|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | **Windows Server 2016 Converged NIC Step-by-Step** |
|  |  |  | The Instruction below provide the detailed steps to deploy and diagnose WS2016 Converged NIC using RoCEv2.  The Instructions marked in **GREEN** are for BASIC single adapter scenarios only. This is the case where only the minimal set of operations and resources are desired.  The instructions in **BLUE** are the recommended configuration Datacenter deployment with multiple RDMA Host vNICs for maximum performance and availability.  Outline:   1. Test Basic connectivity 2. Configure VLAN 3. Configure QoS 4. Test RDMA (Mode1 – Native) 5. Create vSwitch 6. Configure VLAN 7. Configure QoS 8. Test RDMA (Mode2 – Host vNIC)   Terminology:   * TOR – Top of Rack Switch * vSwitch – Hyper-V Virtual Switch * hNIC – Host vNIC – Virtual NIC from vSwitch * pNIC – Physical NIC – In this example both Physical NICs are named Test-40G-\*. |
|  |  |  | **1.0 Test Connectivity** |
| **1** |  |  |  |
| **2** |  |  | # First we ensure the pNIC over which we’ll create a vSwitch has basic connectivity to the destination through the Fabric. This encompasses reachability via L3 (i.e., IP level) as well as L2 VLAN.  **Get-NetAdapter -Name "Test-40G-1" | ft -AutoSize**    Name InterfaceDescription ifIndex Status MacAddress LinkSpeed  ---- -------------------- ------- ------ ---------- ---------  Test-40G-1 Mellanox ConnectX-3 Pro Ethernet Ada... 11 Up E4-1D-2D-07-43-D0 40 Gbps      **Get-NetIPAddress -InterfaceAlias "Test-40G-1"**  or  **Get-NetIPAddress -InterfaceAlias "TEST-40G-1" | Where-Object {$\_.AddressFamily -eq "IPv4"} | fl InterfaceAlias,IPAddress**  .  IPAddress : 192.168.1.3  InterfaceIndex : 11  InterfaceAlias : Test-40G-1  AddressFamily : IPv4  Type : Unicast  PrefixLength : 24  .  . |
| **3** |  |  | # In the case of a Teamed pNIC configuration, also ensure that the other NIC has a valid IP. Note that here we use a separate subnet (xxx.xxx.**2**.xxx vs xxx.xxx.**1**.xxx) to facilitate sending from this adapter to the destination. Otherwise, if we place both pNICs on the same subnet the Windows TCPIP stack will load balance among the interfaces, and simple validation becomes more complicated.  **Get-NetAdapter -Name "Test-40G-2" | ft -AutoSize**  Name InterfaceDescription ifIndex Status MacAddress LinkSpeed  ---- -------------------- ------- ------ ---------- ---------  TEST-40G-2 Mellanox ConnectX-3 Pro Ethernet A...#2 13 Up E4-1D-2D-07-40-70 40 Gbps  **Get-NetIPAddress -InterfaceAlias "Test-40G-2"**  or  **Get-NetIPAddress -InterfaceAlias "Test-40G-2" | Where-Object {$\_.AddressFamily -eq "IPv4"} | fl InterfaceAlias,IPAddress**  .  IPAddress : 192.168.2.3  InterfaceIndex : 13  InterfaceAlias : TEST-40G-2  AddressFamily : IPv4  Type : Unicast  PrefixLength : 24  .  . |
| **4** |  |  | # Ensure bi-directional reachability (ping from source to destination and vice-versa on both systems). The Test-NetConnection command is used, but ping will suffice as well.    **Test-NetConnection 192.168.1.5**  ComputerName : 192.168.1.5  RemoteAddress : 192.168.1.5  InterfaceAlias : Test-40G-1  SourceAddress : 192.168.1.3  PingSucceeded : False  PingReplyDetails (RTT) : 0 ms  # In some cases, it may be necessary to edit firewall settings. This should be done with security in mind, according to your organizational requirements. In this example, we use the simplistic “disable all” firewall setting.  **Set-NetFirewallProfile -All -Enabled False**  **Test-NetConnection 192.168.1.5**  ComputerName : 192.168.1.5  RemoteAddress : 192.168.1.5  InterfaceAlias : Test-40G-1  SourceAddress : 192.168.1.3  PingSucceeded : True  PingReplyDetails (RTT) : 0 ms |
| **5** |  |  | # Repeat the steps above for all subsequent pNICs which will be under the team.  **Test-NetConnection 192.168.2.5**  ComputerName : 192.168.2.5  RemoteAddress : 192.168.2.5  InterfaceAlias : TEST-40G-2  SourceAddress : 192.168.2.3  PingSucceeded : True  PingReplyDetails (RTT) : 0 ms |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | **2.0 Configure VLANs** |
|  |  |  | Note that in this case the NICs are in ACCESS mode. However when a switch is created later, the VLAN properties are applied at the vSwitch port level. Given a switch will host multiple VLANs, it is necessary for the Physical Switch (ToR) to have it’s port configured in Trunk mode. Consult the switch vendor documentation for instructions. |
