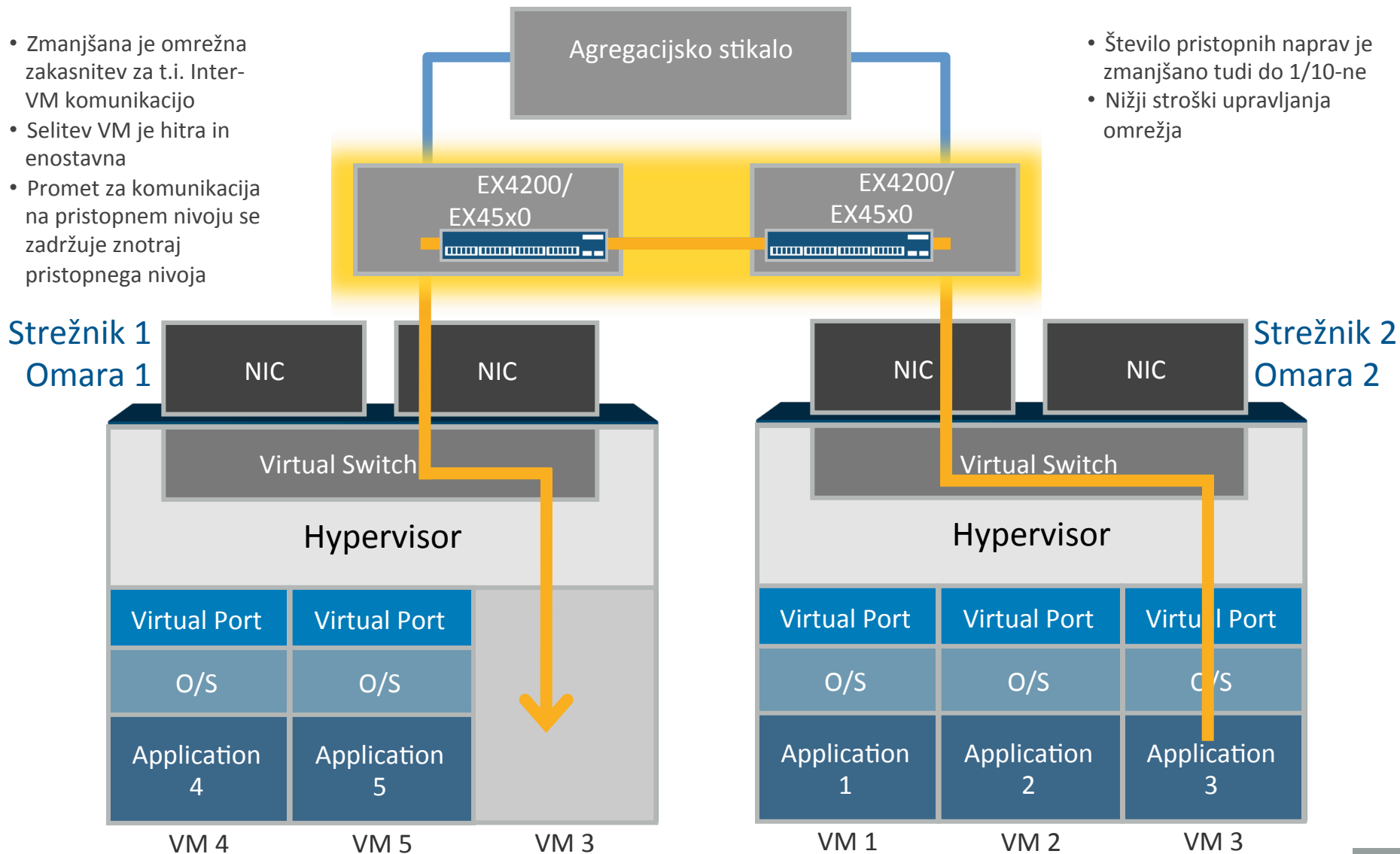


Primer horizontalnega aplikacijskega toka (vzhod-zahod)

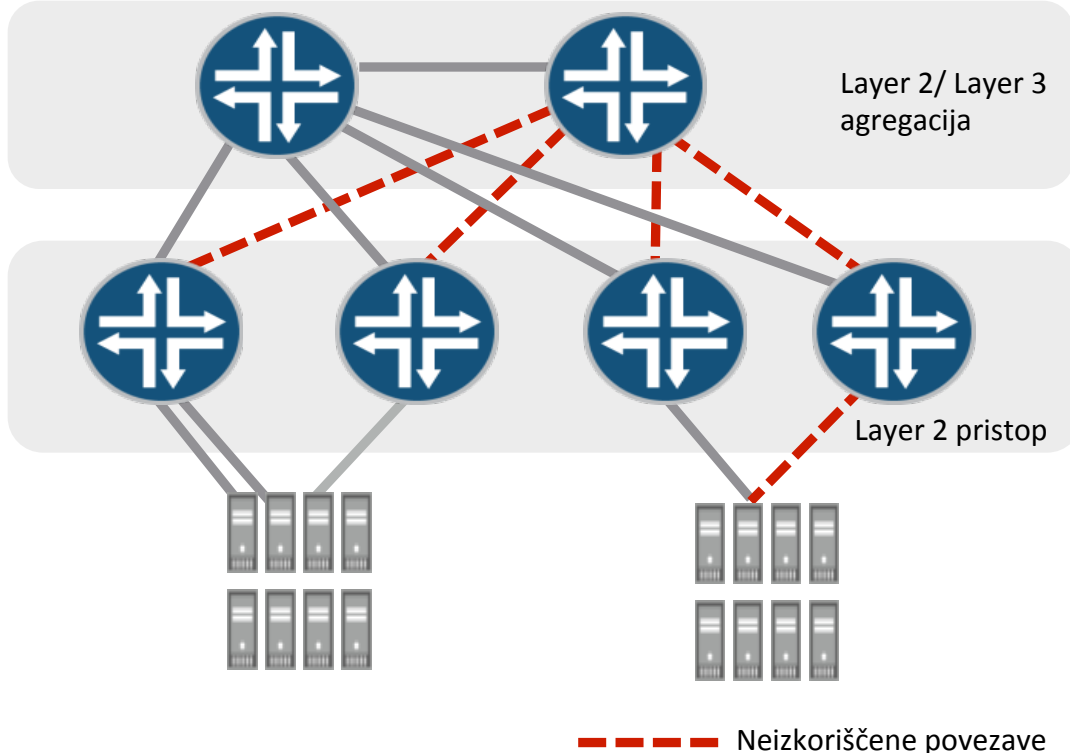
Selitev virtualnega strežnika

- Zmanjšana je omrežna zakasnitev za t.i. Inter-VM komunikacijo
- Selitev VM je hitra in enostavna
- Promet za komunikacija na pristopnem nivoju se zadržuje znotraj pristopnega nivoja

- Število pristopnih naprav je zmanjšano tudi do 1/10-ne
- Nižji stroški upravljanja omrežja



