

Proxmox Cisco VLAN LACP SN3 Config

Devices

- Mikrotik CCR1009 acting as core router
 - `http://10.69.7.13` for the web interface, login admin
 - Also available at `10.70.90.1` as the gateway
 - Manages DHCP leases, VLAN routing
- Cisco Nexus 5000 acting as a fiber switch
 - `ssh -oHostKeyAlgorithms=+ssh-rsa admin@10.70.95.10` to SSH
- Dell R640 box 1
 - Service tag `12F11Q2`
 - Reports one of the power supplies is unpowered
 - 384GB ECC DDR4 2666MHz RAM, 2x Xeon Gold 6152 22C/44T CPUs, 8x 1TB Samsung 870 SATA SSDs configured as RAID-6
 - Plugged with 10G fiber into ports 1/7 and 1/8 of the Nexus 5000 with LACP 802.3ad LAG configured
 - hostname `nycmesh-713-r640-01`
 - iDRAC `https://10.70.90.64` with username root
 - Proxmox WebUI `https://10.70.90.195:8006`
 - SSH with user root
- Dell R640 box 2
 - Service tag `G60FXQ2`
 - hostname `nycmesh-713-r640-02`
 - 384GB ECC DDR4 2666MHz RAM, 2x Xeon Gold 6152 22C/44T CPUs, 8x 1TB Samsung 870 SATA SSDs (NOTE: 2 configured as 1TB RAID-1 and 6 configured as 6TB RAID-0. We should fix this and reimage Proxmox ASAP)
 - Plugged with 10G fiber into ports 1/5 and 1/6 of the Nexus 5000 with LACP 802.3ad LAG configured
 - iDRAC `https://10.70.90.65` with username root
 - Proxmox WebUI `https://10.70.90.196:8006`
 - SSH with user root

iDRAC setup

- Thankfully had DHCP enabled, so it showed up in the Mikrotik DHCP leases, hostnames were `idrac- <servicetag>`
- Went to Maintenance, System Update, Manual Update, Location Type HTTPS, use default address, and then Check for Update. It will load some Update Details at the bottom which can then be selected and installed. The Lifecycle log will show updates of the process
 - NOTE: Don't use automatic updates, we can just manually trigger updates when they are useful
 - NOTE: SUP029 warning may indicate that all the firmware is already up to date and of a matching version to what's trying to be updated
 - <https://www.dell.com/support/kbdoc/en-us/000134013/dell-powerededge-update-the-firmware-of-single-system-components-remotely-using-the-idrac>
- Went to iDRAC settings and configured the DNS name as `nycmesh-713-r640- <id>- idrac- <servicetag>`
- Disabled iDRAC IPV6
- Enabled iDRAC NTP for 10.10.10.123 and changed timezone to America/New_York
- Went to Maintenance/Job Queue and System Event Log and cleared both
- no BIOS settings to change
- Go to Storage/Virtual Disks and confirm that RAID-0 is not used (it is not redundant). Delete RAID-0 virtual disks and reformat with RAID-6 if possible. Only four disks may be added to RAID-6 at creation, but more can be added in a subsequent step. Cancel initialization (the option might not pop up if there are pending operations), then Edit Disk Capacity and add the remaining drives. Edit the Cache policy, use the (safer) Write-through caching, which means the disk is successfully written to before confirming the write, as opposed to write-back caching where it makes it into cache first and disk later

Cisco Switch Setup

- Goals
 - Create a LAG/LACP/802.3ad port-channel channel-group for each pair of interfaces going to each server
 - Make the ports trunk ports, meaning all traffic is expected to be tagged with the destination VLAN. Any untagged traffic will be ignored
 - Label the parts of the switch configuration
- First step: connect to the switch using `ssh -oHostKeyAlgorithms=+ssh-rsa admin@10.70.95.10`, using the older SSH host key to avoid the error `no matching host key type found. Their offer: ssh-rsa` <https://askubuntu.com/questions/836048/ssh-returns-no-matching-host-key-type-found-their-offer-ssh-dss>
- Then run `configure terminal` to get into edit mode