| **1** |  |  |  |
| **2** |  |  | # Per standards, the QoS Properties in the Physical NIC act on the 802.1p header embedded within the 802.1Q (VLAN) header.      **Set-NetAdapterAdvancedProperty -Name "Test-40G-1" -RegistryKeyword VlanID -RegistryValue "101"**  **Get-NetAdapterAdvancedProperty -Name "Test-40G-1" | Where-Object {$\_.RegistryKeyword -eq "VlanID"} | ft -AutoSize**    Name DisplayName DisplayValue RegistryKeyword RegistryValue  ---- ----------- ------------ --------------- -------------  TEST-40G-1 VLAN ID 101 VlanID {101}  # IMPORTANT, Ensure the VLANID takes effect independent of adapter implementation.  **Restart-NetAdapter -Name "Test-40G-1"**  # Ensure adapter status is up before proceeding.  **Get-NetAdapter -Name "Test-40G-1" | ft -AutoSize**    Name InterfaceDescription ifIndex Status MacAddress LinkSpeed  ---- -------------------- ------- ------ ---------- ---------  Test-40G-1 Mellanox ConnectX-3 Pro Ethernet Ada... 11 Up E4-1D-2D-07-43-D0 40 Gbps |
| **3** |  |  | **Set-NetAdapterAdvancedProperty -Name "Test-40G-2" -RegistryKeyword VlanID -RegistryValue "102"**  **Get-NetAdapterAdvancedProperty -Name "Test-40G-2" | Where-Object {$\_.RegistryKeyword -eq "VlanID"} | ft -AutoSize**  Name DisplayName DisplayValue RegistryKeyword RegistryValue  ---- ----------- ------------ --------------- -------------  TEST-40G-2 VLAN ID 101 VlanID {102}    # IMPORTANT, Ensure the VLANID takes effect independent of adapter implementation.  **Restart-NetAdapter -Name "Test-40G-2"**  # Ensure adapter status is up before proceeding.  **Get-NetAdapter -Name "Test-40G-1" | ft -AutoSize**    Name InterfaceDescription ifIndex Status MacAddress LinkSpeed  ---- -------------------- ------- ------ ---------- ---------  Test-40G-2 Mellanox ConnectX-3 Pro Ethernet Ada... 11 Up E4-1D-2D-07-43-D1 40 Gbps |
| **4** |  |  | Note that it may take several seconds for the device to be visible on the network. A Test-NetConnection failure or ping failure \_IMMEDIATELY\_ after Restart-NetAdapter is not uncommon, simply try again.    Reconfirm connectivity after applying the VLAN tag to both adapters. If connectivity breaks, then inspect switch VLAN configuration or destination participation in the same VLAN.  Diagnosis: if the VLAN 101 connections work, but the VLAN 102 connections don’t, the problem may be that the switch needs to be configured to allow port traffic on the desired VLAN. You could test check for this by temporarily setting the failing adapters to VLAN 101, and repeating the connectivity test.  **Test-NetConnection 192.168.1.5**  ComputerName : 192.168.1.5  RemoteAddress : 192.168.1.5  InterfaceAlias : Test-40G-1  SourceAddress : 192.168.1.5  PingSucceeded : True  PingReplyDetails (RTT) : 0 ms |
| **5** |  |  | # Review instruction in prior step.  **Test-NetConnection 192.168.2.5**  ComputerName : 192.168.2.5  RemoteAddress : 192.168.2.5  InterfaceAlias : TEST-40G-2  SourceAddress : 192.168.2.3  PingSucceeded : True  PingReplyDetails (RTT) : 0 ms |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | **3.0 Configure QoS** |
| **1** |  |  |  |
| **2** |  |  | # Turn on Data Center Bridging (DCB; optional for iWarp in most cases, but necessary at fabric scale such as cross-rack scenarios.)  **Install-WindowsFeature Data-Center-Bridging**    Success Restart Needed Exit Code Feature Result  ------- -------------- --------- --------------  True No Success {Data Center Bridging} |
| **3** |  |  | # Set the policies for SMB-Direct  **New-NetQosPolicy "SMB" -NetDirectPortMatchCondition 445 -PriorityValue8021Action 3**    Name : SMB  Owner : Group Policy (Machine)  NetworkProfile : All  Precedence : 127  JobObject :  NetDirectPort : 445  PriorityValue : 3 |
| **4** |  |  | # Set policies for other traffic on the interface  **New-NetQosPolicy "DEFAULT" -Default -PriorityValue8021Action 0**    Name : DEFAULT  Owner : Group Policy (Machine)  NetworkProfile : All  Precedence : 127  Template : Default  JobObject :  PriorityValue : 0 |
| **5** |  |  | # Turn on Flow Control for SMB  **Enable-NetQosFlowControl -priority 3**  **Get-NetQosFlowControl**    Priority Enabled PolicySet IfIndex IfAlias  -------- ------- --------- ------- -------  0 False Global  1 False Global  2 False Global  3 True Global  4 False Global  5 False Global  6 False Global  7 False Global |