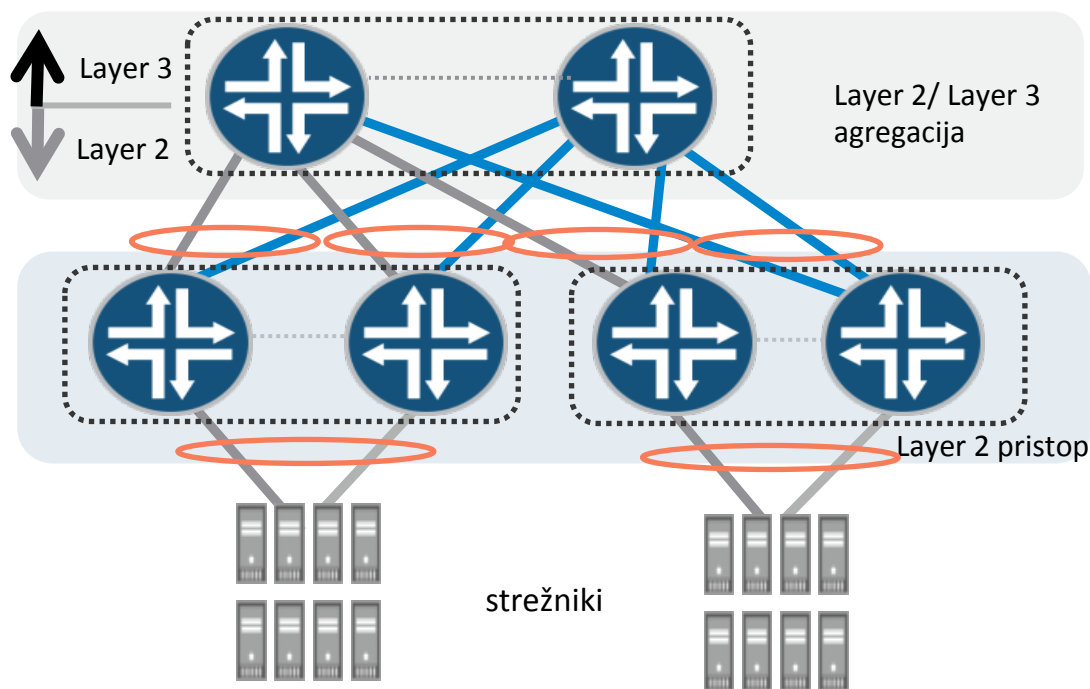
Tradicionala fizična topologija



- Neizkoriščena pasovna širina na povezavah navzgor
- Počasnejša Layer 2 konvergenca
- Neizkoriščena pasovna širina na povezavah navzgor
- Počasnejša Layer 2 konvergenca
- Dvojno vpetje na stikalo
- Osamljena točka napake
- Ni redundance na nivoju naprave
- Active-standby povezave do strežnikov

Primer uporabe: Layer 2 dostop in Layer 3 agregacija

Layer 3 povezljivost do prvega usmerjevalnika



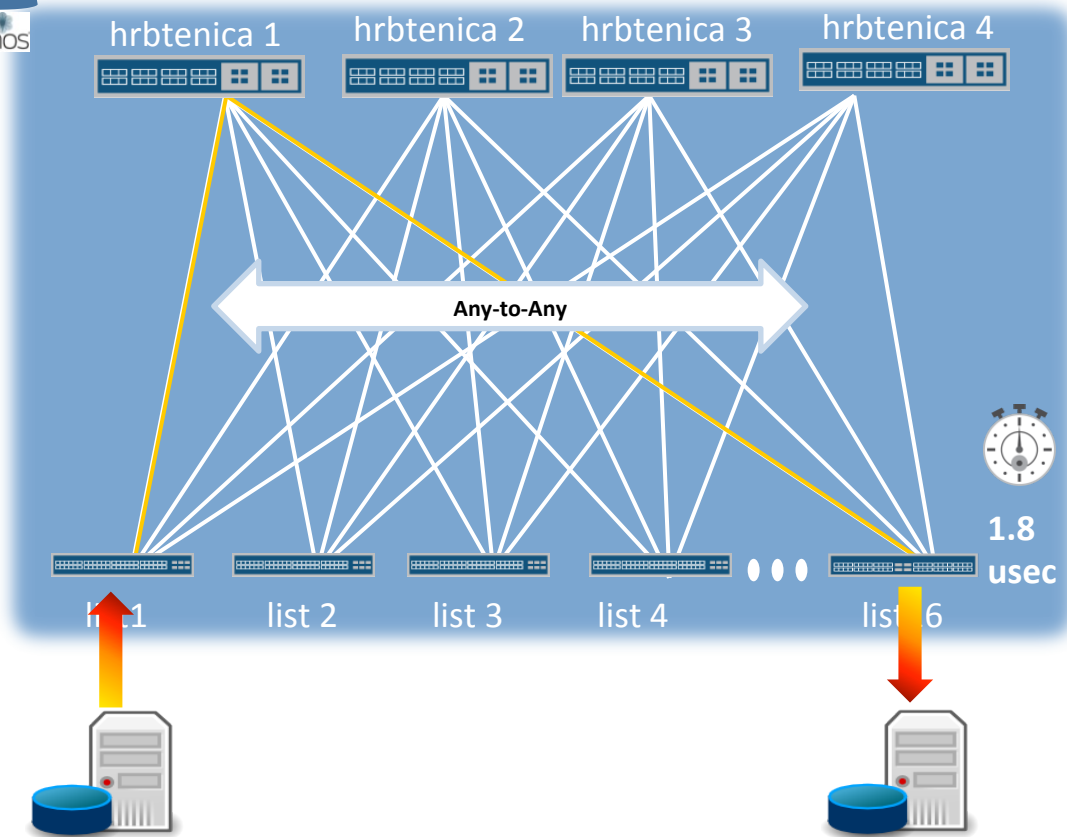
Prednosti:

- Odpravi potrebo po STP
- Vse povezave so active-active
- Hitrejša Layer 2 konvergenca
- Active/active dvojno vpetje strežnikov
- Ni osamljene točke napake
- Layer 3 prehod podpira MC-LAG

 = LAG

Tehnologija Virtual Chassis Fabric (VCF)

Arhitektura hrbtnica in list (Spine and Leaf) Ravninska razporeditev, Any-to-Any povezljivost



Upravljanje ene naprave

Princip priklopi in uporabi

Predvidljiva nizka zakasnitev

Napovedljiva učinkovitost

Pametno upravljanje omrežja

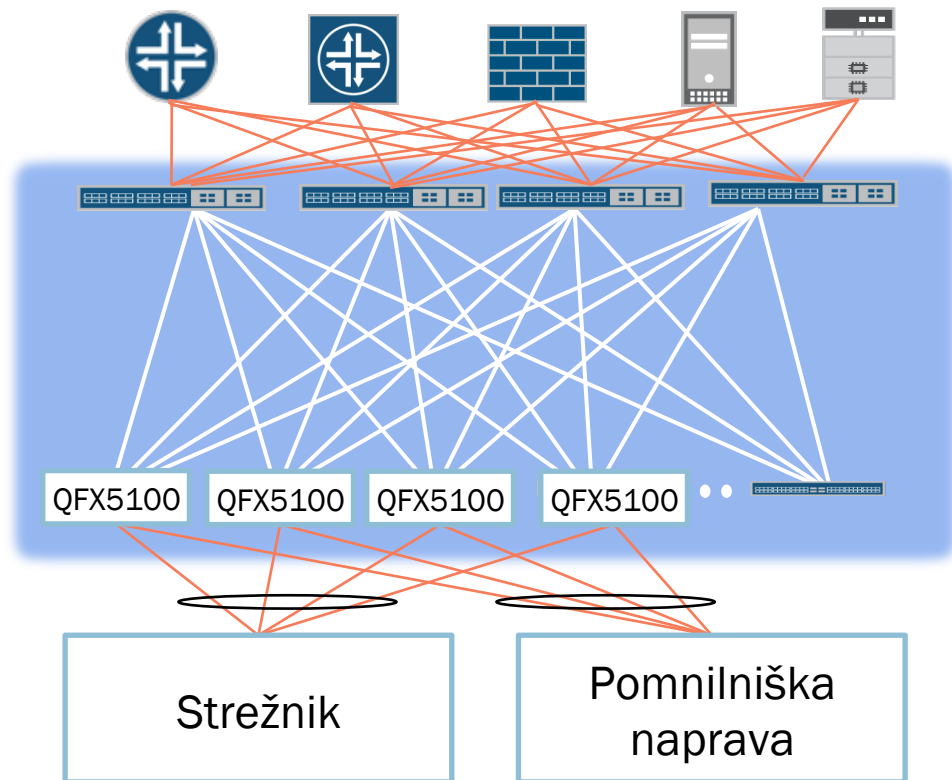
Podpora Layer 2 in Layer 3

Prožnost in visoka razpoložljivost

Topology-Independent In-Service

Software Upgrade (TISSU)

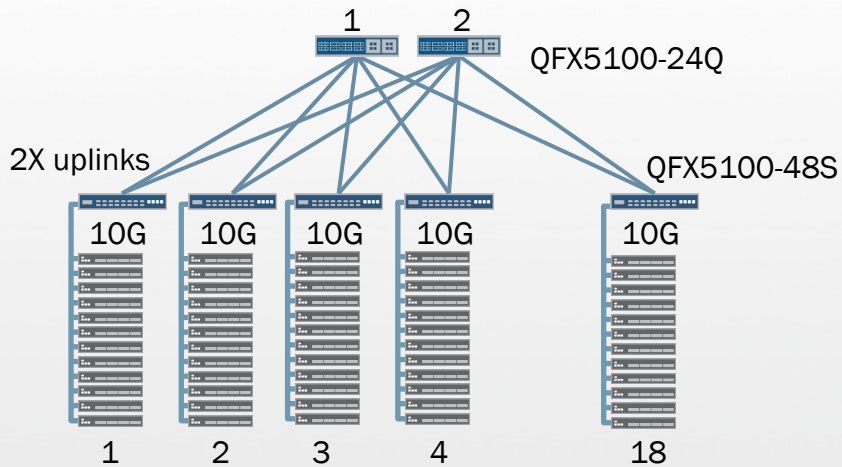
Struktura do 32 stikal



- Poljuben Ethernet medij, visoka prožnost, prilagodljiva uporaba
 - 10/100/1000M baker
 - 10/100/1000M optika
 - 10G baker
 - 10G optika
 - 10G ali 40G fabric
 - Priklop na katerikoli vmesnik
 - In-Service Software Upgrade
 - N-way multi-homing
 - Active-active poti
 - FCoE tranzit
 - Lossless Ethernet / DCB
 - iSCSI / NFS / CIFS
 - Strojna podpora za SDN
 - VTEP, OVSD, EVPN

2 hrbtnižni vozlišči

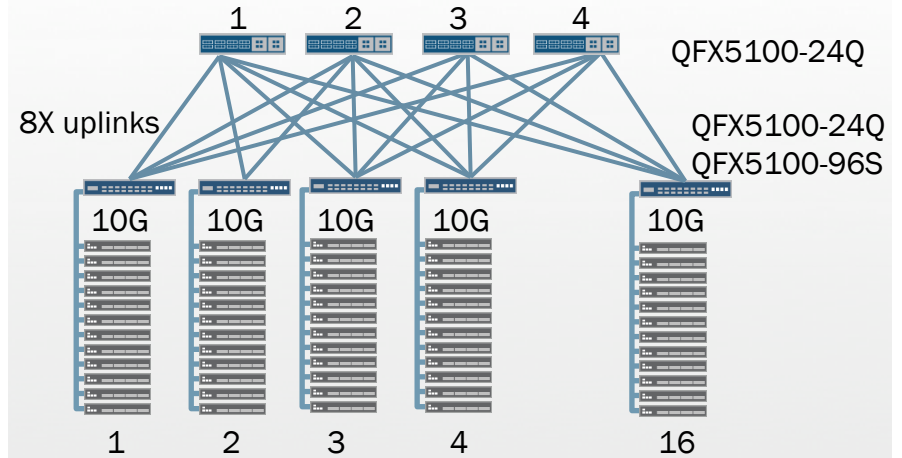
— 40G



- 18x 10GbE ohišji
- 864x 10GbE vmesnikov 6:1 OS

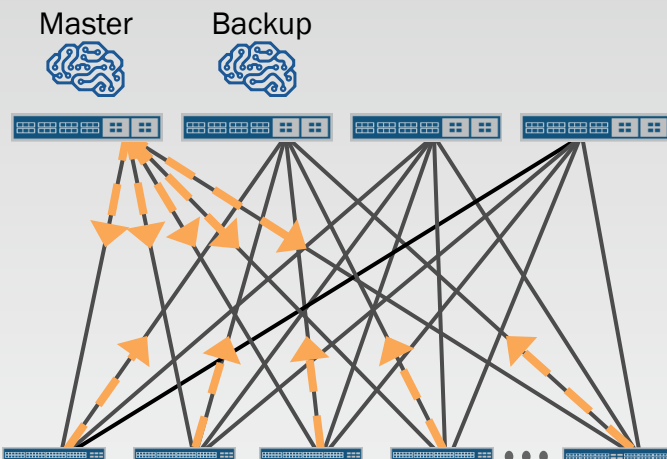
4 hrbtnižna vozlišča

— 40G



- 16x 10GbE ohišji
- 1,536x 10GbE vmesnikov 3:1 OS

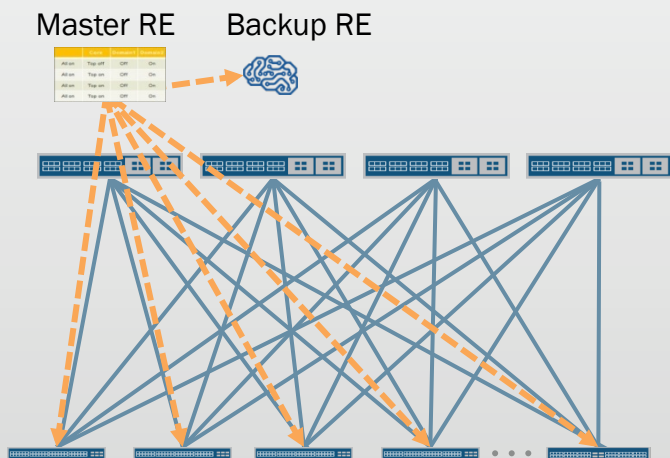
Kontrolna ravnina



Integriran usmerjevalnik in-band kontrolna ravnina

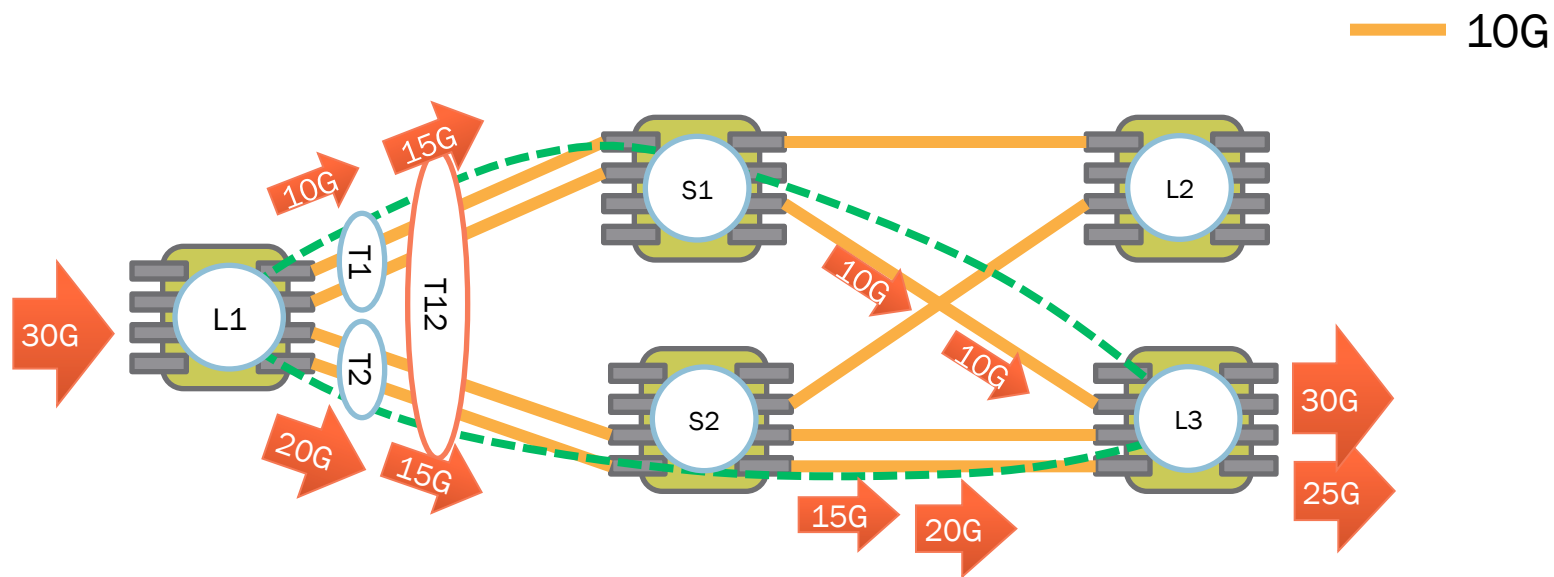
- Dvojni usmerjevalnik za redundanco
- Distribuirana serijska kontrolna ravnina
 - VCCPD teče na vseh članih
 - Avtomatsko odkrivanje topologije fabrica
 - Omogoča izgradnjo poti posredovanja brez zank v fabricu.
 - Sestavljena iz 40 Gbps povezav
- Zaščita kontrolnega prometa za konvergirani fabric

Podatkovna ravnina

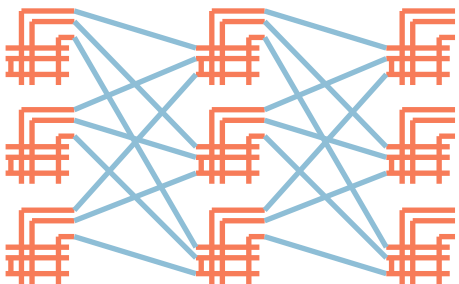


Pametna vozlišča hrbtnenice in lista Zvezna stanja Distribuirano posredovanje

- Nizke zakasnitve in predvidljiva učinkovitost
- Vse povezave v fabricu so active-active
- Promet je enakomerno razporejen med vsemi povezavami
- 550 nsec je zakasnitev v ohišju
- 1.8 μ sec je zakasnitev med ohišji
- 16-kratna večdomna povezava strežnikov
- podpora t.i. smart trunks



- Avtomatski fabric trunki
- Vrste fabric trunkov
 - Next Hop (NH)-trunki: do neposrednega soseda
 - Remote Destination (RD)-trunki: do ponorne točke
- Utežno razmerje pasovne širine z upoštevanjem več poti (namesto NH povezave); z namenom, da se izognemo preobremenjenosti fabrica



sredina



vhod

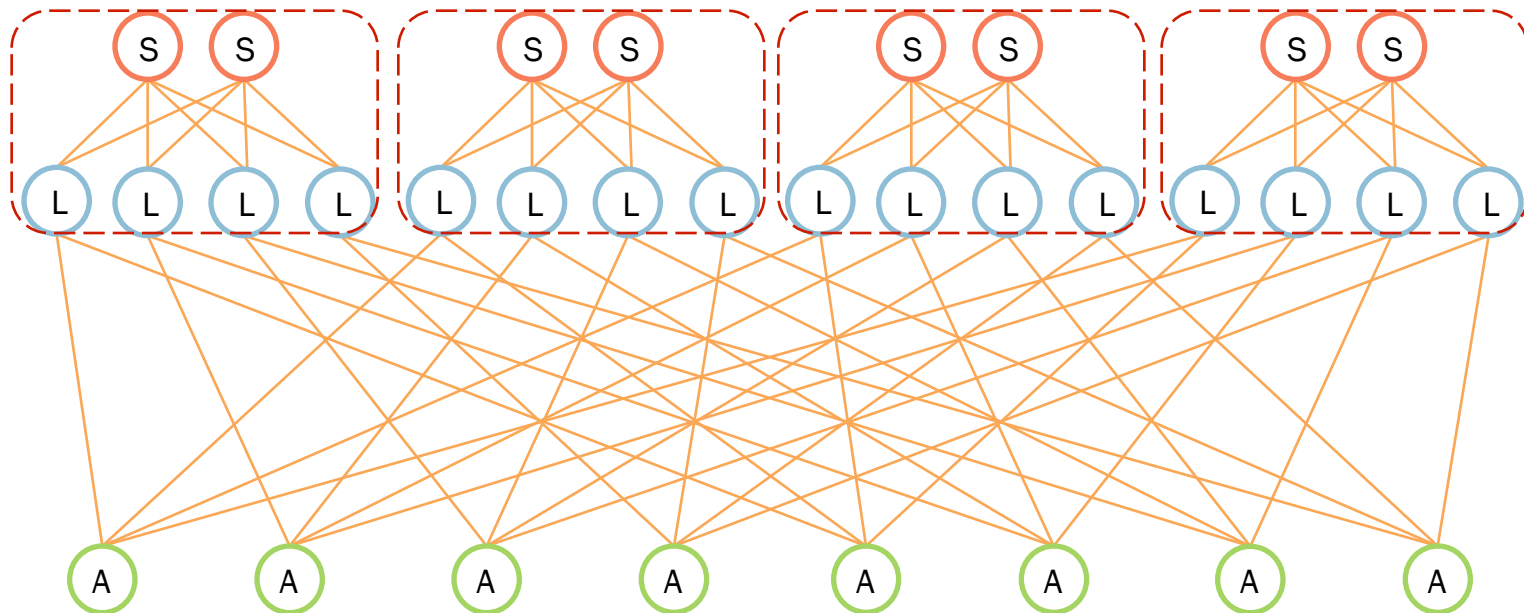
izhod


smer širitve




5 stopenjski Clos s pomočjo t.i. vSpine zasnove

- vSpine je lahko VCF ali 3 stopenjski Close
- Pristopna stikala A so npr. predimenzionirana v razmerju 3:1
- vSpine ohranja razmerje 1:1



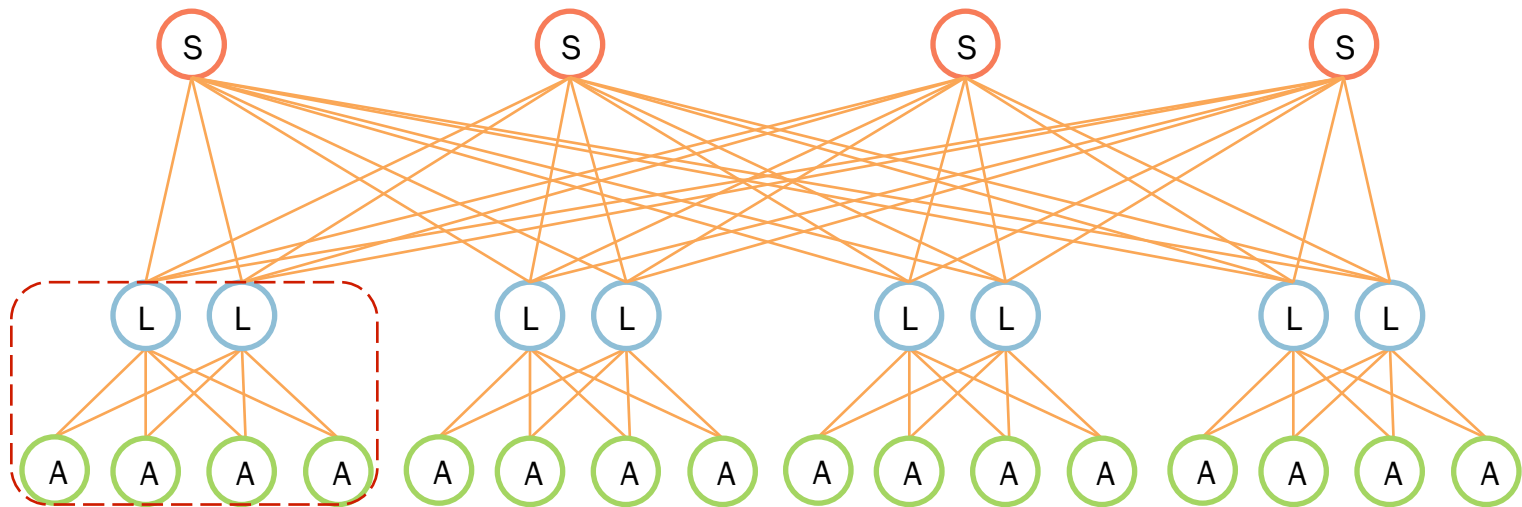
 S QFX5100-24Q


 L QFX5100-24Q


 A QFX5100-48S

5 stopenjski Clos s pomočjo t.i. POD zasnove

- POD je sestavljen iz dveh listov in štirih pristopnih stikal
- POD je npr. predimenzionirana v razmerju 3:1
- Predimenzioniranost(oversubscription) pri prehajanju iz PODa v POD je večja



 QFX5100-24Q

 QFX5100-24Q

 QFX5100-48S

Spine

- Centralna ravnina – večstopenjski Clos
- Vedno 1:1 over-subscription
- Omogoča BGP route reflection
- Uporaba iBGP proti listom (Leaf)

Leaf

- NNI iz večstopenjskega Closa
- Prilagodljiv over-subscription
- Uporablja iBGP proti hrbtenici (spine)
- Uporablja eBGP proti dostopu

Access

- Omogoča dostop do končnih točk (strežniki in pomnilniške naprave)
- Tipični 3:1 over-subscription v poslovnih okoljih in pri ponudnikih storitev, 1:1 v HPC
- Uporablja eBGP proti vSpine vozliščem
- Omogoča storitve Layer 3 prehoda za končne točke
- Omogoča agregacijo povezav do končnih točk

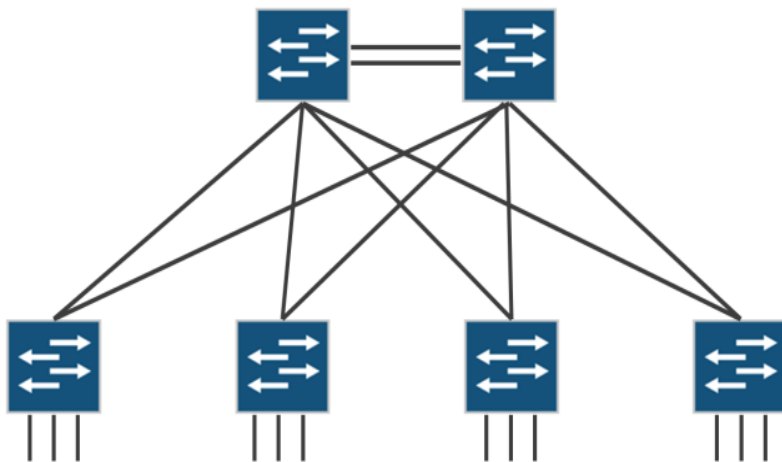
vSpine

- Kombinacija hrbtenice in lista (spine and leaf)
- Se obnaša kot logično stikalo
- Navidezna peering točka za pristopni nivo
- Over-subscription je odvisen od vloge hrbtenice in lista
- Samo ena BGP številka avtonomnega sistema
- Uporablja eBGP do pristopnih stikal



- Velik obseg: preko 73,000 10GbE pristopnih vmesnikov
- Visoka učinkovitost: prilagodljiv over-subscription, 1:1 do N:1
- Investiraj glede na potrebe: začni z majhnimi koraki in nato širi
- Nizka zakasnitev s fiksnimi stikali
- Zelo omejen vpliv napak na omrežje
- Standardizirana postavitvev: podpora različnih proizvajalcev
- Poznana zakasnitev s fiksno hrbtnica list topologijo
- Zelo prilagodljiva fizična postavitvev: TOR, EOR in MOR

Junos Fusion



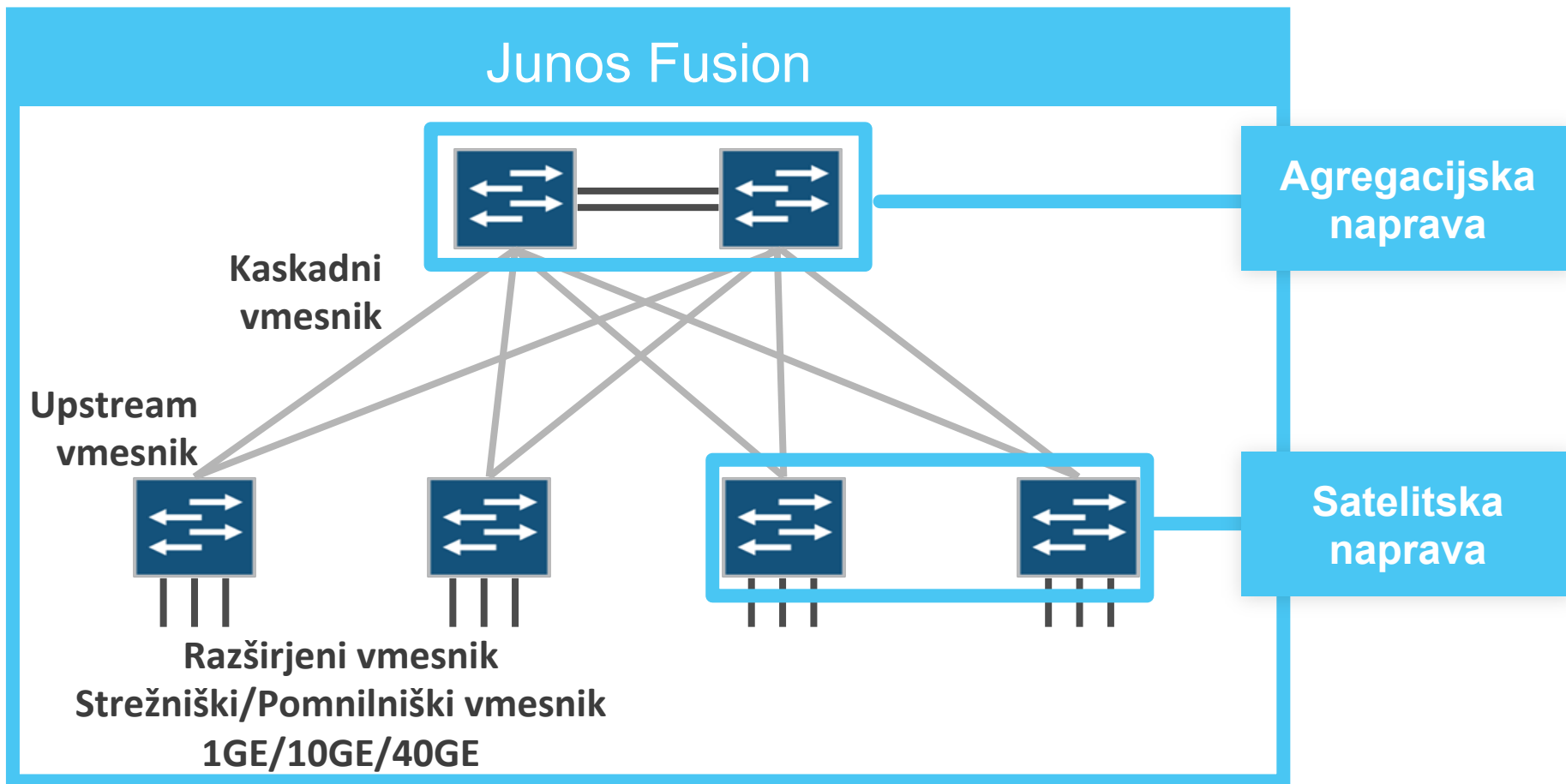
Vzpostavitev omrežja podatkovnega centra s poenostavljenim upravljanjem in možnostjo hitre širitve

Odprti standardi in programabilnost IEEE 802.1BR in JSON-RPC API-ji

Prožnost

Priklopi in uporabi zagotavljanje virov

1GE-100GE

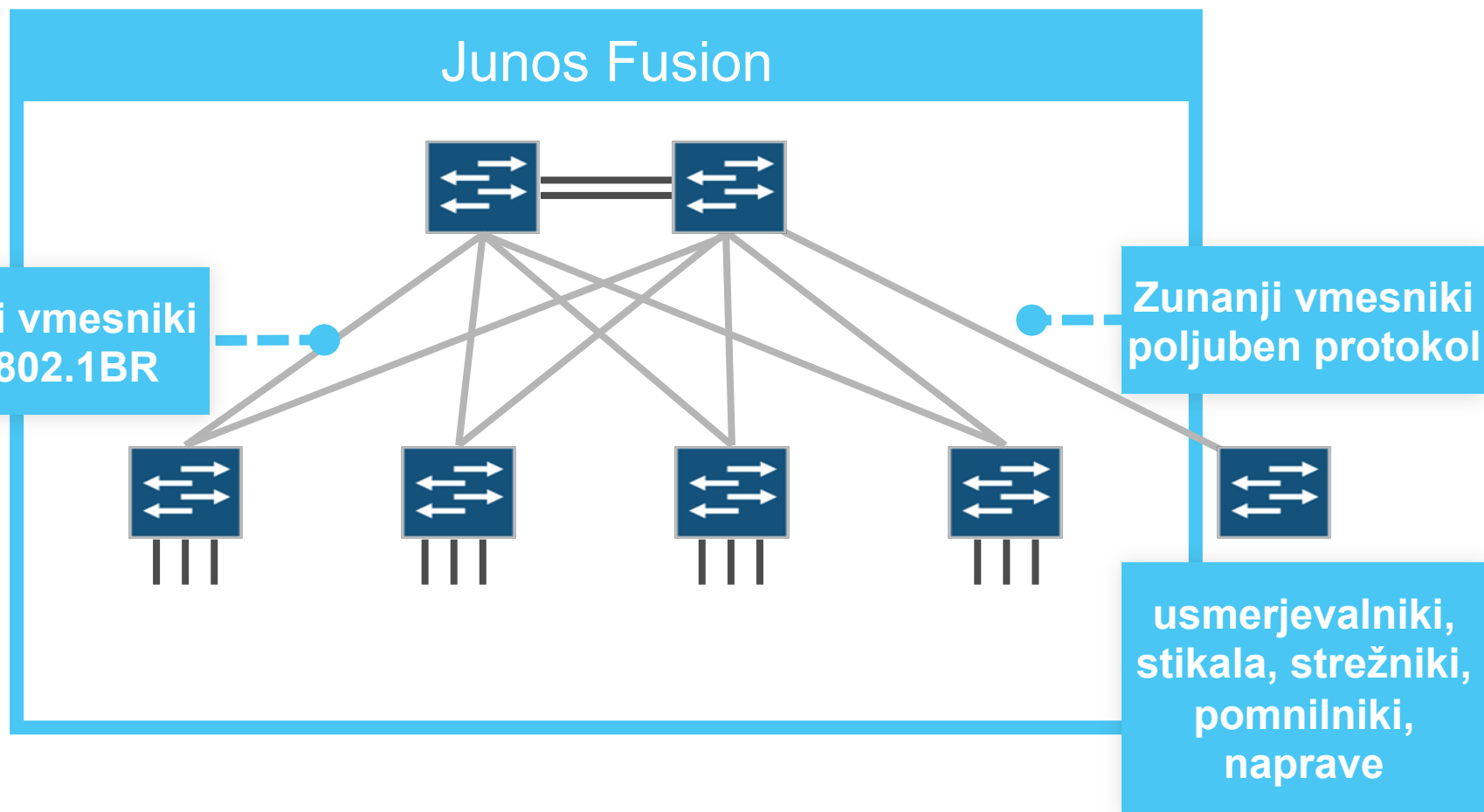


AGREGACIJSKA NAPRAVA

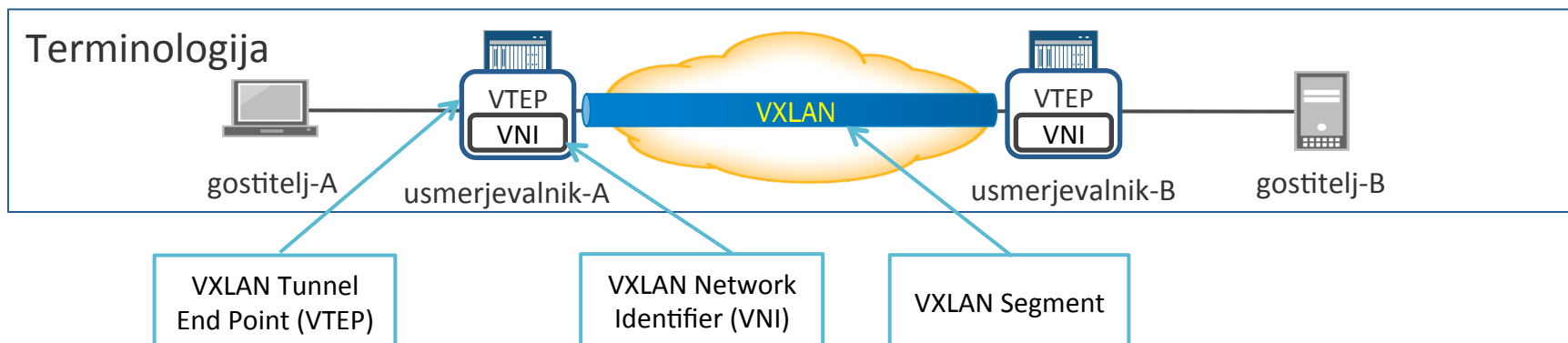
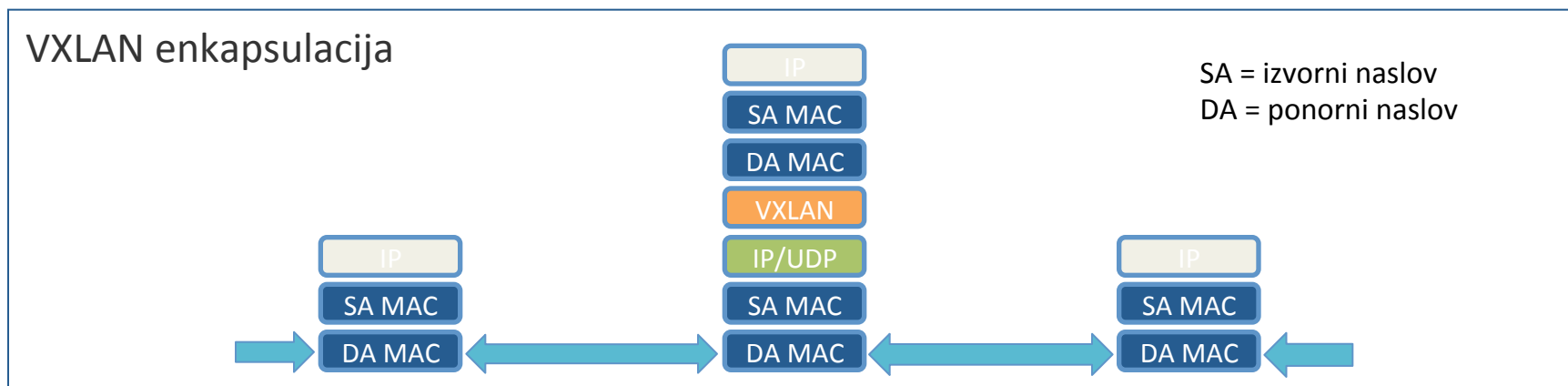
- Eno ali več stikal, ki upravlja satelitske naprave (top-of-rack stikala) na daljavo
- Konfiguracija, upravljanje programske opreme, pobiranje statistike
- Avtomatsko odkrivanje in upravljanje satelitskih naprav
- Omogoča notranjo povezljivost za upravljanje satelitskih naprav
- Uporablja Junos OS

SATELITSKA NAPRAVA

- Standardna top-of-rack stikala, ki se upravljajo na daljavo s pomočjo agregacijskih stikal
- Ni potrebno lokalno upravljanje
- Priklop na AN je lahko eno ali dvodomen
- Podpira večdomen priklop kateregakoli gostitelja (strežnik, naprava) na par satelitskih naprav
- Uporablja Windriver Yocto Linux

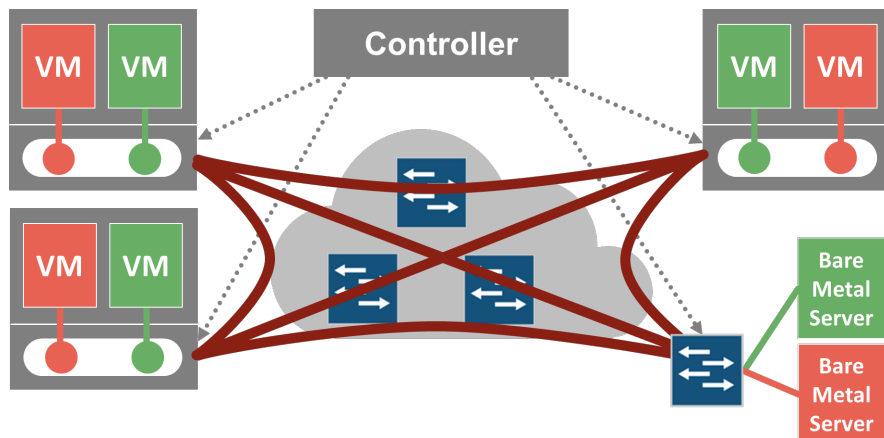


- Virtual Extensible Local Area Network
- Raztegne Layer 2 povezavo preko Layer 3 omrežja s pomočjo enkapsulacije Ethernet okvirja v VXLAN paket, ki vsebuje IP naslov
- VTEP skrbijo za enkapsulacijo/de-kapsulacijo prometa, ki se prenaša preko VXLANov

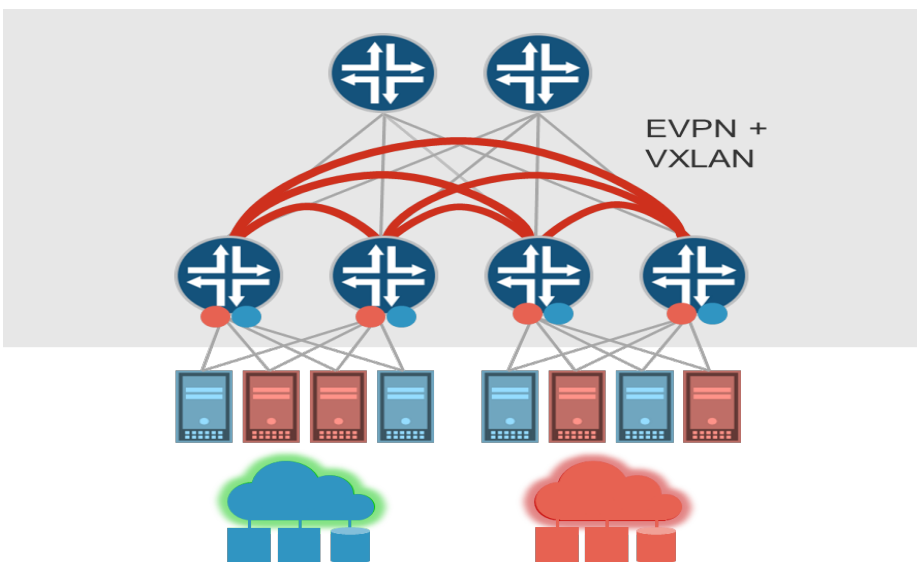


- **Prednosti uporabe VXLAN-ov:**
 - Teoretično je možno kreirati 16 milijonov VXLAN-ov v administrativnih domenah
 - Omogoča selitev virtualnega računalnika med strežniki, ki se nahajajo v ločenih Layer 2 domenah, s pomočjo tuneliranja preko Layer 3 omrežja
 - Ni potreben STP za pripravo same topologije
 - Vse povezave se lahko uporabijo
 - Promet je lahko deljen preko več povezav
 - Poveča učinkovitost

Prekrivno omrežje z in brez krmilnika



- Skupinska politika
- Visoko virtualizirana infrastruktura
- Večja skalabilnost za uporabnike
- Možnost širitve omrežja



- Kontrolna ali podatkovna ravnina imata nadzor nad gostitelji
- Prekrivanje na robu omrežja
- Priklop baremetal gostiteljev
- Okolje za gostovanje

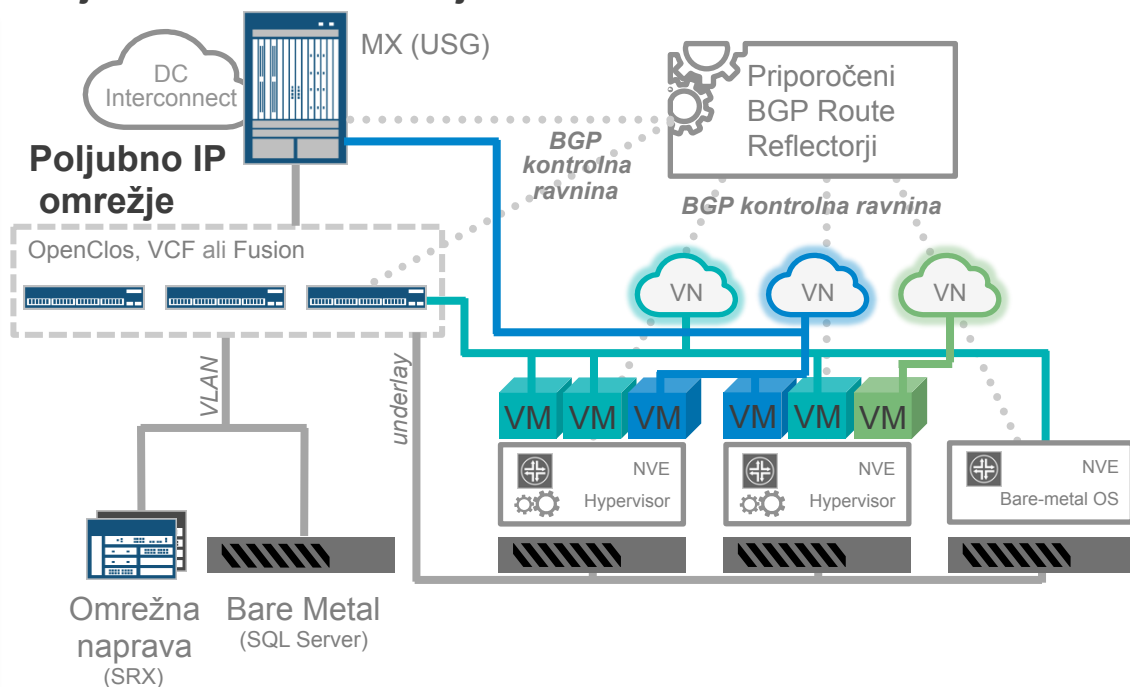
Prikazan bo primer brez krmilnika

Ethernet VPN znotraj podatkovnega centra

Skalabilno L2 prekrivno omrežje z VXLAN

- Brez preplavljanja. VXLAN in BGP komunicirata preko IP omrežja
- IETF standardizirana zvezna kontrolna ravnina, ki je distribuirana in skrbi za integrirano usmerjanje in mostičenje
- Končne postaje so lahko gostitelji (hypervisor) ali ToR
- Skalabilno za velike podatkovne centre z logično centraliziranim route reflectorjem (npr. Juniper (v)MX ali Cisco APIC)
- Vse povezave posredujejo promet s pomočjo ECMP znotraj samega omrežja
- EVPN podpira različne tipe enkapsulacij: NVGRE, MPLSoGRE in MPLS

Poljubni DC robni usmerjevalnik

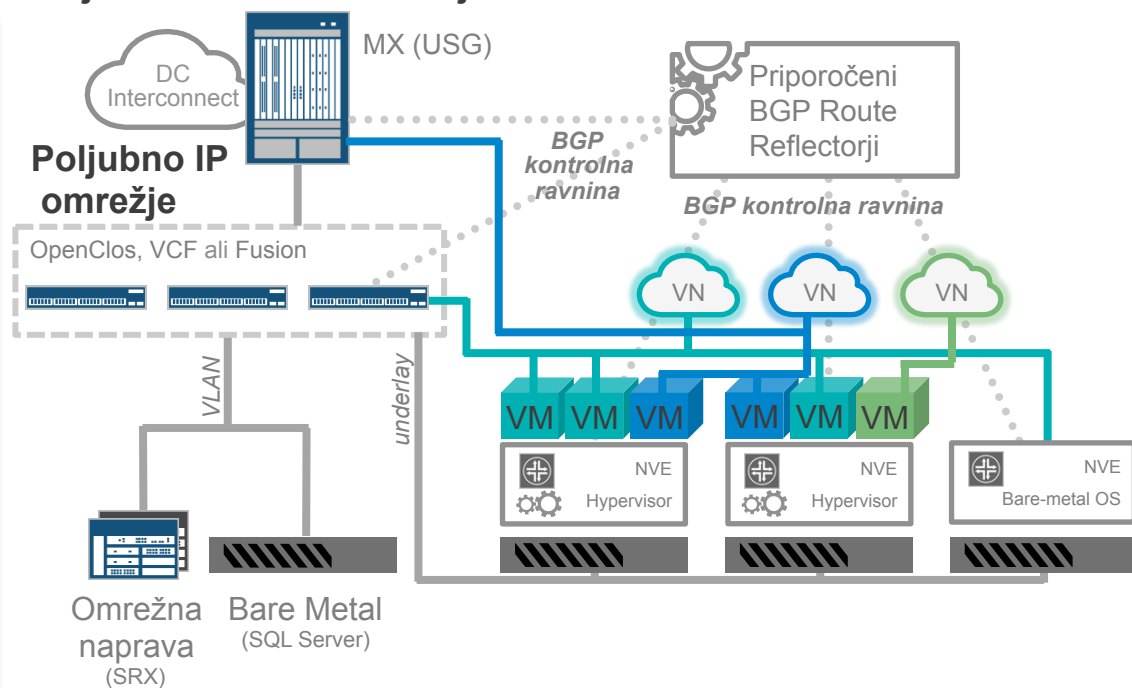


NVO = Network Virtualization Overlay
 NVE = NVO Endpoint (ali Network Virtualization Edge)

PREDNOSTI REŠITVE

- V IP omrežju je lahko uporabljeno katerokoli stikalo
- Uporabljena je EVPN tehnologija, ki se uporablja za povezavo med podatkovnimi centri s pomočjo MPLS z TE
- Odprt standard široko uporabljen s strani Juniper, Cisco in ALU
- Omogoča enostavno povezavo stikal, usmerjevalnikov in SDN krmilnikov

Poljubno DC robni usmerjevalnik



NVO = Network Virtualization Overlay

NVE = NVO Endpoint (ali Network Virtualization Edge)



Smart Com d.o.o.

Brnčičeva ulica 45

1231 Ljubljana – Črnuče

www.smart-com.si, info@smart-com.si