- `show run` or `show running-config` will show the full switch configuration. `show run interface` will show just the part of the configuration for the interfaces, both physical and LAG/LACP
- `show port-channel summary` will show already-configured LAG/LACP setups. Same with `show port-channel usage`. This can help determine which ID number to pick for the new port-channel
- The following can be pasted in. It will create a port-channel with ID 1, set a text description, set it to trunk mode (all traffic expected to be tagged, versus access mode where traffic is expected to be untagged) and allows it access only to VLAN32. It then duplicates the setup for port-channel ID 2

```
interface port-channel1
  description nycmesh-713-r640-01 johnb
  switchport mode trunk
  switchport trunk allowed vlan 32

interface port-channel2
  description nycmesh-713-r640-02 johnb
  switchport mode trunk
  switchport trunk allowed vlan 32
```

- Then configure the switch ethernet ports for the first server. Set the description, set it to trunk mode and VLAN 32 same as the LACP, and then add it to the channel-group with ID 2 with active mode, indicating 802.3ad LACP

```
interface Ethernet1/5
  description nycmesh-713-r640-02 wilnil johnb
  switchport mode trunk
  switchport trunk allowed vlan 32
  channel-group 2 mode active

interface Ethernet1/6
  description nycmesh-713-r640-02 wilnil johnb
  switchport mode trunk
  switchport trunk allowed vlan 32
  channel-group 2 mode active
```

- Then configure the switch ethernet ports for the second server. Note the shorthand used to configure two ethernet ports at once, `int eth 1/7 -8`

```
interface Ethernet1/7-8
  description nycmesh-713-r640-01 wilnil johnb
```

```
switchport mode trunk
switchport trunk allowed vlan 32
channel-group 1 mode active
```

- Use `show lacp neighbor` to look to see if a connection was made. If there is nothing connected at the other end, the Partner System ID will probably be all 0s.

```
nycmesh-sn3-n5k(config-if-range) # show lacp neighbor
Flags:  S - Device is sending Slow LACPDUs F - Device is sending Fast LACPDUs
        A - Device is in Active mode      P - Device is in Passive mode

port-channel1 neighbors
Partner's information

Port      Partner
System ID Partner
Eth1/7    0, 0- 0- 0- 0- 0
          0x0      0      SP

          LACP Partner Partner
          Port Priority Oper Key Port State
          0          0x0      0x0

Partner's information

Port      Partner
System ID Partner
Eth1/8    0, 0- 0- 0- 0- 0
          0x0      0      SP

          LACP Partner Partner
          Port Priority Oper Key Port State
          0          0x0      0x0
```

- Once the other end is configured, the MAC address should show up, as well as the flag showing Active mode

```
port-channel1 neighbors
Partner's information

Port      Partner
System ID Partner
Eth1/7    65535, e4- 43- 4b- 18- 25- f00x1
          1488      SA

          LACP Partner Partner
          Port Priority Oper Key Port State
```

	255	0xf	0x3d
Partner's information			
	Partner	Partner	Partner
Port	System ID	Port Number	Age
Eth1/8	65535, e4-43-4b-18-25-f00x2		1488
	LACP Partner	Partner	Partner
	Port Priority	Oper Key	Port State
	255	0xf	0x3d

- Save the configuration to permanent memory so it will persist between reboots with `copy running-config startup-config`
- Other useful commands
 - Get rid of unused port-channels with `no interface port-channel 5`
 - Trying to set `channel-group 1 mode active` may result in an error `command failed: port not compatible [port allowed VLAN list]` which means the allowed VLAN list between the port-channel and the ethernet interface does not match. Get them to match with `switchport trunk allowed vlan 32` and then rerun the command and it should work
 - `no shutdown` apparently will bring a port back online if it were previously disabled for some reason
 - Get a port or set of ports back to its default configuration with `default int eth 1/5-8`
 - View a port's counters and other info with `show int eth 1/5`.
 - Clear the port's counters with `clear count int eth 1/5`
- A variety of different commands can be used to check status on things as well

```
nycmesh-sn3-n5k(config)# show int eth1/5-8 brief
```

Ethernet Interface	VLAN	Type	Mode	Status	Reason	Speed	Port Ch #
Eth1/5	1	eth	trunk	down	suspended(no LACP PDUs)	10G(D)	2
Eth1/6	1	eth	trunk	down	suspended(no LACP PDUs)	10G(D)	2
Eth1/7	1	eth	trunk	up	none	10G(D)	1
Eth1/8	1	eth	trunk	up	none	10G(D)	1

```
nycmesh-sn3-n5k(config)#
```

```
nycmesh-sn3-n5k(config)# show lacp port-channel
port-channel1
```

System Mac=8c- 60- 4f- 50- 45- fc
Local System Identifier=0x8000, 8c- 60- 4f- 50- 45- fc
Admin key=0x8000
Operational key=0x8000
Partner System Identifier=0xffff, e4- 43- 4b- 18- 25- f0
Operational key=0xf
Max delay=0
Aggregate or individual=1
Member Port List=7- 8

port- channel2
System Mac=8c- 60- 4f- 50- 45- fc
Local System Identifier=0x8000, 8c- 60- 4f- 50- 45- fc
Admin key=0x1
Operational key=0x1
Partner System Identifier=0x0, 0- 0- 0- 0- 0- 0
Operational key=0x0
Max delay=0
Aggregate or individual=0
Member Port List=5- 6

nycmesh- sn3- n5k# show lacp counters

Port	LACPDUs		Marker		Marker Response		LACPDUs	
	Sent	Recv	Sent	Recv	Sent	Recv	Pkts	Err

port- channel1								
Ethernet1/7	1245	122	0	0	0	0	0	
Ethernet1/8	1230	90	0	0	0	0	0	
port- channel2								
Ethernet1/5	150	6	0	0	0	0	0	
Ethernet1/6	150	6	0	0	0	0	0	
port- channel16								
Ethernet1/27	1567680	1567441	0	0	0	0	0	
Ethernet1/28	1523391	1523177	0	0	0	0	0	

- Some resources

- Cisco PDF for port-channel configuration with examples, details, etc. Very useful
https://www.cisco.com/c/en/us/td/docs/switches/datacenter/nexus9000/sw/7-x/interfaces/configuration/guide/b_Cisco_Nexus_9000_Series_NX-OS_Interfaces_Configuration_Guide_7x/b_Cisco_Nexus_9000_Series_NX-OS_Interfaces_Configuration_Guide_7x_chapter_0111.pdf
- https://www.cisco.com/en/US/docs/switches/datacenter/nexus5000/sw/configuration/nxos/Cisco_Nexus_5000_Series_NX-OS_Software_Configuration_Guide_chapter9.html has info on how to set up a VLAN in trunk mode, which I think is what we want
- select multiple interfaces
https://www.cisco.com/c/en/us/td/docs/ios/interface/configuration/guide/ir_ifrange.html
- clear interface counters <https://community.cisco.com/t5/switching/how-to-clear-the-interfaces-counter-in-cisco-switches/td-p/2709137>
- briefly describes cisco lacp active vs passive vs on
https://www.grandmetric.com/knowledge-base/design_and_configure/how-to-configure-lacp-on-cisco/
- cisco command reference for different LACP configuration options, I didn't need this
https://www.cisco.com/c/en/us/td/docs/optical/cpt/r9_3/command/reference/cpt93_cr/cpt93_cr_chapter_01000.html
- similar but not exact web version of the PDF for LACP cisco, has the general gist of what I needed
https://www.cisco.com/c/en/us/td/docs/ios/12_2sb/feature/guide/gigeth.html
- Apparently I need to run `no shutdown` and that resets the up/down state of ports
<https://networkengineering.stackexchange.com/questions/56548/what-does-no-shutdown-command-do-cisco>
- https://www.cisco.com/c/en/us/td/docs/switches/datacenter/nexus9000/sw/6-x/interfaces/configuration/guide/b_Cisco_Nexus_9000_Series_NX-OS_Interfaces_Configuration_Guide/b_Cisco_Nexus_9000_Series_NX-OS_Interfaces_Configuration_Guide_chapter_0110.html another guide on how to set up LACP, to some extent

Proxmox/Debian Setup

- Goals
 - The host has four 10G network adapters, two of which are connected to the Cisco switch. The two connected adapters should run in an 802.3ad LACP LAG for redundant + aggregated link speed.
 - The host should also be able to talk using tagged traffic on VLAN 32, which is where the other Proxmox host Jon communicates. This will be configured as a trunk port, meaning all traffic that is not tagged will be ignored
- Configuration to some degree can occur in the Proxmox GUI, but if the switch is already configured for VLANs and LACP, a connection may not be possible. Instead, use the virtual console of the iDRAC to type and configure via the CLI of Proxmox

- All configuration occurs within `/etc/network/interfaces`
- After modifying, the settings can be applied by restarting the service `systemctl restart networking`
- The following config is pulled from one of the hosts. Note that the `iface` names might vary from host to host, so while one host may be `enl1np0` another host may just be `eno0`

```
root@nycmesh-713-r640-01: ~# cat /etc/network/interfaces
```

```
auto lo
iface lo inet loopback

iface eno1np0 inet manual
iface eno2np1 inet manual
iface eno3np2 inet manual
iface eno4np3 inet manual

auto bond0
iface bond0 inet manual
    bond-slaves eno1np0 eno2np1
    bond-miimon 100
    bond-mode 802.3ad
    bond-xmit-hash-policy layer2+3

iface bond0.32 inet manual

auto vmbr0v32
iface vmbr0v32 inet static
    address 10.70.90.195
    gateway 10.70.90.1
    bridge-ports bond0.32
    bridge-stp off
    bridge-fd 0

iface idrac inet manual

source /etc/network/interfaces.d/*
```

- The following version is marked up to indicate the purpose of each line

```
# declares and configures the loopback interface
auto lo
iface lo inet loopback
```



```

# declares and configures the "raw" adapter interfaces
iface eno1np0 inet manual
iface eno2np1 inet manual
iface eno3np2 inet manual
iface eno4np3 inet manual

# declares the LACP 802.3ad LAG port-channel bond
auto bond0
iface bond0 inet manual
    bond-slaves eno1np0 eno2np1 # defines which adapters are part of the LACP bond
    bond-miimon 100 # default, defines the link monitoring frequency
    bond-mode 802.3ad # defines the type of bond (round-robin, active-backup, etc) and
802.3ad is the LACP standard
    bond-xmit-hash-policy layer2+3 # defines the method by which traffic will be
transmitted across the interfaces of the bond

# declares an interface configured for VLAN 32 on the LACP bond interface
iface bond0.32 inet manual

# declares a Linux Bridge
auto vmbr0v32
iface vmbr0v32 inet static
    address 10.70.90.195 # defines the static IP (and subnet if desired, via /24, /28,
etc.)
    gateway 10.70.90.1 # defines the gateway through which traffic will pass
    bridge-ports bond0.32 # defines the interface the Bridge is connected to. In this case
VLAN 32 on the LACP bond is desired, and an interface was set up a few lines above for this
    bridge-stp off # turns off spanning tree protocol for loop prevention
    bridge-fd 0 # default forwarding delay, seen in all tutorials

# defines the interface that the iDRAC operates on. This won't be used
iface idrac inet manual

# Fetches further configuration from files within the following folder
source /etc/network/interfaces.d/*

```

- Resources

- Very helpful Proxmox official guide for VLANs and LACP bonding all on one page at the same time
https://pve.proxmox.com/wiki/Network_Configuration#sysadmin_network_vlan

- Proxmox community sample LACP config, I didn't end up using this but good to know <https://forum.proxmox.com/threads/proxmox-lACP-bonding-and-tp-link-managed-switch.130789/>
- Proxmox another community sample LACP config <https://forum.proxmox.com/threads/lACP-trunk-on-two-ethernet-ports.146386/>
- General debian guide for bonding/LACP <https://wiki.debian.org/Bonding>
- Another community LACP and VLAN guide <https://forum.proxmox.com/threads/setting-up-lACP-bond-with-vlan-trunk-and-bridge.93331/>
- https://wiki.debian.org/NetworkConfiguration#Reinitialize_new_network_setup once a change is made in /etc/network/interfaces, apply it with `systemctl restart networking`
- I left the subnet off the end of the linux /etc/network/interfaces address line, which seems to resolve to a /32 when I run `ip a`. According to <https://networkengineering.stackexchange.com/questions/34587/why-do-i-need-to-specify-a-subnet-mask-on-interface> it just seems to dictate which pings/traffic need to hit the router and which can go directly to another device without a router involved. Since we use a mix of /26, /30/ 24, etc I figure the /32 is safe

Revision #5

Created 3 July 2024 20:39:36 by johnb

Updated 9 August 2024 02:35:43 by johnb