| **6** |  |  | # If traffic classes other than 3 are enabled for FlowControl, make sure flow control is disabled for them. Under more complex configuration, the other traffic classes may require flow control, however said scenarios are outside the scope of this document.  **Disable-NetQosFlowControl -priority 0,1,2,4,5,6,7**  **Get-NetQosFlowControl**    Priority Enabled PolicySet IfIndex IfAlias  -------- ------- --------- ------- -------  0 False Global  1 False Global  2 False Global  3 True Global  4 False Global  5 False Global  6 False Global  7 False Global |
| **7** |  |  | # Turn this on for the target adapter as well as destination adapters.  **Enable-NetAdapterQos -InterfaceAlias "Test-40G-1"**  **Get-NetAdapterQos -Name "Test-40G-1"**      Name : TEST-40G-1  Enabled : True  Capabilities : Hardware Current  -------- -------  MacSecBypass : NotSupported NotSupported  DcbxSupport : None None  NumTCs(Max/ETS/PFC) : 8/8/8 8/8/8    OperationalTrafficClasses : TC TSA Bandwidth Priorities  -- --- --------- ----------  0 Strict 0-7    OperationalFlowControl : Priority 3 Enabled  OperationalClassifications : Protocol Port/Type Priority  -------- --------- --------  Default 0  NetDirect 445 3 |
| **8** |  |  | **Enable-NetAdapterQos -InterfaceAlias "Test-40G-2"**  **Get-NetAdapterQos -Name "Test-40G-2"**      Name : TEST-40G-2  Enabled : True  Capabilities : Hardware Current  -------- -------  MacSecBypass : NotSupported NotSupported  DcbxSupport : None None  NumTCs(Max/ETS/PFC) : 8/8/8 8/8/8    OperationalTrafficClasses : TC TSA Bandwidth Priorities  -- --- --------- ----------  0 Strict 0-7    OperationalFlowControl : Priority 3 Enabled  OperationalClassifications : Protocol Port/Type Priority  -------- --------- --------  Default 0  NetDirect 445 3 |
| **9** |  |  | # Give SMB Direct (RDMA) 50% of the bandwidth reservation  **New-NetQosTrafficClass "SMB" -priority 3 -bandwidthpercentage 50 -algorithm ETS**    Name Algorithm Bandwidth(%) Priority PolicySet IfIndex IfAlias  ---- --------- ------------ -------- --------- ------- -------  SMB ETS 50 3 Global    **Get-NetQosTrafficClass | ft -AutoSize**    Name Algorithm Bandwidth(%) Priority PolicySet IfIndex IfAlias  ---- --------- ------------ -------- --------- ------- -------  [Default] ETS 50 0-2,4-7 Global  SMB ETS 50 3 Global |
| **10** |  |  | OPTIONAL  # Create two more traffic classes for tenant IP traffic. Omit the “IP1” and “IP2” creation if not desired.  **New-NetQosTrafficClass "IP1" -Priority 1 -bandwidthpercentage 10 -algorithm ETS**      Name Algorithm Bandwidth(%) Priority PolicySet IfIndex IfAlias  ---- --------- ------------ -------- --------- ------- -------  IP1 ETS 10 1 Global        **New-NetQosTrafficClass "IP2" -Priority 2 -bandwidthpercentage 10 -algorithm ETS**      Name Algorithm Bandwidth(%) Priority PolicySet IfIndex IfAlias  ---- --------- ------------ -------- --------- ------- -------  IP2 ETS 10 2 Global        **Get-NetQosTrafficClass | ft -AutoSize**    Name Algorithm Bandwidth(%) Priority PolicySet IfIndex IfAlias  ---- --------- ------------ -------- --------- ------- -------  [Default] ETS 30 0,4-7 Global  SMB ETS 50 3 Global  IP1 ETS 10 1 Global  IP2 ETS 10 2 Global |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | **4.0 Configure Debugger (Optional)** |
| **1** |  |  | # Override the Debugger - by default the attached debugger blocks NetQos  **Set-ItemProperty HKLM:"\SYSTEM\CurrentControlSet\Services\NDIS\Parameters" AllowFlowControlUnderDebugger -type DWORD -Value 1 –Force**  **Get-ItemProperty HKLM:"\SYSTEM\CurrentControlSet\Services\NDIS\Parameters" | ft AllowFlowControlUnderDebugger**    AllowFlowControlUnderDebugger  -----------------------------  1 |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | **5.0 Test RDMA (Mode 1)** |
|  |  |  | This steps ensures the fabric is correctly configured prior to creating a vSwitch and transitioning to RDMA (Mode 2). |
| **1** |  |  |  |
| **2** |  |  | Download the DiskSpd.exe utility and extract into C:\TEST\  <http://tinyurl.com/z68h3rc>  Download the Test-RDMA powershell script to C:\TEST\  <https://github.com/Microsoft/SDN/blob/master/Diagnostics/Test-Rdma.ps1> |
| **3** |  |  | **Get-NetAdapterRdma | ft -AutoSize**  Name InterfaceDescription Enabled  ---- -------------------- -------  TEST-40G-1 Mellanox ConnectX-4 VPI Adapter #2 True  TEST-40G-2 Mellanox ConnectX-4 VPI Adapter True |
| **4** |  |  | # Determine the ifIndex value of your target adapter:  **Get-NetIPConfiguration -InterfaceAlias "TEST\*" | ft InterfaceAlias,InterfaceIndex,IPv4Address**  InterfaceAlias InterfaceIndex IPv4Address  -------------- -------------- -----------  TEST-40G-1 14 {192.168.1.3}  TEST-40G-2 13 {192.168.2.3}  # Pass the ifIndex value to the Test-RDMA.ps1 script, along with the IP address of the remote adapter on the same VLAN:  **C:\TEST\Test-RDMA.PS1 -IfIndex 14 -IsRoCE $true -RemoteIpAddress 192.168.1.5 -PathToDiskspd C:\TEST\Diskspd-v2.0.17\amd64fre\**  VERBOSE: Diskspd.exe found at C:\TEST\Diskspd-v2.0.17\amd64fre\\diskspd.exe  VERBOSE: The adapter TEST-40G-1 is a physical adapter  VERBOSE: Underlying adapter is RoCE. Checking if QoS/DCB/PFC is configured on each physical adapter(s)  VERBOSE: QoS/DCB/PFC configuration is correct.  VERBOSE: RDMA configuration is correct.  VERBOSE: Checking if remote IP address, 192.168.1.5, is reachable.  VERBOSE: Remote IP 192.168.1.5 is reachable.  VERBOSE: Disabling RDMA on adapters that are not part of this test. RDMA will be enabled on them later.  VERBOSE: Testing RDMA traffic now for. Traffic will be sent in a parallel job. Job details:  VERBOSE: 0 RDMA bytes written per second  VERBOSE: 0 RDMA bytes sent per second  VERBOSE: 662979201 RDMA bytes written per second  VERBOSE: 37561021 RDMA bytes sent per second  VERBOSE: 1023098948 RDMA bytes written per second  VERBOSE: 8901349 RDMA bytes sent per second  VERBOSE: Enabling RDMA on adapters that are not part of this test. RDMA was disabled on them prior to sending RDMA traffic.  VERBOSE: RDMA traffic test SUCCESSFUL: RDMA traffic was sent to 192.168.1.5  **Note that if the RDMA traffic fails, for the RoCE case specifically, consult your TOR (Physical Switch) configuration for proper PFC/ETS settings which match the Host side settings. Refer to the QoS section in this document for reference values.** |
| **5** |  |  | # Determine the ifIndex value of your target adapter:  **Get-NetIPConfiguration -InterfaceAlias "TEST\*" | ft InterfaceAlias,InterfaceIndex,IPv4Address**  InterfaceAlias InterfaceIndex IPv4Address  -------------- -------------- -----------  TEST-40G-1 14 {192.168.1.3}  TEST-40G-2 13 {192.168.2.3}  # See index v  **C:\TEST\Test-RDMA.PS1 -IfIndex 13 -IsRoCE $true -RemoteIpAddress 192.168.2.5 -PathToDiskspd C:\TEST\Diskspd-v2.0.17\amd64fre\**  VERBOSE: Diskspd.exe found at C:\TEST\Diskspd-v2.0.17\amd64fre\\diskspd.exe  VERBOSE: The adapter TEST-40G-2 is a physical adapter  VERBOSE: Underlying adapter is RoCE. Checking if QoS/DCB/PFC is configured on each physical adapter(s)  VERBOSE: QoS/DCB/PFC configuration is correct.  VERBOSE: RDMA configuration is correct.  VERBOSE: Checking if remote IP address, 192.168.2.5, is reachable.  VERBOSE: Remote IP 192.168.2.5 is reachable.  VERBOSE: Disabling RDMA on adapters that are not part of this test. RDMA will be enabled on them later.  VERBOSE: Testing RDMA traffic now for. Traffic will be sent in a parallel job. Job details:  VERBOSE: 0 RDMA bytes written per second  VERBOSE: 0 RDMA bytes sent per second  VERBOSE: 541185606 RDMA bytes written per second  VERBOSE: 34821478 RDMA bytes sent per second  VERBOSE: 954717307 RDMA bytes written per second  VERBOSE: 35040816 RDMA bytes sent per second  VERBOSE: Enabling RDMA on adapters that are not part of this test. RDMA was disabled on them prior to sending RDMA traffic.  VERBOSE: RDMA traffic test SUCCESSFUL: RDMA traffic was sent to 192.168.2.5 |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | **6.0 vSwitch Creation** |
| **1** |  |  |  |
| **2** |  |  | # Create a simple virtual switch over RDMA capable adapter  **New-VMSwitch -Name VMSTEST -NetAdapterName "Test-40G-1" -AllowManagementOS $true**    Name SwitchType NetAdapterInterfaceDescription  ---- ---------- ------------------------------  VMSTEST External Mellanox ConnectX-3 Pro Ethernet Adapter |
| **3** |  |  | # AllowManagementOS will create a Host vNIC which inherits the MAC and IP of the Physical NIC under the vSwitch  **Get-NetAdapter | ft -AutoSize**    Name InterfaceDescription ifIndex Status MacAddress LinkSpeed  ---- -------------------- ------- ------ ---------- ---------  vEthernet (VMSTEST) Hyper-V Virtual Ethernet Adapter #2 27 Up E4-1D-2D-07-40-71 40 Gbps  # A Host vNIC is managed two ways, one representation is the NetAdapter view which operates on the "vEthernet (VMSTEST)" Name  # The other mechanism is the VMNetworkAdapter view which drops the "vEthernet" prefix and simply uses the vmswitch name.  # The VMNetworkAdapter view which allows for setting some properties not accessible via NetAdapter.  **Get-VMNetworkAdapter –ManagementOS | ft -AutoSize**    Name IsManagementOs VMName SwitchName MacAddress Status IPAddresses  ---- -------------- ------ ---------- ---------- ------ -----------  CORP-External-Switch True CORP-External-Switch 001B785768AA {Ok}  VMSTEST True VMSTEST E41D2D074071 {Ok}  # Remove the ACCESS VLAN Setting from the physical NIC to prevent it from both auto-tagging the egress traffic with incorrect VLAN ID and filtering ingress traffic which doesn’t match the ACCESS VLAN ID.  **Set-NetAdapterAdvancedProperty -Name "Test-40G-1" -RegistryKeyword VlanID -RegistryValue "0"**  # Confirm the VlanID setting:  **Get-NetAdapterAdvancedProperty -Name 'TEST-10G-1' | Where-Object {$\_.RegistryKeyword -eq 'VlanID'} | ft -AutoSize**  Name DisplayName DisplayValue RegistryKeyword RegistryValue  ---- ----------- ------------ --------------- -------------  TEST-10G-1 VLAN ID 0 VlanID {0} |
| **4** |  |  | # When we test the Connection again, it will fail. This is because the physical NIC was placed on an Access VLAN in earlier steps, and when a vSwitch is created on the NIC, the Host vNIC does not inherit the VLAN properties.  **Test-NetConnection 192.168.1.5**  WARNING: Ping to 192.168.1.5 failed -- Status: DestinationHostUnreachable    ComputerName : 192.168.1.5  RemoteAddress : 192.168.1.5  InterfaceAlias : vEthernet (CORP-External-Switch)  SourceAddress : 10.199.48.170  PingSucceeded : False  PingReplyDetails (RTT) : 0 ms      **Set-VMNetworkAdapterVlan -VMNetworkAdapterName "VMSTEST" -VlanId "101" -Access -ManagementOS**  **Get-VMNetworkAdapterVlan -ManagementOS -VMNetworkAdapterName "VMSTEST"**    VMName VMNetworkAdapterName Mode VlanList  ------ -------------------- ---- --------  VMSTEST Access 101    # Recall that the change may take a few seconds to complete before you can ping the other adapter.  **Test-NetConnection 192.168.1.5**    ComputerName : 192.168.1.5  RemoteAddress : 192.168.1.5  InterfaceAlias : vEthernet (VMSTEST)  SourceAddress : 192.168.1.3  PingSucceeded : True  PingReplyDetails (RTT) : 0 ms |
| **5** |  |  |  |
| **6** |  |  | # Create a vSwitch in Switch Embedded Teaming mode (ie. SET).  **New-VMSwitch –Name "VMSTEST" –NetAdapterName "TEST-40G-1","TEST-40G-2" -EnableEmbeddedTeaming $true -AllowManagementOS $true**  Name SwitchType NetAdapterInterfaceDescription  ---- ---------- ------------------------------  VMSTEST External Teamed-Interface  # List the Physical adapter team in SET  **Get-VMSwitchTeam -Name "VMSTEST" | fl**  Name : VMSTEST  Id : ad9bb542-dda2-4450-a00e-f96d44bdfbec  NetAdapterInterfaceDescription : {Mellanox ConnectX-3 Pro Ethernet Adapter, Mellanox ConnectX-3 Pro Ethernet Adapter #2}  TeamingMode : SwitchIndependent  LoadBalancingAlgorithm : Dynamic |
| **7** |  |  | # Display the two (2) views of the Host vNIC.  **Get-NetAdapter**  Name InterfaceDescription ifIndex Status MacAddress LinkSpeed  ---- -------------------- ------- ------ ---------- ---------  vEthernet (VMSTEST) Hyper-V Virtual Ethernet Adapter #2 28 Up E4-1D-2D-07-40-71 80 Gbps  **Get-VMNetworkAdapter -ManagementOS**  Name IsManagementOs VMName SwitchName MacAddress Status IPAddresses  ---- -------------- ------ ---------- ---------- ------ -----------  VMSTEST True VMSTEST E41D2D074071 {Ok} |
| **8** |  |  | **Test-NetConnection 192.168.1.5** #The remote VLAN 101 adapter.  WARNING: Ping to 192.168.1.5 failed -- Status: DestinationHostUnreachable  ComputerName : 192.168.1.5  RemoteAddress : 192.168.1.5  InterfaceAlias : vEthernet (CORP-External-Switch)  SourceAddress : 10.199.48.170  PingSucceeded : False  PingReplyDetails (RTT) : 0 ms  # Remove the ACCESS VLAN Setting from the physical NIC to prevent it from both auto-tagging the egress traffic with incorrect VLAN ID and filtering ingress traffic which doesn’t match the ACCESS VLAN ID.  **Set-NetAdapterAdvancedProperty -Name "Test-40G-1" -RegistryKeyword VlanID -RegistryValue "0"**  **Set-NetAdapterAdvancedProperty -Name "Test-40G-2" -RegistryKeyword VlanID -RegistryValue "0"**  # Now set the VLANID using the vSwitch specific CMDLETs  **Set-VMNetworkAdapterVlan -VMNetworkAdapterName "VMSTEST" -VlanId "101" -Access -ManagementOS**  **Get-VMNetworkAdapterVlan -ManagementOS -VMNetworkAdapterName "VMSTEST"**  VMName VMNetworkAdapterName Mode VlanList  ------ -------------------- ---- --------  VMSTEST Access 101  **Test-NetConnection 192.168.1.5**  ComputerName : 192.168.1.5  RemoteAddress : 192.168.1.5  InterfaceAlias : vEthernet (VMSTEST)  SourceAddress : 192.168.1.3  PingSucceeded : True  PingReplyDetails (RTT) : 0 ms  # If ping still fails, confirm that the “vEthernet (VMSTEST)” has the proper IP address set.  **Get-NetIPAddress -InterfaceAlias "vEthernet (VMSTEST)"**  # If not set, use the following.  **New-NetIPAddress -InterfaceAlias "vEthernet (VMSTEST)" -IPAddress 192.168.1.3 -PrefixLength 24**  IPAddress : 192.168.1.3  InterfaceIndex : 37  InterfaceAlias : vEthernet (VMSTEST)  AddressFamily : IPv4  Type : Unicast  PrefixLength : 24  PrefixOrigin : Manual  SuffixOrigin : Manual  AddressState : Tentative  ValidLifetime : Infinite ([TimeSpan]::MaxValue)  PreferredLifetime : Infinite ([TimeSpan]::MaxValue)  SkipAsSource : False  PolicyStore : ActiveStore |
| **9** |  |  | # Now that we’ve verified the NIC has functionality, we rename the Management NIC and later use separate Host vNICs instances for RDMA  **Rename-VMNetworkAdapter -ManagementOS -Name “VMSTEST” -NewName “MGT”**  **Get-VMNetworkAdapter -ManagementOS**  Name IsManagementOs VMName SwitchName MacAddress Status IPAddresses  ---- -------------- ------ ---------- ---------- ------ -----------  CORP-External-Switch True CORP-External-Switch 001B785768AA {Ok}  MGT True VMSTEST E41D2D074071 {Ok}  **Get-NetAdapter**  Name InterfaceDescription ifIndex Status MacAddress LinkSpeed  ---- -------------------- ------- ------ ---------- ---------  vEthernet (MGT) Hyper-V Virtual Ethernet Adapter #2 28 Up E4-1D-2D-07-40-71 80 Gbps |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | **7.0 Test vSwitch RDMA (Mode 2)** |
| **1** |  |  |  |
| **2** |  |  | # Set Priority tagging on the Host vNIC to complement the previous VLAN settings.  **Set-VMNetworkAdapter -ManagementOS -Name "VMSTEST" -IeeePriorityTag on**  **Get-VMNetworkAdapter -ManagementOS -Name "VMSTEST" | fl Name,IeeePriorityTag**  Name : VMSTEST  IeeePriorityTag : On |
| **3** |  |  | **Get-NetAdapterRdma -Name "vEthernet (VMSTEST)"**  Name InterfaceDescription Enabled  ---- -------------------- -------  vEthernet (VMSTEST) Hyper-V Virtual Ethernet Adapter #2 False  # Now enable RDMA on the Host vNIC  **Get-NetAdapter**  Name InterfaceDescription ifIndex Status MacAddress LinkSpeed  ---- -------------------- ------- ------ ---------- ---------  vEthernet (VMSTEST) Hyper-V Virtual Ethernet Adapter #2 27 Up E4-1D-2D-07-40-71 40 Gbps  **Enable-NetAdapterRdma -Name "vEthernet (VMSTEST)"**  **Get-NetAdapterRdma -Name "vEthernet (VMSTEST)"**  Name InterfaceDescription Enabled  ---- -------------------- -------  vEthernet (VMSTEST) Hyper-V Virtual Ethernet Adapter #2 True |
| **4** |  |  | **Get-NetAdapter**  Name InterfaceDescription ifIndex Status MacAddress LinkSpeed  ---- -------------------- ------- ------ ---------- ---------  vEthernet (VMSTEST) Hyper-V Virtual Ethernet Adapter #2 27 Up E4-1D-2D-07-40-71 40 Gbps  **C:\TEST\Test-RDMA.PS1 -IfIndex 27 -IsRoCE $true -RemoteIpAddress 192.168.1.5 -PathToDiskspd C:\TEST\Diskspd-v2.0.17\amd64fre\**  VERBOSE: Diskspd.exe found at C:\TEST\Diskspd-v2.0.17\amd64fre\\diskspd.exe  VERBOSE: The adapter vEthernet (VMSTEST) is a virtual adapter  VERBOSE: Retrieving vSwitch bound to the virtual adapter  VERBOSE: Found vSwitch: VMSTEST  VERBOSE: Found the following physical adapter(s) bound to vSwitch: TEST-40G-1  VERBOSE: Underlying adapter is RoCE. Checking if QoS/DCB/PFC is configured on each physical adapter(s)  VERBOSE: QoS/DCB/PFC configuration is correct.  VERBOSE: RDMA configuration is correct.  VERBOSE: Checking if remote IP address, 192.168.1.5, is reachable.  VERBOSE: Remote IP 192.168.1.5 is reachable.  VERBOSE: Disabling RDMA on adapters that are not part of this test. RDMA will be enabled on them later.  VERBOSE: Testing RDMA traffic now for. Traffic will be sent in a parallel job. Job details:  VERBOSE: 0 RDMA bytes written per second  VERBOSE: 0 RDMA bytes sent per second  VERBOSE: 0 RDMA bytes written per second  VERBOSE: 0 RDMA bytes sent per second  VERBOSE: 0 RDMA bytes written per second  VERBOSE: 0 RDMA bytes sent per second  VERBOSE: 0 RDMA bytes written per second  VERBOSE: 9162492 RDMA bytes sent per second  VERBOSE: 938797258 RDMA bytes written per second  VERBOSE: 34621865 RDMA bytes sent per second  VERBOSE: 933572610 RDMA bytes written per second  VERBOSE: 35035861 RDMA bytes sent per second  VERBOSE: Enabling RDMA on adapters that are not part of this test. RDMA was disabled on them prior to sending RDMA traffic.  VERBOSE: RDMA traffic test SUCCESSFUL: RDMA traffic was sent to 192.168.1.5 |
| **5** |  |  |  |
| **6** |  |  | **Set-VMNetworkAdapter -ManagementOS -Name "MGT" -IeeePriorityTag on**  **Get-VMNetworkAdapter -ManagementOS -Name "MGT" | fl Name,IeeePriorityTag**  Name : MGT  IeeePriorityTag : On |
| **7** |  |  | # Now create two Host vNIC for RDMA  **Add-VMNetworkAdapter –SwitchName "VMSTEST" –Name SMB1 –ManagementOS**  **Add-VMNetworkAdapter –SwitchName "VMSTEST" –Name SMB2 –ManagementOS**  **Get-VMNetworkAdapter -ManagementOS**  Name IsManagementOs VMName SwitchName MacAddress Status IPAddresses  ---- -------------- ------ ---------- ---------- ------ -----------  CORP-External-Switch True CORP-External-Switch 001B785768AA {Ok}  Mgt True VMSTEST E41D2D074071 {Ok}  SMB1 True VMSTEST 00155D30AA00 {Ok}  SMB2 True VMSTEST 00155D30AA01 {Ok} |
| **8** |  |  | # Assign an IP address to the SMB Host vNICs. Note that TEST-40G-1 and TEST-40G-2 physical adapters still have ACCESS VLAN of 101 and 102 set. Hence they’ll tag the traffic and ping succeeds. We’ll remove this in later steps.   |  | | --- | | On page 10, we set both pNIC VLAN IDs to zero, then set the VMSTEST vSwitch to VLAN 101. After that, we can still ping the remote VLAN 101 adapter via the MGT vNIC, but we don’t have any VLAN 102 members.  “# Remove the ACCESS VLAN Setting from the physical NIC to prevent it from both auto-tagging the egress traffic with incorrect VLAN ID and filtering ingress traffic which doesn’t match the ACCESS VLAN ID.  ” |   **New-NetIPAddress -InterfaceAlias "vEthernet (SMB1)" -IPAddress 192.168.2.111 -PrefixLength 24**  IPAddress : 192.168.2.111  InterfaceIndex : 40  InterfaceAlias : vEthernet (SMB1)  AddressFamily : IPv4  Type : Unicast  PrefixLength : 24  PrefixOrigin : Manual  SuffixOrigin : Manual  AddressState : Invalid  ValidLifetime : Infinite ([TimeSpan]::MaxValue)  PreferredLifetime : Infinite ([TimeSpan]::MaxValue)  SkipAsSource : False  PolicyStore : PersistentStore  #Target the remote VLAN 102 adapter.  **Test-NetConnection 192.168.2.5**  ComputerName : 192.168.2.5  RemoteAddress : 192.168.2.5  InterfaceAlias : vEthernet (SMB1)  SourceAddress : 192.168.2.111  PingSucceeded : True  PingReplyDetails (RTT) : 0 ms  **New-NetIPAddress -InterfaceAlias "vEthernet (SMB2)" -IPAddress 192.168.2.222 -PrefixLength 24**  IPAddress : 192.168.2.222  InterfaceIndex : 44  InterfaceAlias : vEthernet (SMB2)  AddressFamily : IPv4  Type : Unicast  PrefixLength : 24  PrefixOrigin : Manual  SuffixOrigin : Manual  AddressState : Invalid  ValidLifetime : Infinite ([TimeSpan]::MaxValue)  PreferredLifetime : Infinite ([TimeSpan]::MaxValue)  SkipAsSource : False  PolicyStore : PersistentStore  # No need to test connection again since we have reachability already |
| **9** |  |  | # Since we placed the “MGT” Host vNIC on VLAN 101, we’ll place the RDMA Host vNICs on the pre-existing VLAN 102.  **Set-VMNetworkAdapterVlan -VMNetworkAdapterName "SMB1" -VlanId "102" -Access -ManagementOS**  **Set-VMNetworkAdapterVlan -VMNetworkAdapterName "SMB2" -VlanId "102" -Access -ManagementOS**  **Get-VMNetworkAdapterVlan -ManagementOS**  VMName VMNetworkAdapterName Mode VlanList  ------ -------------------- ---- --------  SMB1 Access 102  Mgt Access 101  SMB2 Access 102  CORP-External-Switch Untagged |
| **10** |  |  | # Inspect the Mapping of SMB1 and SMB2 to the underlying Physical NICs under the vSwitch Team. The association of Host vNIC to Physical NICs is random and subject to rebalancing during creation and destruction.  # Here we use an indirect mechanism to check the current association. Note that the MAC address of SMB1 and SMB2 are associated with NIC Team member “TEST-40G-2”. This is not ideal since Test-40G-1 will does not have an associated SMB Host vNIC and will not allow for utilization of RDMA traffic over the link until an SMB Host vNIC is mapped to it.  **Get-NetAdapterVPort** (Preferred)  **Get-NetAdapterVmqQueue**  Name QueueID MacAddress VlanID Processor VmFriendlyName  ---- ------- ---------- ------ --------- --------------  TEST-40G-1 1 E4-1D-2D-07-40-71 101 0:17  TEST-40G-2 1 00-15-5D-30-AA-00 102 0:17  TEST-40G-2 2 00-15-5D-30-AA-01 102 0:17  **Get-VMNetworkAdapter -ManagementOS**  Name IsManagementOs VMName SwitchName MacAddress Status IPAddresses  ---- -------------- ------ ---------- ---------- ------ -----------  CORP-External-Switch True CORP-External-Switch 001B785768AA {Ok}  Mgt True VMSTEST E41D2D074071 {Ok}  SMB1 True VMSTEST 00155D30AA00 {Ok}  SMB2 True VMSTEST 00155D30AA01 {Ok}  # Both commands below should return no info since no mapping has been performed.  **Get-VMNetworkAdapterTeamMapping -ManagementOS -SwitchName VMSTEST -VMNetworkAdapterName SMB1**  **Get-VMNetworkAdapterTeamMapping -ManagementOS -SwitchName VMSTEST -VMNetworkAdapterName SMB2**  # Map SMB1 and SMB2 to separate physical NIC team members – It \_MUST\_ be done here before proceeding with further configuration steps.  **Set-VMNetworkAdapterTeamMapping -ManagementOS -SwitchName VMSTEST -VMNetworkAdapterName "SMB1" -PhysicalNetAdapterName "Test-40G-1"**  **Set-VMNetworkAdapterTeamMapping -ManagementOS -SwitchName VMSTEST -VMNetworkAdapterName "SMB2" -PhysicalNetAdapterName "Test-40G-2"**  **Get-VMNetworkAdapterTeamMapping -ManagementOS -SwitchName VMSTEST**  NetAdapterName : Test-40G-1  NetAdapterDeviceId : {BAA9A00F-A844-4740-AA93-6BD838F8CFBA}  ParentAdapter : VMInternalNetworkAdapter, Name = 'SMB1'  IsTemplate : False  CimSession : CimSession: .  ComputerName : 27-3145G0803  IsDeleted : False  NetAdapterName : Test-40G-2  NetAdapterDeviceId : {B7AB5BB3-8ACB-444B-8B7E-BC882935EBC8}  ParentAdapter : VMInternalNetworkAdapter, Name = 'SMB2'  IsTemplate : False  CimSession : CimSession: .  ComputerName : 27-3145G0803  IsDeleted : False  # Indirectly confirm the MAC associations  **Get-NetAdapterVmqQueue**  Name QueueID MacAddress VlanID Processor VmFriendlyName  ---- ------- ---------- ------ --------- --------------  TEST-40G-1 1 E4-1D-2D-07-40-71 101 0:17  TEST-40G-1 2 00-15-5D-30-AA-00 102 0:17  TEST-40G-2 1 00-15-5D-30-AA-01 102 0:17 |
| **11** |  |  | **!!!IMPORTANT!!! – Run commands on remote system**  # Since both Host vNICs reside on the same subnet \_AND\_ same VLAN ID (102), then validate the reachability from the remote system.  **Test-NetConnection 192.168.2.111**  ComputerName : 192.168.2.111  RemoteAddress : 192.168.2.111  InterfaceAlias : Test-40G-2  SourceAddress : 192.168.2.5  PingSucceeded : True  PingReplyDetails (RTT) : 0 ms  **Test-NetConnection 192.168.2.222**  ComputerName : 192.168.2.222  RemoteAddress : 192.168.2.222  InterfaceAlias : Test-40G-2  SourceAddress : 192.168.2.5  PingSucceeded : True  PingReplyDetails (RTT) : 0 ms |
| **12** |  |  | **Set-VMNetworkAdapter -ManagementOS -Name "SMB1" -IeeePriorityTag on**  **Set-VMNetworkAdapter -ManagementOS -Name "SMB2" -IeeePriorityTag on**  **Get-VMNetworkAdapter -ManagementOS -Name "SMB\*" | fl Name,SwitchName,IeeePriorityTag,Status**  Name : SMB1  SwitchName : VMSTEST  IeeePriorityTag : On  Status : {Ok}  Name : SMB2  SwitchName : VMSTEST  IeeePriorityTag : On  Status : {Ok} |
| **13** |  |  | **Get-NetAdapterRdma -Name "vEthernet\*" | sort Name | ft -AutoSize**  Name InterfaceDescription Enabled  ---- -------------------- -------  vEthernet (SMB2) Hyper-V Virtual Ethernet Adapter #4 False  vEthernet (SMB1) Hyper-V Virtual Ethernet Adapter #3 False  vEthernet (MGT) Hyper-V Virtual Ethernet Adapter #2 False  **Enable-NetAdapterRdma -Name "vEthernet (SMB1)"**  **Enable-NetAdapterRdma -Name "vEthernet (SMB2)"**  **Get-NetAdapterRdma -Name "vEthernet\*" | sort Name | fl \***  Name InterfaceDescription Enabled  ---- -------------------- -------  vEthernet (SMB2) Hyper-V Virtual Ethernet Adapter #4 True  vEthernet (SMB1) Hyper-V Virtual Ethernet Adapter #3 True  vEthernet (MGT) Hyper-V Virtual Ethernet Adapter #2 False |
| **14** |  |  | First Run these commands from the **remote system.**  # The intent here is to validate RDMA functionality from the remote system to our system which has a vSwitch, thus testing both adapter under the vSwitch team  # Since we placed both Host vNICs (SMB1 and SMB2) on VLAN 102, we need to select the VLAN 102 adapter on the remote system. Here Test-40G-2 does RDMA to SMB1 (192.168.2.111) and SMB2 (192.168.2.222).  # Optional: You may need to disable the Firewall on this system. