Openstack Networking Design

Pete Lumbis – CCIE #28677, CCDE 2012::3

Cumulus Networks Technical Marketing Engineer

Openstack Overview

Takes a pool of servers

Deploys VMs (OS, disk, memory, CPU cores, etc)

Attaches VM to networks

Resource management

“Microservice” style. Each thing is a stand alone component

- Openstack “Projects” (i.e., Neutron, Nova, Cinder, etc)
- Why Openstack is seen as complex
- Networking vs Compute vs Storage vs GUI
- Work via APIs, not tightly coupled

Openstack Nova – Compute Services

Deploys VMs

Manages CPU, memory, disk size

“Nova Nodes” are servers that can run VMs

Openstack Neutron – Networking Services

Manages list of tenant networks

- Tenant = VLAN/VxLAN

Assigns tenant network to VM

- Programs network stack on “Nova Nodes” based on user config

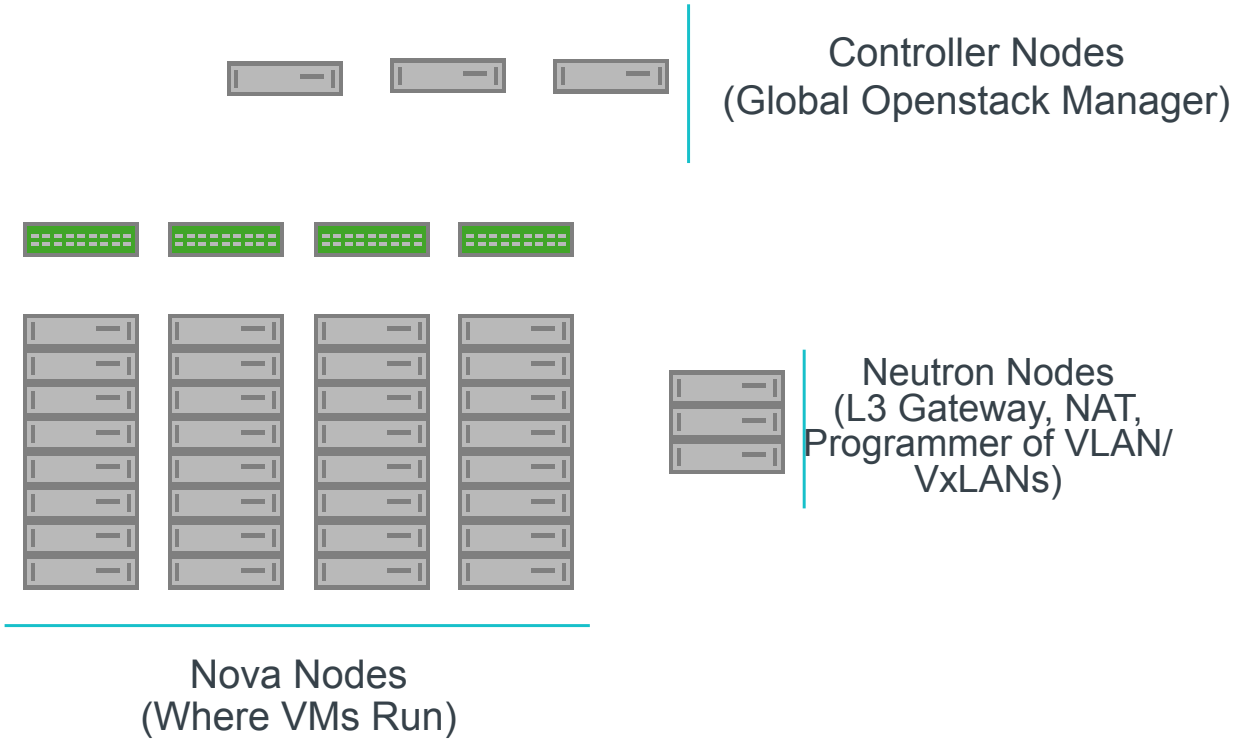
(Optional) ML2 Driver: Neutron server to switch API

- Switch runs the driver to translate Openstack API to local config/state

Neutron focus is layer 2. L3 generally done on server software

- L3 gateway, NAT, ACLs done on a centralized Neutron x86 Node
- DVR (Distributed vRouter) allows for L3 on Local Compute Node (Nova)

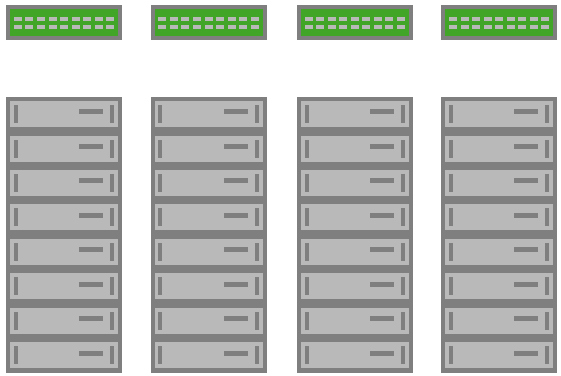
Relevant Components



Relevant Components

Controller Nodes
(Global Openstack Manager)

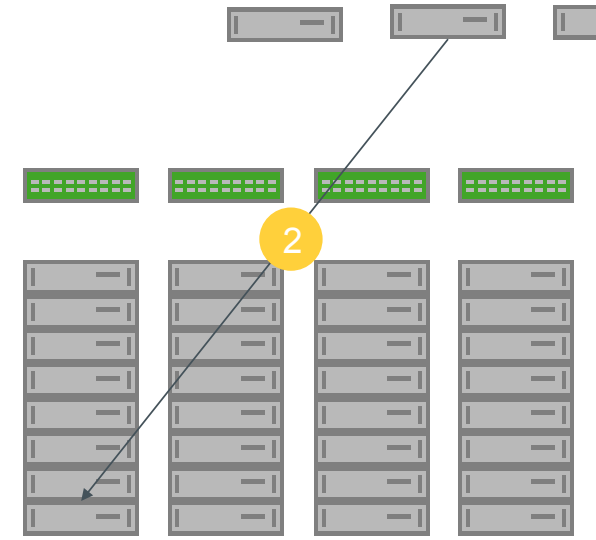
1 User requests a new VM, supplying parameters like amount of RAM and the network (vlan/l2 segment) they want to be on



Nova Nodes
(Where VMs Run)

Neutron Nodes
(L3 Gateway, NAT, Programmer of VLAN/ VxLANs)

Relevant Components



Nova Nodes
(Where VMs Run)

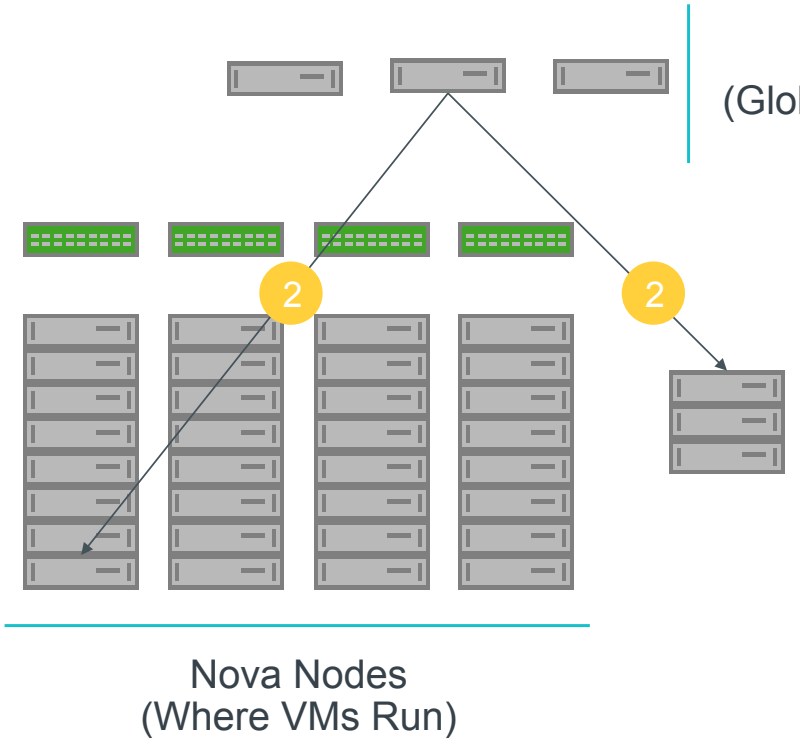
Controller Nodes
(Global Openstack Manager)

Neutron Nodes
(L3 Gateway, NAT,
Programmer of VLAN/
VxLANs)

1. User requests a new VM, supplying parameters like amount of RAM and the network (vlan/l2 segment) they want to be on

2. Openstack Controller magically selects a Nova Node to deploy the VM on.

Relevant Components



Controller Nodes
(Global Openstack Manager)

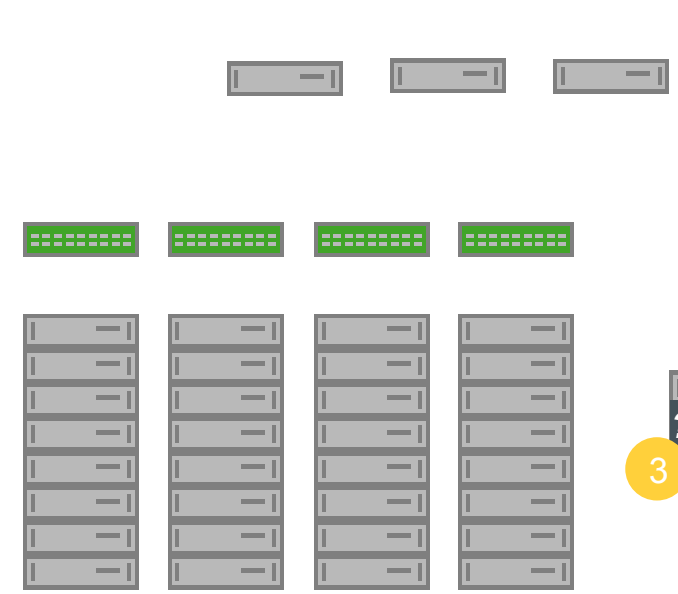
Neutron Nodes
(L3 Gateway, NAT,
Programmer of VLAN/
VxLANs)

1 User requests a new VM, supplying parameters like amount of RAM and the network (vlan/l2 segment) they want to be on

2 Openstack Controller magically selects a Nova Node to deploy the VM on.

2 Simultaneously Openstack Controller tells Neutron Node about a new VM, the Nova node and desired VLAN

Relevant Components



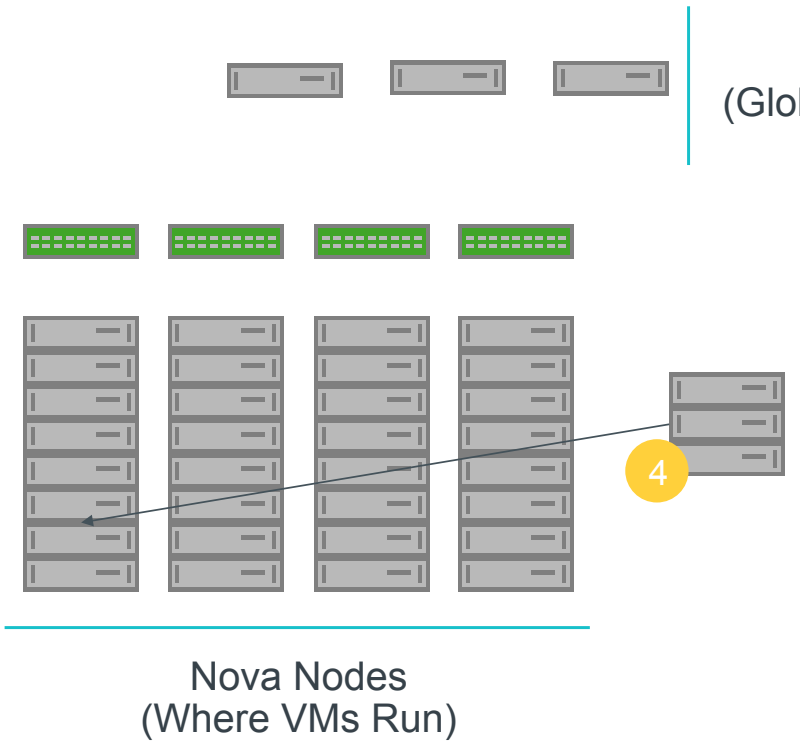
Nova Nodes
(Where VMs Run)

Controller Nodes
(Global Openstack Manager)

Neutron Nodes
(L3 Gateway, NAT,
Programmer of VLAN/
VxLANs)

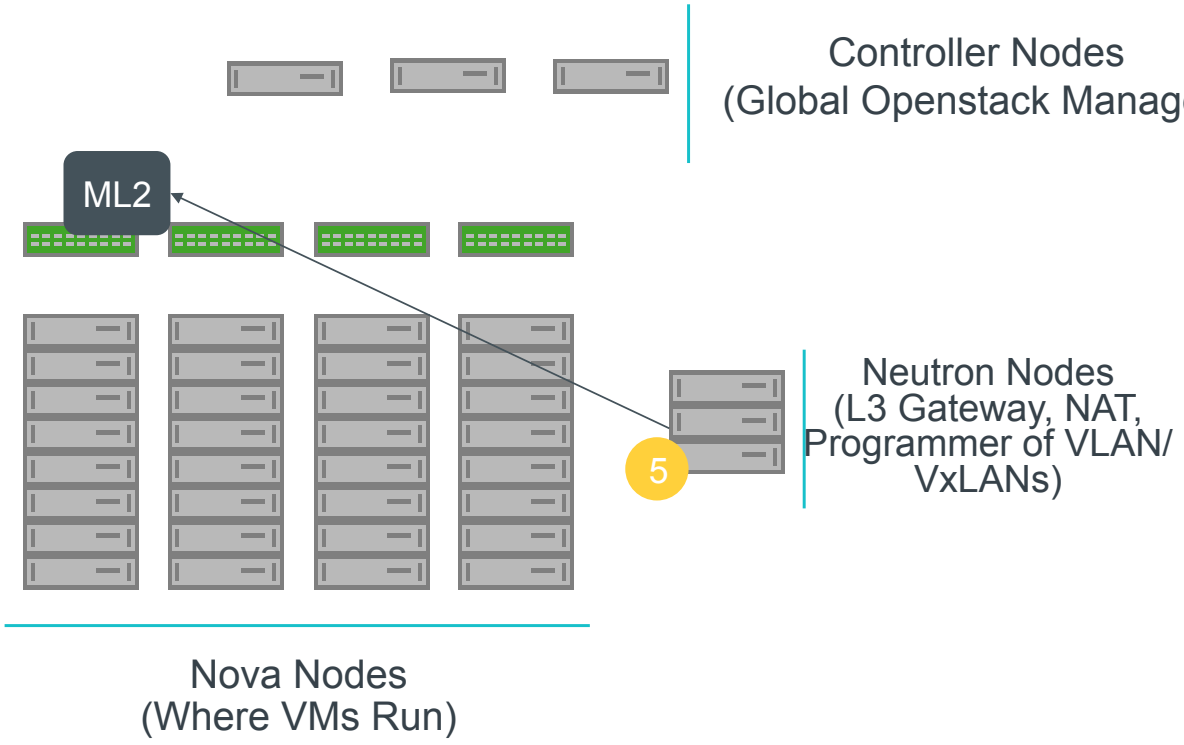
1. User requests a new VM, supplying parameters like amount of RAM and the network (vlan/l2 segment) they want to be on
2. Openstack Controller magically selects a Nova Node to deploy the VM on.
2. Simultaneously Openstack Controller tells Neutron Node about a new VM, the Nova node and desired VLAN
3. Neutron decides if the network already exists. If no, a new L3 gateway is created on the Neutron Node.

Relevant Components



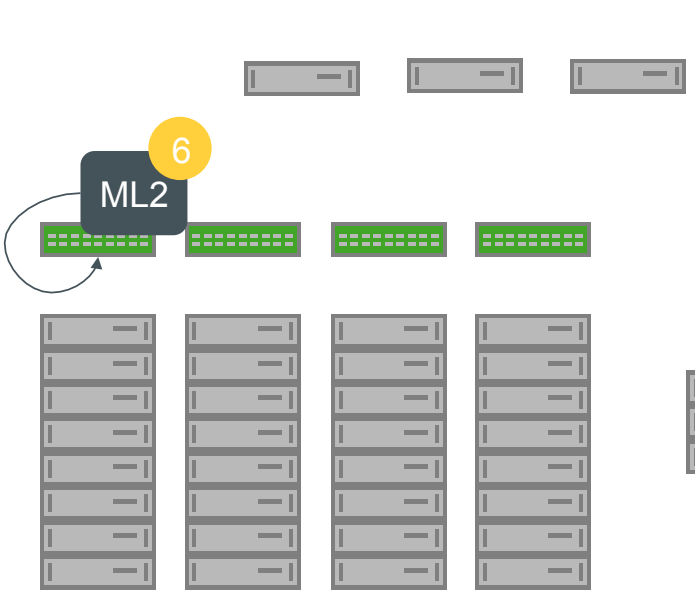
1. User requests a new VM, supplying parameters like amount of RAM and the network (vlan/l2 segment) they want to be on
2. Openstack Controller magically selects a Nova Node to deploy the VM on.
2. Simultaneously Openstack Controller tells Neutron Node about a new VM, the Nova node and desired VLAN
3. Neutron decides if the network already exists. If no, a new L3 gateway is created on the Neutron Node.
4. Neutron builds L2 config on the Nova device

Relevant Components



1. User requests a new VM, supplying parameters like amount of RAM and the network (vlan/l2 segment) they want to be on
2. Openstack Controller magically selects a Nova Node to deploy the VM on.
2. Simultaneously Openstack Controller tells Neutron Node about a new VM, the Nova node and desired VLAN
3. Neutron decides if the network already exists. If no, a new L3 gateway is created on the Neutron Node.
4. Neutron builds L2 config on the Nova device
5. If deployed to do so, Neutron will speak, via the ML2 Driver, to a hardware switch to provision a VLAN or VxLAN (along with or instead of step 4)

Relevant Components



Nova Nodes
(Where VMs Run)

Controller Nodes
(Global Openstack Manager)

Neutron Nodes
(L3 Gateway, NAT,
Programmer of VLAN/
VxLANs)

1. User requests a new VM, supplying parameters like amount of RAM and the network (vlan/l2 segment) they want to be on
2. Openstack Controller magically selects a Nova Node to deploy the VM on.
2. Simultaneously Openstack Controller tells Neutron Node about a new VM, the Nova node and desired VLAN
3. Neutron decides if the network already exists. If no, a new L3 gateway is created on the Neutron Node.
4. Neutron builds L2 config on the Nova device
5. If deployed to do so, Neutron will speak, via the ML2 Driver, to a hardware switch to provision a VLAN or VxLAN (along with or instead of step 4)
6. Local ML2 plugin translates OpenStack config into device specific config

Openstack Network Design Options

VLANs

- Most common
- Most fragile

EVPN-VxLAN

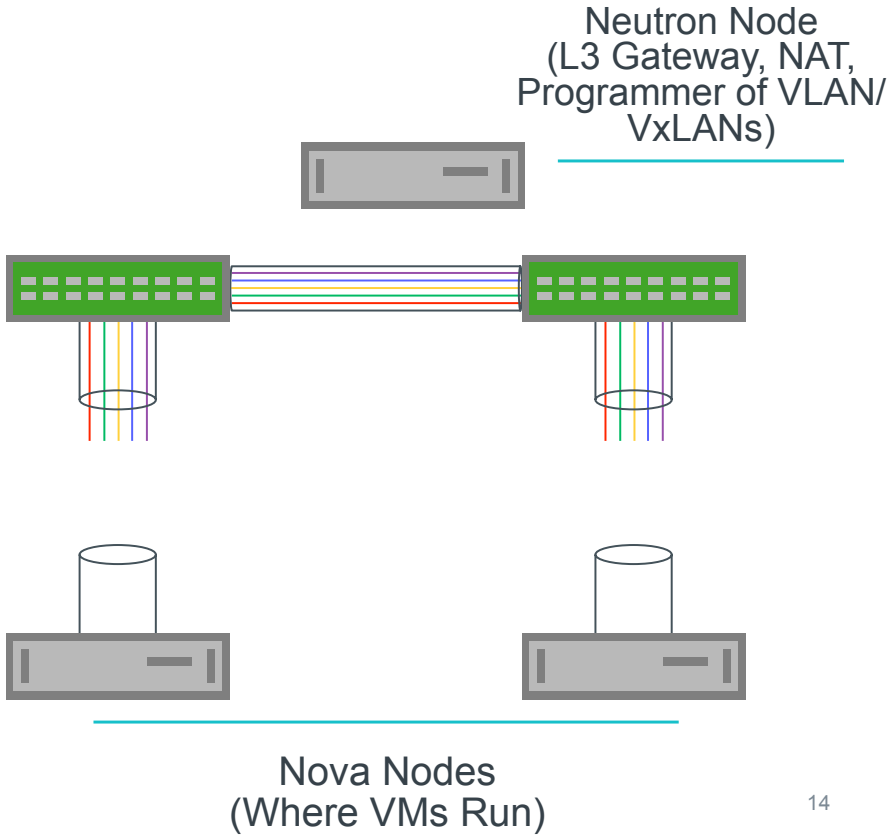
- Network centric
- Scalable, resilient

VxLAN on Servers

- Most scalable
- Simplest network
- More complex servers

Openstack Networking: Preprovisioned VLANs

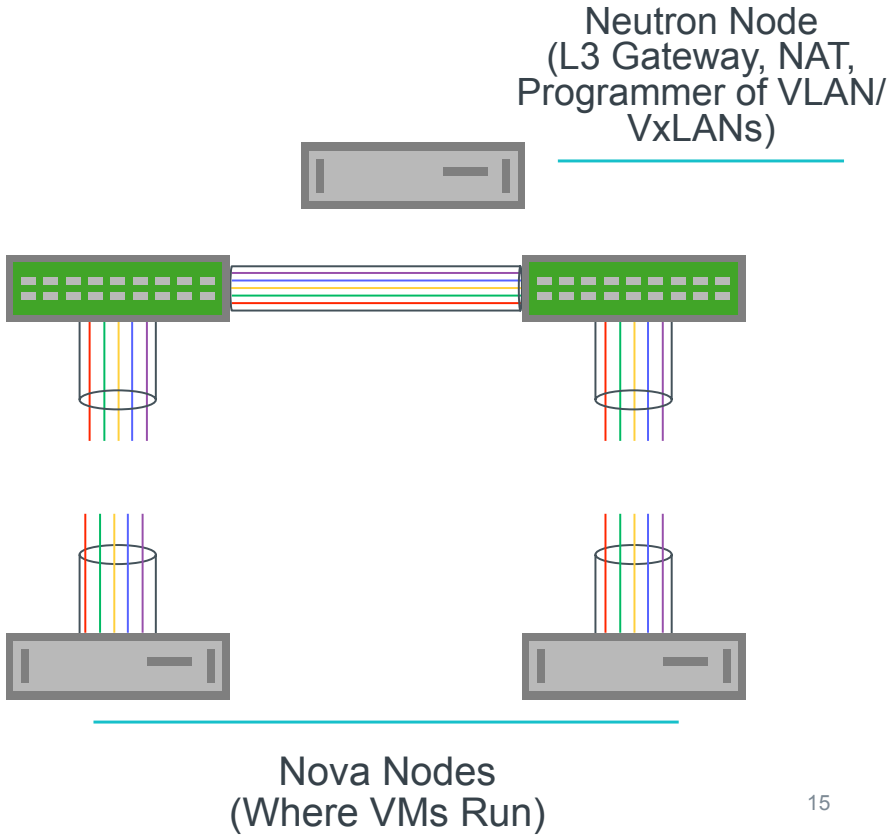
Network trunks all VLANs



Openstack Networking: Preprovisioned VLANs

Network trunks all VLANs

Servers *may* trunk all VLANs

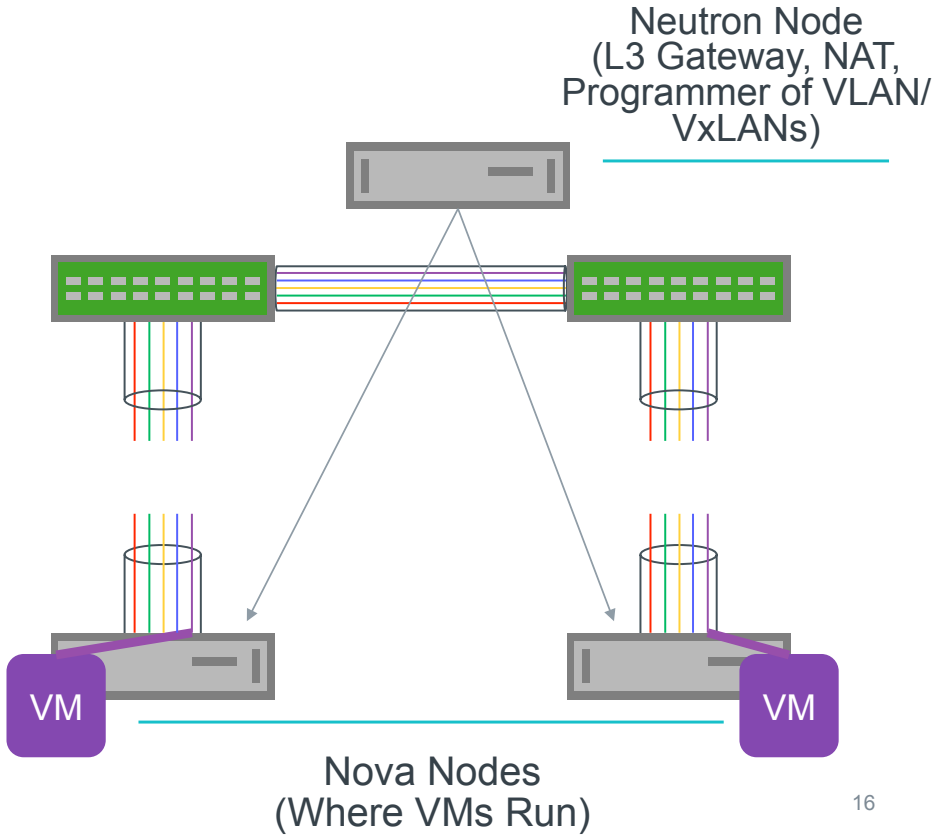


Openstack Networking: Preprovisioned VLANs

Network trunks all VLANs

Servers *may* trunk all VLANs

New server creation only links physical trunk to VM



Openstack Networking: Preprovisioned VLANs

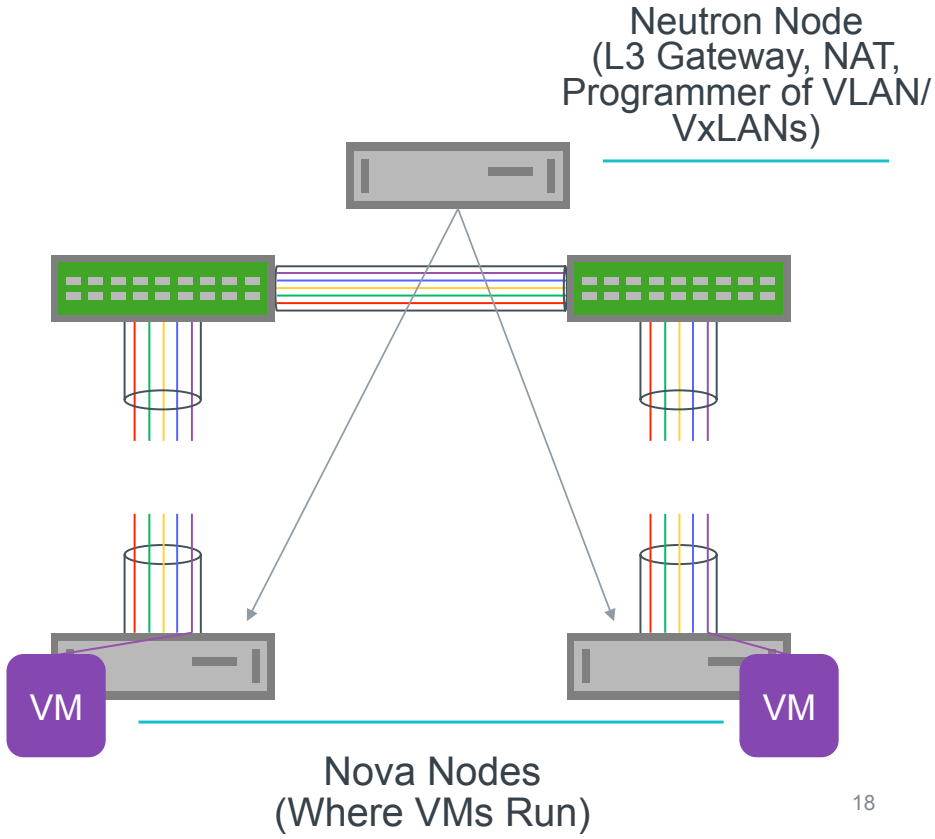
Openstack Networking: Preprovisioned VLANs

Pros:

- **Easiest deployment**
- **Physical network is static**
- **No ML2**

Cons:

- **Limited scale**
- **Very large blast radius**



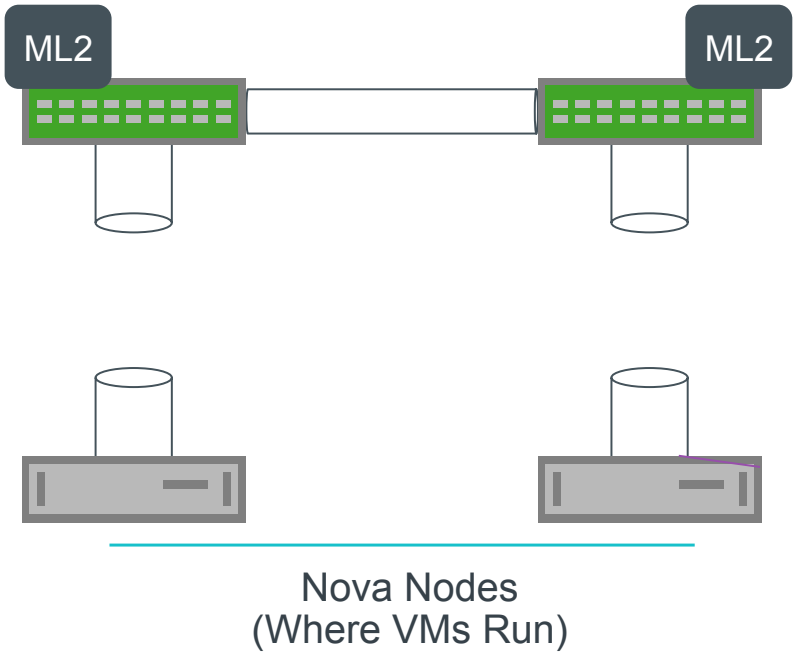
Variation on a Theme: ML2 Provisioned VLANs

Nothing pre-configured

Empty trunks on network and compute

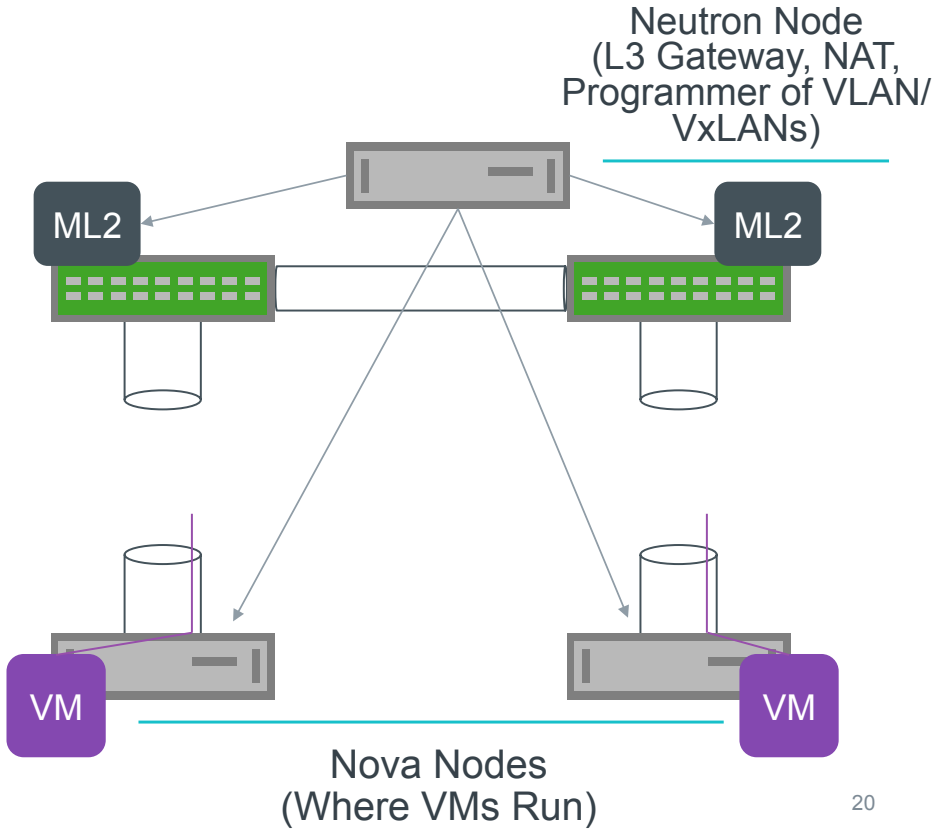
ML2 Agent running on all switches

- Including Core/Spines



Variation on a Theme: ML2 Provisioned VLANs

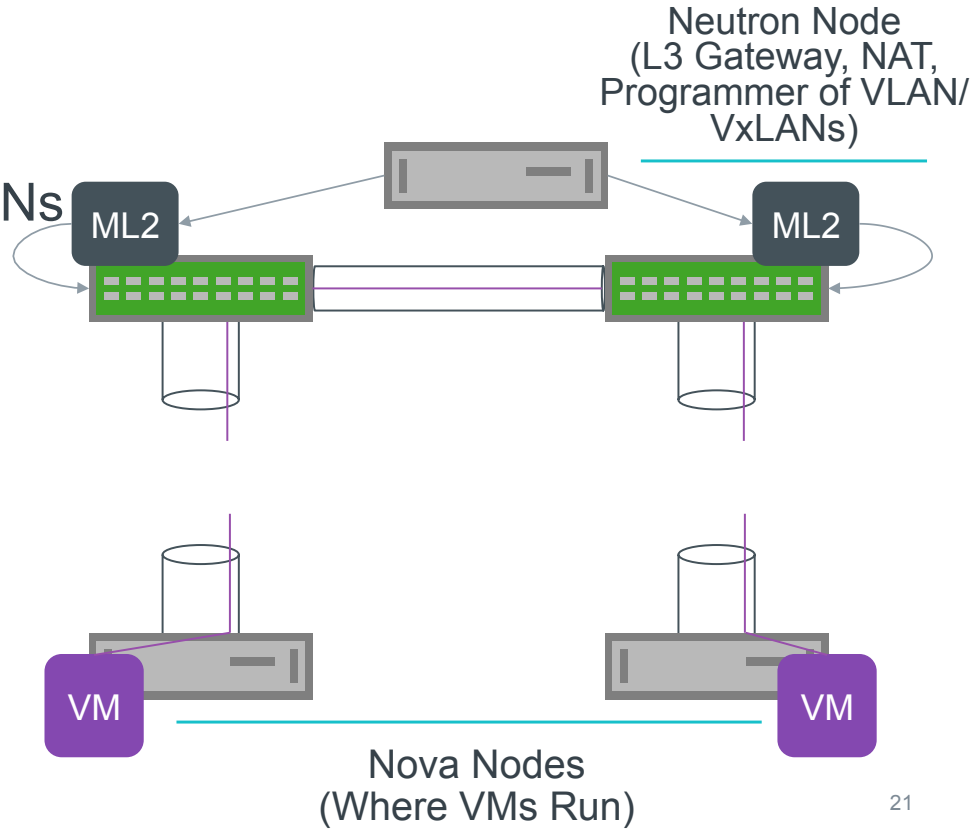
Neutron provisions VLANs on switches via ML2 agent



Variation on a Theme: ML2 Provisioned VLANs

Neutron provisions VLANs on switches via ML2 agent

ML2 Agent provisions switch VLANs



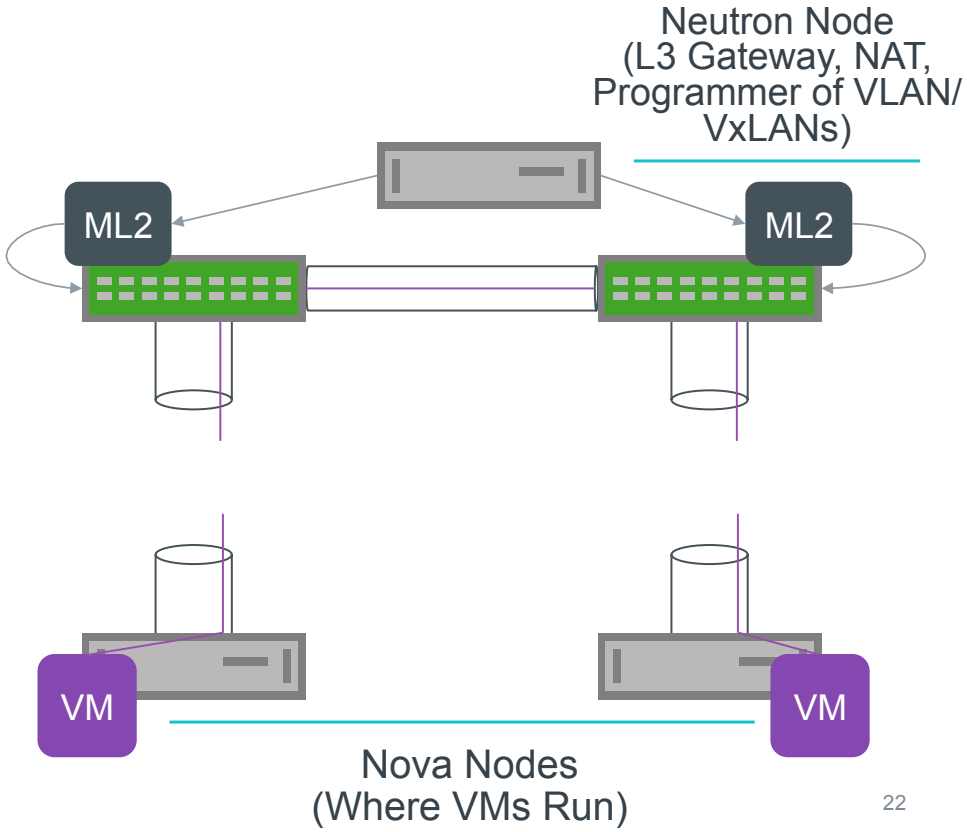
Variation on a Theme: ML2 Provisioned VLANs

Pros:

- Simple server networking
- Slightly more scalable*

Cons:

- Requires ML2 on switches
- Still limited by L2 scale
- Likely to still have large blast radius as environment grows



Sidebar: Openstack Agents and ML2 on Switches

Openstack is designed for “clouds”

Everything is ephemeral

- It can die at any time and no one should care
- This includes networking
- **No one actually accepts this fact**

There may be no “config” for ML2 state

A reloaded switch may lose all provisioned state

- Depends on vendor implementation

Lost state requires all rack VMs to be destroyed and recreated

Sidebar: Openstack Agents and ML2 on Switches

VxLAN-EVPN for Better L2

VxLAN provides L2 over L3

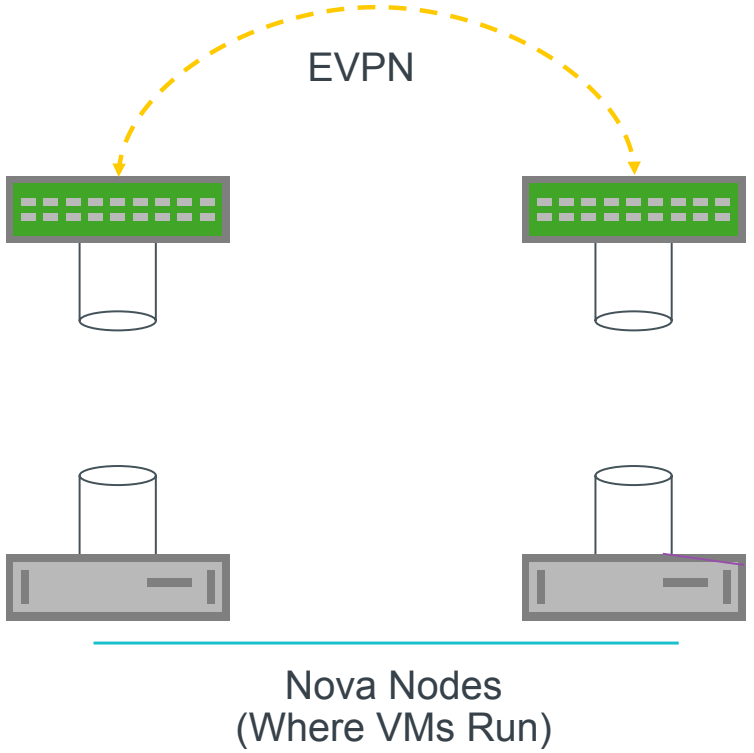
EVPN provides VxLAN control plane (Where MACs live)

EVPN configured from TOR to TOR

Pre-Provision VxLAN Tunnels

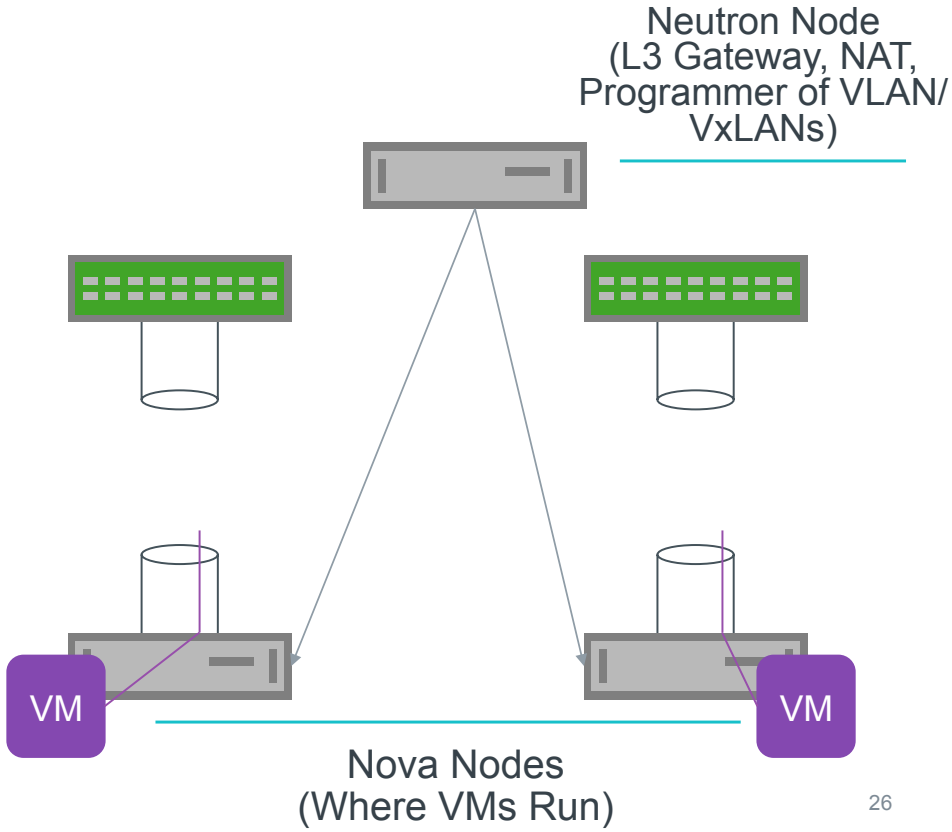
- Scale improvement since MACs are not pushed local a host exists

Openstack doesn't care about EVPN



VxLAN-EVPN for Better L2

Neutron provisions VLAN on host

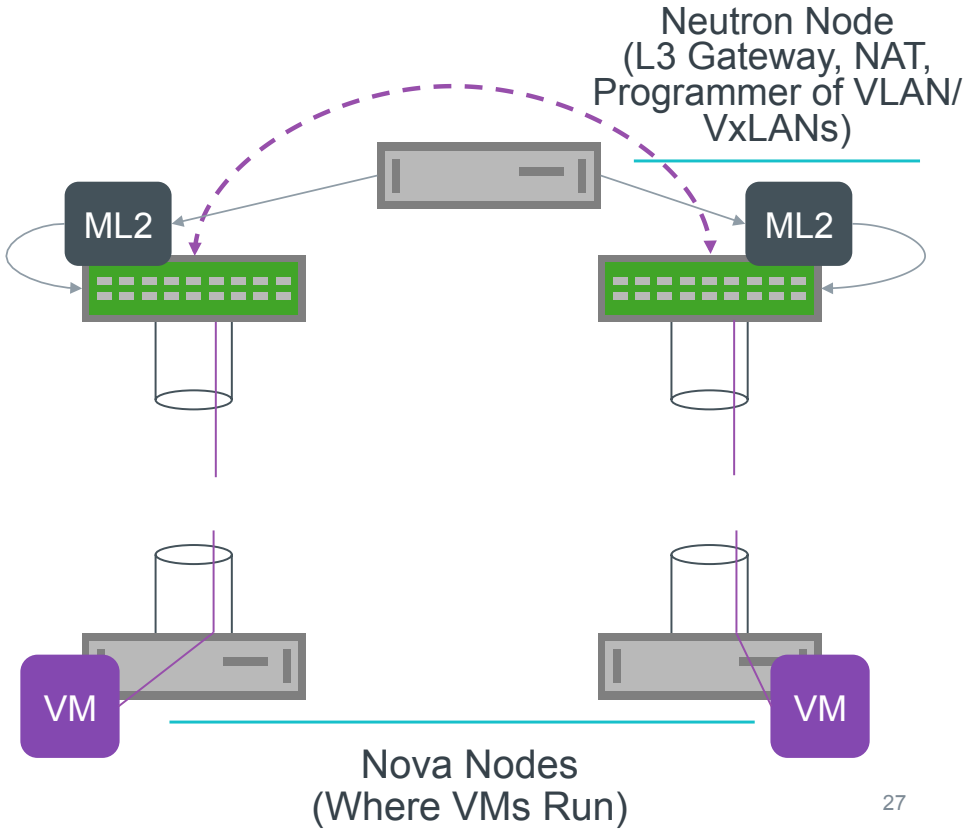


VxLAN-EVPN for Better L2

Neutron provisions VLAN on host

Neuron also provisions VLAN on network via ML2

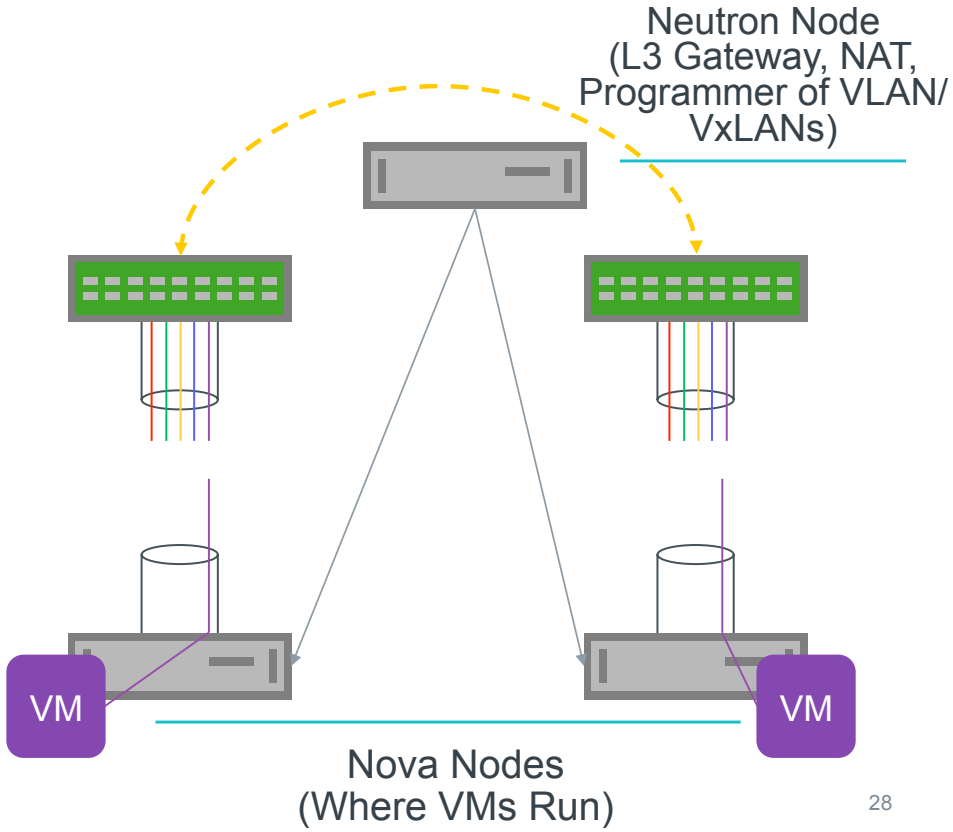
Switch pre-provisioning maps VLAN to VxLAN



VxLAN-EVPN for Better L2

Alternative deployment:

- Pre provision switch VLANs
- Neutron only deploys server VLANs



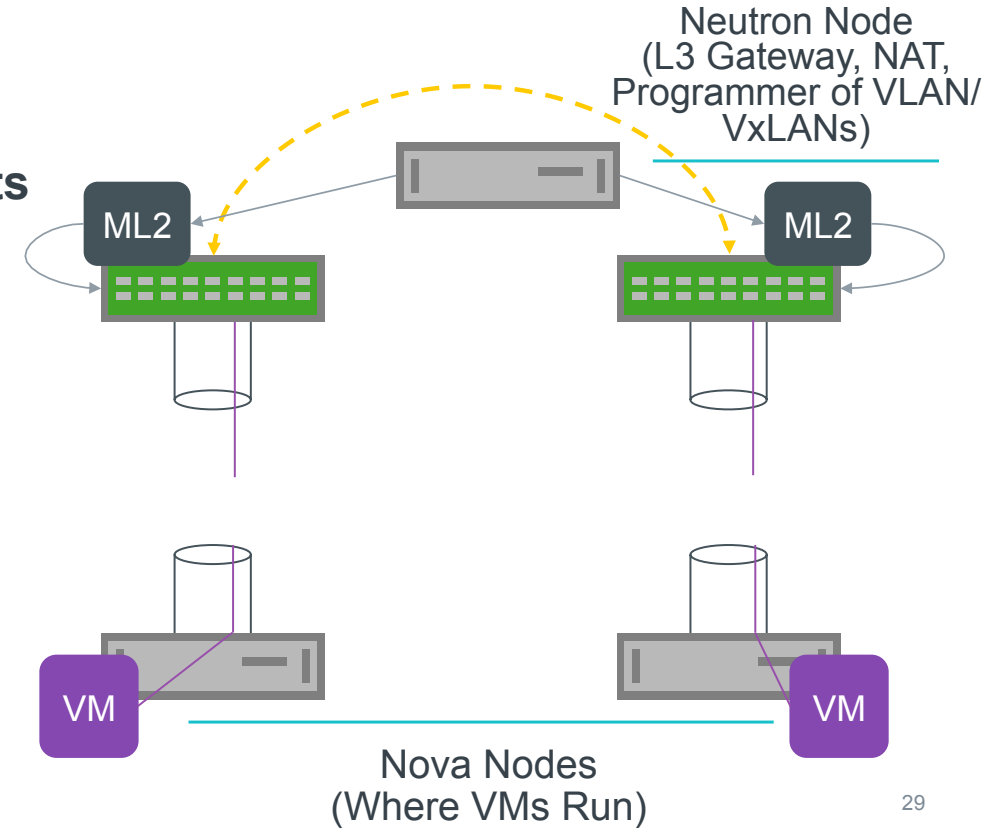
Server Networking: Pre-provisioned EVPN

Pros:

- **Allows for L3 underlay**
- **Easily scales 100-1000 tenants**
- **Preprovisioning is easy**

Cons:

- **May require ML2 on switches**
- **Network still involved**



Server Networking: Pre-provisioned EVPN

Scalable Openstack: Server based VxLAN

Host connects to TOR on routed port

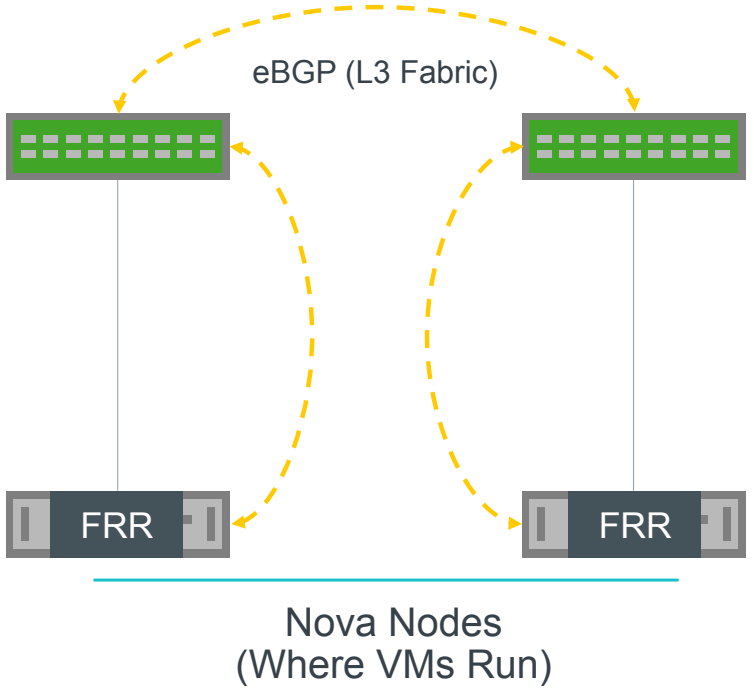
Host runs Free Range Routing (FOSS routing suite)

Host and TOR run eBGP unnumbered

- Dual attach does not require mLAG

Server advertises /32 loopback into the network

No relationship between Openstack and BGP



Scalable Openstack: Server based VxLAN

Neutron programs VxLAN tunnel from host to host

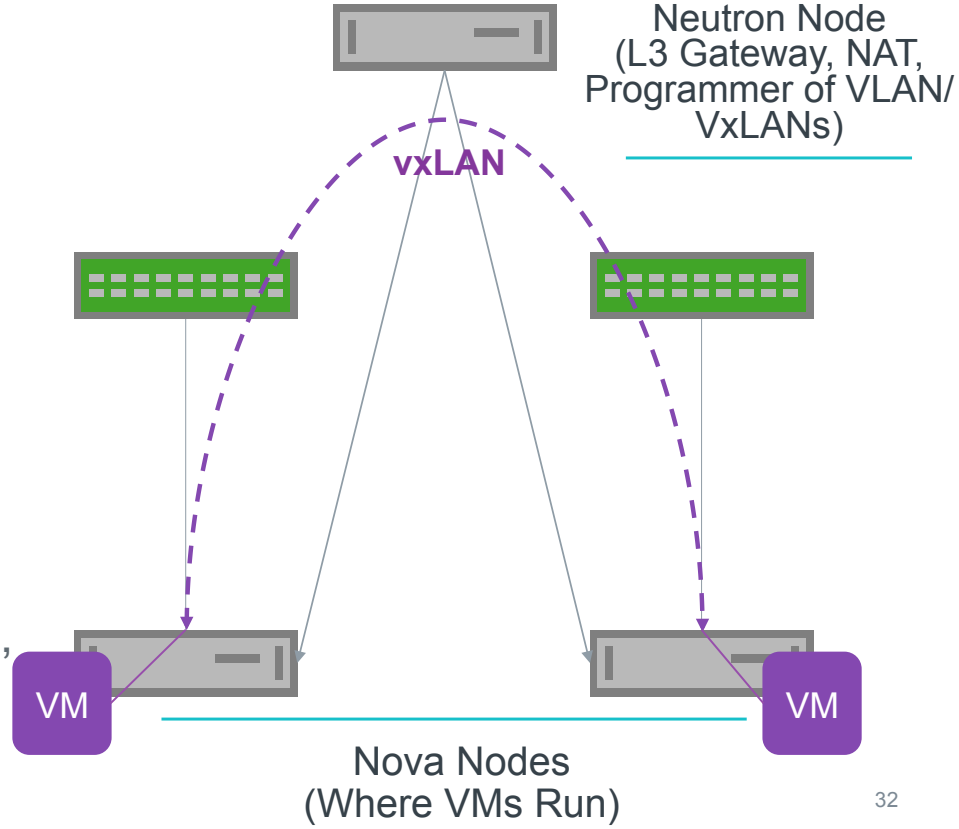
Server loopback interface is the Tunnel Endpoint (VTEP)

Host only sends encapsulated VxLAN traffic into the network

Switches only do basic L3 routing

Openstack links VxLAN to VM

- VM still only has normal ethernet, no VM based VxLAN or VLAN tag



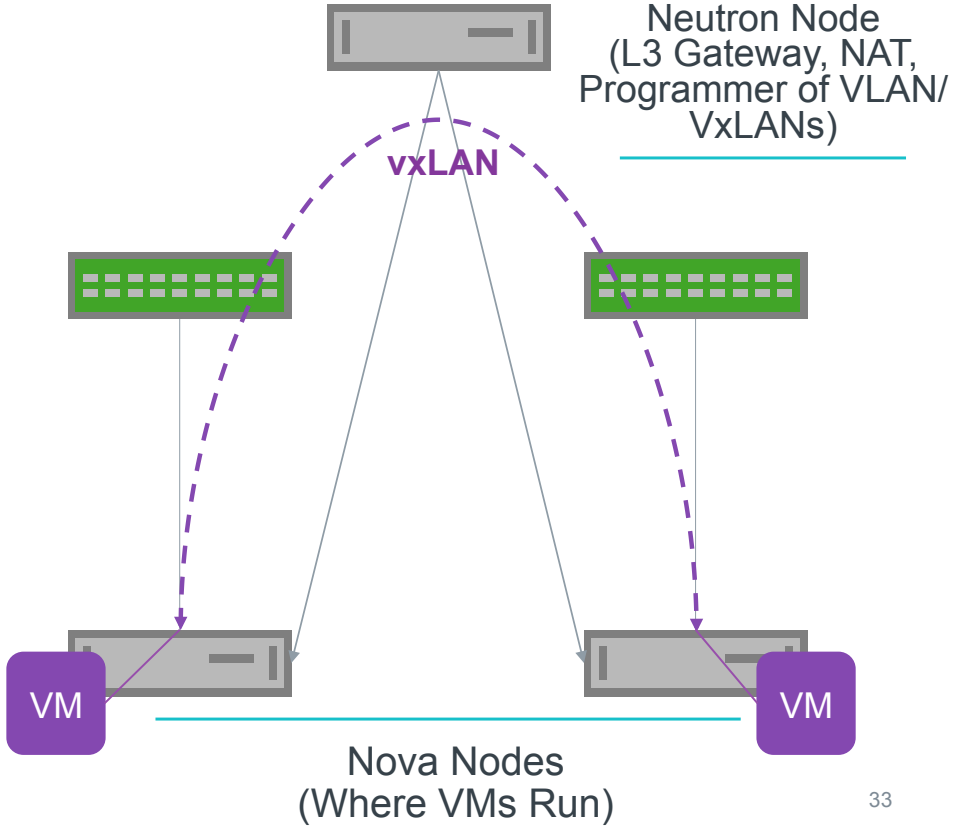
Scalable Openstack: Server based VxLAN

Pros:

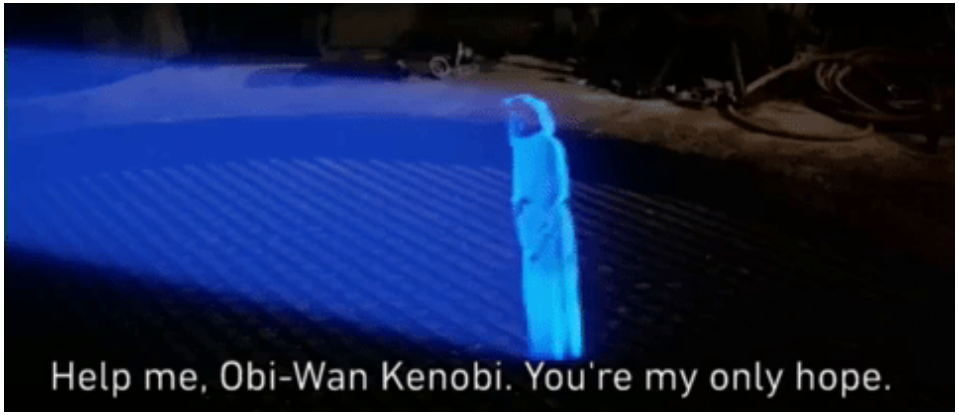
- **Operationally easy, no mLAG**
  Plug and play servers, no IPAM
  0 packet loss network changes
- **Extremely scalable**

Cons:

- **CPU performance hit if NICs don't support VxLAN**
- **Requires FRR on servers**
- **Ironic still requires ML2**



What Should I Do?



What Should I Do?

How many tenants?

- 1-50: L2 everywhere, pre-provisioned will be easiest
- 50-1000: consider pre-provisioning VxLAN-EVPN
- >1000: dynamic (ML2, server VxLAN) options are required to scale

Do you want the network to be programmed by Openstack?

- Yes: ML2 is acceptable
- No: Pre-provision or use server-server VxLAN

VLANs or VxLANs?

- Always prefer VxLANs
- Network hardware needs VxLAN support
- Server NICs need VxLAN offload

A Final Note about L3

All this is about L2 connectivity

L3 is a different plugin (Layer 3 Plugin)

- Less network vendor support for L3 vs L2 plugin

L3 usually requires NAT

- Most Network Hardware L3 plugins don’t support NAT functionality

Other services often required (FWaaS, LBaaS)

- Easy to scale out with FRR on hosts

Thank you!

Visit us at cumulusnetworks.com or follow us @cumulusnetworks