GUIDE

By: Humza Altaf, SONiC Network Engineer

Revision No.	Description	Editor	Date
1.0	Guide for VxLAN BGP EVPN Between SONiC and Cumulus Linux	Humza Altaf	Nov 14, 2023

Table of Contents

Introduction to VxLAN BGP EVPN

In traditional Layer 2 networks, reachability information is distributed in the data plane through flooding. With EVPN-VXLAN networks, this activity moves to the control plane.

EVPN is an extension to Border Gateway Protocol (BGP) that allows the network to carry endpoint reachability information such as Layer 2 MAC addresses and Layer 3 IP addresses. This control plane technology uses MP-BGP for MAC and IP address endpoint distribution, where MAC addresses are treated as routes.

Because MAC learning is now handled in the control plane, it avoids the flooding typical with layer 2 networks. EVPN can support different data-plane encapsulation technologies between EVPN-VXLAN-enabled switches. With EVPN-VXLAN architectures, VXLAN provides the overlay data-plane encapsulation.

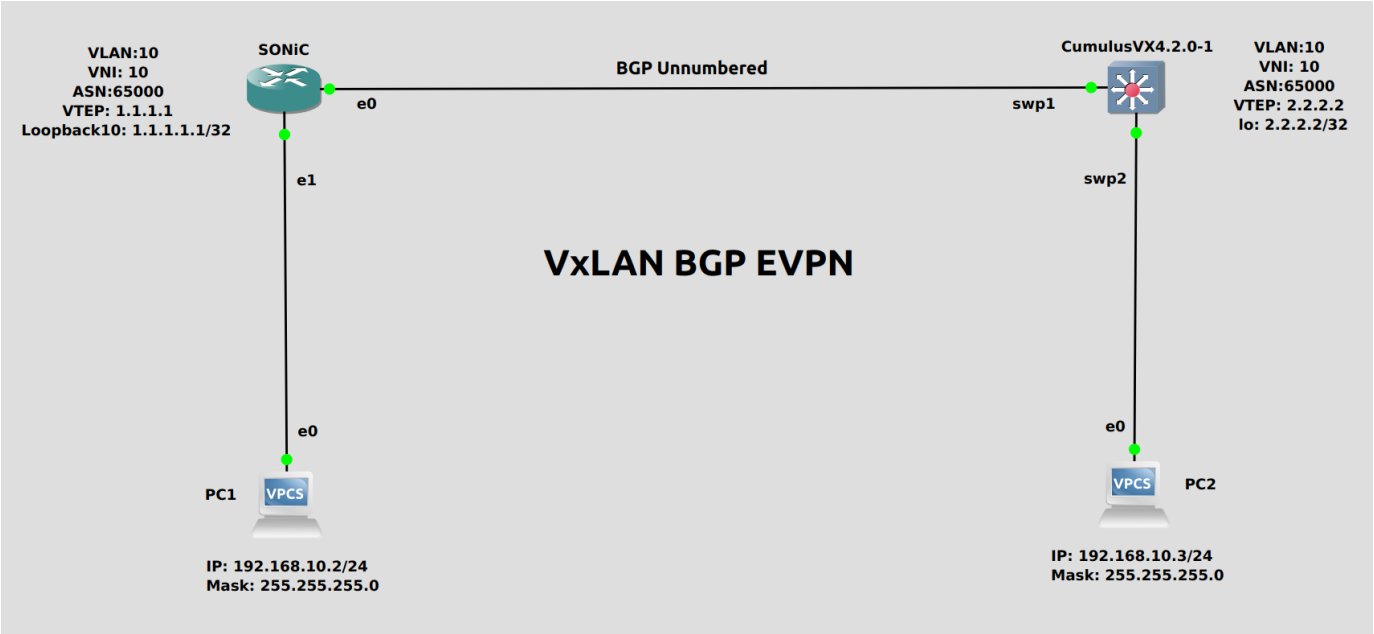
Network overlays are created by encapsulating traffic and tunneling it over a physical network. The VXLAN tunneling protocol encapsulates Layer 2 Ethernet frames in Layer 3 UDP packets, enabling Layer 2 virtual networks or subnets that can span the underlying physical Layer 3 network. The device that performs VXLAN encapsulation and decapsulation is called a VXLAN tunnel endpoint (VTEP). EVPN enables devices acting as VTEPs to exchange reachability information with each other about their endpoints.

In a VXLAN overlay network, each Layer 2 subnet or segment is uniquely identified by a virtual network identifier (VNI). A VNI segments traffic the same way that a VLAN ID segments traffic - endpoints within the same virtual network can communicate directly with each other, while endpoints in different virtual networks require a device that supports inter-VNI (inter-VXLAN) routing.

Term	Meaning
VNI (VxLAN Network Identifier)	Virtual extension of VLAN over IP network
VTEP (VXLAN Tunnel End Point)	An entity that originates and/or terminates VXLAN tunnels which is specified by a source IP address.
NVO	Network Virtualization Overlay

Network Topology

The GNS3 network topology consists of two switches SONiC (202205) and Cumulus (4.2). Both switches are kept in the same ASN which is 65000. No IP address is assigned between switches. PC1 and PC2 are assigned untagged VLAN 10.



Port Mapping

GNS3	SONiC
Ethernet 0	Ethernet 0
Ethernet 1	Ethernet 4
Ethernet 2	Ethernet 8
Ethernet 3	Ethernet 12

Configurations

For the above topology, all hosts and switches are first configured before sending traffic. First, switch Cumulus is configured and then SONiC. Command Reference guide is also available on GitHub for SONiC, whose link is given here.

Follow these steps to configure Cumulus.

Step 1

Assign IP address to Loopback interface by using the following command given below:

- `net add loopback lo ip address 2.2.2.2/32`

Create VLAN, VNI, VxLAN tunnel and make “swp2” port as access port by using the following commands given below in the screenshot.

```
cumulus@cumulus:mgmt:~$ net add interface swp2 br
    breakout  :  Split an interface into multiple interface via a breakout cable
    bridge    :  a layer2 bridge
cumulus@cumulus:mgmt:~$ net add interface swp2 bridge access 10
cumulus@cumulus:mgmt:~$ net add vxlan vni10 vxlan id 10
cumulus@cumulus:mgmt:~$ net add bridge bridge ports vni10
cumulus@cumulus:mgmt:~$ net add bridge bridge vids 10
cumulus@cumulus:mgmt:~$ net add vxlan vni10 bridge access 10
cumulus@cumulus:mgmt:~$ net add bridge bridge ports swp2
cumulus@cumulus:mgmt:~$ net add loopback lo vxlan local-tunnelip 2.2.2.2
```

Before going to vtysh mode, use “ sudo -i ” command to go into root. Make all the interfaces swp1, spw2 and lo “up” by using the following commands given below in the screenshot.

```
cumulus@cumulus:mgmt:~$ vtysh
% Can't open configuration file /etc/frr/vtysh.conf due to 'Permission denied'.
Exiting: failed to connect to any daemons.
Hint: if this seems wrong, try running me as a privileged user!
cumulus@cumulus:mgmt:~$ sudo -i
root@cumulus:mgmt:~# vtysh

Hello, this is FRRouting (version 7.4+cl4u1).
Copyright 1996-2005 Kunihiro Ishiguro, et al.

cumulus# configure
cumulus(config)# interface sw
swp1  swp2  swp3  swp4  swp5  swp6
cumulus(config)# interface swp1
cumulus(config-if)# no shutdown
cumulus(config-if)# interface swp2
cumulus(config-if)# no shutdown
cumulus(config-if)# interface lo
cumulus(config-if)# no shutdown
```

Step 2

In the topology, no IP address is assigned on the interfaces between the switches. So, configure BGP Neighbor Session with unnumbered and announce the network by using the following commands given below in the screenshot.

```
cumulus(config)# router
router     router-id
cumulus(config)# router bgp 65000
cumulus(config-router)# bgp rou
route-map        route-reflector  router-id
cumulus(config-router)# bgp router-id 2.2.2.2
cumulus(config-router)# neighbor swp1 interface remote-as internal
cumulus(config-router)# address-family ipv4
cumulus(config-router-af)# network 2.2.2.2/32
cumulus(config-router-af)# exit
```

Now announce L2VPN EVPN routes by using the following commands given below in the screenshot and save configurations by using “ wr ” command.

```
cumulus(config-router)# address-family l2vpn evpn
cumulus(config-router-af)# neighbor swp1 activate
cumulus(config-router-af)# adve
advertise             advertise-default-gw  advertise-svi-ip
advertise-all-vni     advertise-pip
cumulus(config-router-af)# advertise-all-vni
cumulus(config-router-af)# end
cumulus# wr
Note: this version of vtysh never writes vtysh.conf
Building Configuration...
Integrated configuration saved to /etc/frr/frr.conf
[OK]
cumulus#
```

Follow these steps to configure the SONiC switch.

Step 3

By default, all interfaces are routed (L3) and IP is assigned to them. To check the status of IP addresses, use the following command given below:

- `show ip interfaces`

Step 3 (Continued)

```
admin@sonic:~$ show ip interfaces
Interface    Master    IPv4 address/mask    Admin/Oper    BGP Neighbor    Neighbor IP
-----------  --------  -------------------  ------------  --------------  -------------
Ethernet0              10.0.0.0/31          up/up         ARISTA01T2      10.0.0.1
Ethernet4              10.0.0.2/31          up/up         ARISTA02T2      10.0.0.3
```

Remove the IP addresses and assign IP address to Loopback by using the following commands given below in the screenshot:

```
admin@sonic:~$ sudo config interface ip remove Ethernet0 10.0.0.0/31
admin@sonic:~$ sudo config interface ip remove Ethernet4 10.0.0.2/31
admin@sonic:~$ sudo config interface ip add Loopback10 1.1.1.1/32
```

Note: It is better practice to save configurations after executing two or three commands by using “sudo config save -y” command.

Step 4

Create VLAN and make Ethernet 4 as access port by using the following commands given below in the screen-shot.

```
admin@sonic:~$ sudo config vlan add 10
admin@sonic:~$ sudo config vlan member add -u 10 Ethernet4
```

Create VxLAN, VNI and tunnel by using the following commands given below in the screenshot.

```
admin@sonic:~$ sudo config vxlan add vtep 1.1.1.1
admin@sonic:~$ sudo config vxlan evpn_nvo add nvo vtep
admin@sonic:~$ sudo config vxlan map add vtep 10 10
admin@sonic:~$ sudo config save -y
Running command: /usr/local/bin/sonic-cfggen -d --print-data > /etc/sonic/config_db.json
```

Step 5

In SONiC, by default BGP is running with ASN 65100. First remove that instance and then configure BGP with unnumbered and announce the network by using the following commands given below in the screenshot.

```
admin@sonic:~$ vtysh

Hello, this is FRRouting (version 8.2.2).
Copyright 1996-2005 Kunihiro Ishiguro, et al.

sonic# configure
sonic(config)# no router bgp 65100
sonic(config)# router bgp 65000
sonic(config-router)# neighbor Ethernet0 interface remote 65000
sonic(config-router)# neighbor Ethernet0 interface remote-as internal
sonic(config-router)# bgp router-id 1.1.1.1
sonic(config-router)# address-family ipv4
sonic(config-router-af)# network 1.1.1.1/32
sonic(config-router-af)# exit
```

Now announce L2VPN EVPN routes by using the following commands given below in the screenshot and save configurations by using “ wr ” command.

```
sonic(config-router)# address-family l2vpn evpn
sonic(config-router-af)# neighbor Ethernet0 activate
sonic(config-router-af)# advertise-all-vni
sonic(config-router-af)# end
sonic# wr
Note: this version of vtysh never writes vtysh.conf
Building Configuration...
Configuration saved to /etc/frr/zebra.conf
Configuration saved to /etc/frr/bgpd.conf
Configuration saved to /etc/frr/staticd.conf
```

Step 6

Check the status in Cumulus Linux switch by using the following commands given below in the screenshot.

```
root@cumulus:mgmt:~# net show evpn vni 10
VNI: 10
 Type: L2
 Tenant VRF: default
 VxLAN interface: vni10
 VxLAN ifIndex: 10
 Local VTEP IP: 2.2.2.2
 Mcast group: 0.0.0.0
 Remote VTEPs for this VNI:
  1.1.1.1 flood: HER
 Number of MACs (local and remote) known for this VNI: 2
 Number of ARPs (IPv4 and IPv6, local and remote) known for this VNI: 0
 Advertise-gw-macip: No
```

Step 6 (Continued)

```
root@cumulus:mgmt:~# net show bridge vlan

Interface  VLAN  Flags                  VNI
---------  ----  ---------------------  ---
swp2         10  PVID, Egress Untagged
vni10        10  PVID, Egress Untagged  10

root@cumulus:mgmt:~#
```

```
root@cumulus:mgmt:~# net show evpn mac vni all
VNI 10 #MACs (local and remote) 2

Flags: N=sync-neighs, I=local-inactive, P=peer-active, X=peer-proxy
MAC               Type   Flags Intf/Remote ES/VTEP            VLAN  Seq #'s
00:50:79:66:68:00 remote       1.1.1.1                              0/0
00:50:79:66:68:01 local        swp2                           10    0/0
```

```
root@cumulus:mgmt:~# net show bgp l2vpn evpn route
BGP table version is 2, local router ID is 2.2.2.2
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete
EVPN type-1 prefix: [4]:[ESI]:[EthTag]:[IPlen]:[VTEP-IP]
EVPN type-2 prefix: [2]:[EthTag]:[MAClen]:[MAC]:[IPlen]:[IP]
EVPN type-3 prefix: [3]:[EthTag]:[IPlen]:[OrigIP]
EVPN type-4 prefix: [4]:[ESI]:[IPlen]:[OrigIP]
EVPN type-5 prefix: [5]:[EthTag]:[IPlen]:[IP]

   Network          Next Hop            Metric LocPrf Weight Path
                    Extended Community
Route Distinguisher: 1.1.1.1:2
*>i[2]:[0]:[48]:[00:50:79:66:68:00]
                    1.1.1.1                        100      0 i
                    RT:65000:10 ET:8
*>i[3]:[0]:[32]:[1.1.1.1]
                    1.1.1.1                        100      0 i
                    RT:65000:10 ET:8
Route Distinguisher: 2.2.2.2:2
*> [2]:[0]:[48]:[00:50:79:66:68:01]
                    2.2.2.2                             32768 i
                    ET:8 RT:65000:10
*> [3]:[0]:[32]:[2.2.2.2]
                    2.2.2.2                             32768 i
                    ET:8 RT:65000:10

Displayed 4 prefixes (4 paths)
```

Step 6 (Continued)

```
root@cumulus:mgmt:~# net show bgp summary
show bgp ipv4 unicast summary
=============================
BGP router identifier 2.2.2.2, local AS number 65000 vrf-id 0
BGP table version 4
RIB entries 3, using 576 bytes of memory
Peers 1, using 21 KiB of memory

Neighbor        V         AS   MsgRcvd   MsgSent   TblVer  InQ OutQ  Up/Down State/PfxRcd
sonic(swp1)     4      65000       362       360        0    0    0 00:14:42            1

Total number of neighbors 1


show bgp ipv6 unicast summary
=============================
% No BGP neighbors found


show bgp l2vpn evpn summary
===========================
BGP router identifier 2.2.2.2, local AS number 65000 vrf-id 0
BGP table version 0
RIB entries 3, using 576 bytes of memory
Peers 1, using 21 KiB of memory

Neighbor        V         AS   MsgRcvd   MsgSent   TblVer  InQ OutQ  Up/Down State/PfxRcd
```

```
root@cumulus:mgmt:~# net show route
show ip route
=============
Codes: K - kernel route, C - connected, S - static, R - RIP,
       O - OSPF, I - IS-IS, B - BGP, E - EIGRP, N - NHRP,
       T - Table, v - VNC, V - VNC-Direct, A - Babel, D - SHARP,
       F - PBR, f - OpenFabric,
       > - selected route, * - FIB route, q - queued route, r - rejected route

B>* 1.1.1.1/32 [200/0] via fe80::e55:b6ff:fe42:0, swp1, weight 1, 00:14:18
C>* 2.2.2.2/32 is directly connected, lo, 00:44:04



show ipv6 route
===============
Codes: K - kernel route, C - connected, S - static, R - RIPng,
       O - OSPFv3, I - IS-IS, B - BGP, N - NHRP, T - Table,
       v - VNC, V - VNC-Direct, A - Babel, D - SHARP, F - PBR,
       f - OpenFabric,
       > - selected route, * - FIB route, q - queued route, r - rejected route

C * fe80::/64 is directly connected, swp1, 00:37:39
C>* fe80::/64 is directly connected, bridge, 00:44:03
```

Step 7

Now check the status in SONiC switch by using the following commands given below in the screenshot.

```
admin@sonic:~$ show vlan brief
+-----------+-------------+-----------+---------------+------------+----------------------+
|   VLAN ID | IP Address  | Ports     | Port Tagging  | Proxy ARP  | DHCP Helper Address  |
+===========+=============+===========+===============+============+======================+
|        10 |             | Ethernet4 | untagged      | disabled   |                      |
+-----------+-------------+-----------+---------------+------------+----------------------+
admin@sonic:~$
```

```
sonic# show bgp summary

IPv4 Unicast Summary (VRF default):
BGP router identifier 1.1.1.1, local AS number 65000 vrf-id 0
BGP table version 8
RIB entries 3, using 552 bytes of memory
Peers 1, using 723 KiB of memory

Neighbor        V         AS   MsgRcvd   MsgSent   TblVer  InQ OutQ  Up/Down State/PfxRcd   PfxSnt Desc
Ethernet0       4      65000        67        73        0    0    0 00:00:16            1        1 N/A

Total number of neighbors 1

L2VPN EVPN Summary (VRF default):
BGP router identifier 1.1.1.1, local AS number 65000 vrf-id 0
BGP table version 0
RIB entries 3, using 552 bytes of memory
Peers 1, using 723 KiB of memory

Neighbor        V         AS   MsgRcvd   MsgSent   TblVer  InQ OutQ  Up/Down State/PfxRcd   PfxSnt Desc
Ethernet0       4      65000        67        73        0    0    0 00:00:16            1        1 N/A

Total number of neighbors 1
sonic# exit
```

```
admin@sonic:~$ show vxlan vlanvnimap
+--------+-------+
| VLAN   |   VNI |
+========+=======+
| Vlan10 |    10 |
+--------+-------+
Total count : 1
```

```
admin@sonic:~$ show vxlan remotevtep
+---------+---------+-------------------+--------------+
| SIP     | DIP     | Creation Source   | OperStatus   |
+=========+=========+===================+==============+
| 1.1.1.1 | 2.2.2.2 | EVPN              | oper_down    |
+---------+---------+-------------------+--------------+
Total count : 1
```

```
admin@sonic:~$ show vxlan tunnel
vxlan tunnel name    source ip    destination ip    tunnel map name    tunnel map mapping(vni -> vlan)
-------------------  -----------  ----------------  -----------------  --------------------------------
vtep                 1.1.1.1                        map_10_Vlan10      10 -> Vlan10
```

Step 7 (Continued)

```
sonic# show bgp ipv4 summary

IPv4 Unicast Summary (VRF default):
BGP router identifier 1.1.1.1, local AS number 65000 vrf-id 0
BGP table version 8
RIB entries 3, using 552 bytes of memory
Peers 1, using 723 KiB of memory

Neighbor        V         AS   MsgRcvd   MsgSent   TblVer  InQ OutQ  Up/Down State/PfxRcd   PfxSnt Desc
Ethernet0       4      65000       126       132        0    0    0 00:03:13            1        1 N/A

Total number of neighbors 1
sonic# show bgp l2vpn evpn summary
BGP router identifier 1.1.1.1, local AS number 65000 vrf-id 0
BGP table version 0
RIB entries 3, using 552 bytes of memory
Peers 1, using 723 KiB of memory

Neighbor        V         AS   MsgRcvd   MsgSent   TblVer  InQ OutQ  Up/Down State/PfxRcd   PfxSnt Desc
Ethernet0       4      65000       129       135        0    0    0 00:03:22            1        1 N/A

Total number of neighbors 1
```

```
sonic# sh ip route
Codes: K - kernel route, C - connected, S - static, R - RIP,
       O - OSPF, I - IS-IS, B - BGP, E - EIGRP, N - NHRP,
       T - Table, v - VNC, V - VNC-Direct, A - Babel, F - PBR,
       f - OpenFabric,
       > - selected route, * - FIB route, q - queued, r - rejected, b - backup
       t - trapped, o - offload failure

C>* 1.1.1.1/32 is directly connected, Loopback10, 00:08:21
B>* 2.2.2.2/32 [200/0] via fe80::e91:88ff:fea2:1, Ethernet0, weight 1, 00:03:28
C>* 10.0.0.4/31 is directly connected, Ethernet8, 00:09:54
C>* 10.0.0.6/31 is directly connected, Ethernet12, 00:09:54
C>* 10.0.0.8/31 is directly connected, Ethernet16, 00:09:54
C>* 10.0.0.10/31 is directly connected, Ethernet20, 00:09:54
C>* 10.0.0.12/31 is directly connected, Ethernet24, 00:09:54
C>* 10.0.0.14/31 is directly connected, Ethernet28, 00:09:54
C>* 10.0.0.16/31 is directly connected, Ethernet32, 00:09:54
```

```
sonic# show evpn vni detail
VNI: 10
 Type: L2
 Tenant VRF: default
 VxLAN interface: vtep-10
 VxLAN ifIndex: 50
 SVI interface: Vlan10
 SVI ifIndex: 49
 Local VTEP IP: 1.1.1.1
 Mcast group: 0.0.0.0
 Remote VTEPs for this VNI:
  2.2.2.2 flood: HER
 Number of MACs (local and remote) known for this VNI: 0
 Number of ARPs (IPv4 and IPv6, local and remote) known for this VNI: 0
 Advertise-gw-macip: No
 Advertise-svi-macip: No
```

Step 7 (Continued)

```
sonic# sh bgp l2vpn evpn route type multicast
BGP table version is 1, local router ID is 1.1.1.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete
EVPN type-1 prefix: [1]:[EthTag]:[ESI]:[IPlen]:[VTEP-IP]
EVPN type-2 prefix: [2]:[EthTag]:[MAClen]:[MAC]:[IPlen]:[IP]
EVPN type-3 prefix: [3]:[EthTag]:[IPlen]:[OrigIP]
EVPN type-4 prefix: [4]:[ESI]:[IPlen]:[OrigIP]
EVPN type-5 prefix: [5]:[EthTag]:[IPlen]:[IP]

   Network          Next Hop            Metric LocPrf Weight Path
                    Extended Community
Route Distinguisher: 1.1.1.1:2
*> [3]:[0]:[32]:[1.1.1.1]
                    1.1.1.1                            32768 i
                    ET:8 RT:65000:10
Route Distinguisher: 2.2.2.2:2
*>i[3]:[0]:[32]:[2.2.2.2]
                    2.2.2.2                 100      0 i
                    RT:65000:10 ET:8

Displayed 2 prefixes (2 paths) (of requested type)
sonic# exit
```

Step 8

Assign IP addresses to hosts PC1 and PC2 by using command given below:

- `ip <ip_addr> <subnet mask>`

```
PC1> ip 192.168.10.2/24 255.255.255.0
Checking for duplicate address...
PC1 : 192.168.10.2 255.255.255.0
```

After assigning IP addresses, check the status of IP address using command given below:

- `show ip`

```
PC1> sh ip

NAME        : PC1[1]
IP/MASK     : 192.168.10.2/24
GATEWAY     : 255.255.255.0
DNS         :
MAC         : 00:50:79:66:68:00
LPORT       : 10022
RHOST:PORT  : 127.0.0.1:10023
MTU         : 1500
```

Result

PC1 to PC2

Once the switches and hosts are configured, communication becomes possible among hosts in the same VLAN. As is evident from the provided figure below, PC1 is receiving a response from PC2, as both of them belong to the same VLAN. Furthermore, the TTL (Time-to-Live) value stays at 64 and remains unchanged. Therefore, the VxLAN BGP EVPN has been successfully configured.

```
PC1> ping 192.168.10.3

84 bytes from 192.168.10.3 icmp_seq=1 ttl=64 time=3.484 ms
84 bytes from 192.168.10.3 icmp_seq=2 ttl=64 time=5.693 ms
84 bytes from 192.168.10.3 icmp_seq=3 ttl=64 time=5.831 ms
84 bytes from 192.168.10.3 icmp_seq=4 ttl=64 time=5.711 ms
84 bytes from 192.168.10.3 icmp_seq=5 ttl=64 time=6.198 ms
```

PC2 to PC1

```
PC2> ping 192.168.10.2

84 bytes from 192.168.10.2 icmp_seq=1 ttl=64 time=5.760 ms
84 bytes from 192.168.10.2 icmp_seq=2 ttl=64 time=5.997 ms
84 bytes from 192.168.10.2 icmp_seq=3 ttl=64 time=2.362 ms
84 bytes from 192.168.10.2 icmp_seq=4 ttl=64 time=2.548 ms
84 bytes from 192.168.10.2 icmp_seq=5 ttl=64 time=4.595 ms
```

References

Github Reference 1

Github Reference 2

NVIDIA Reference 1

NVIDIA Reference 2

NVIDIA Reference 3

We connect ideas, people, and technology.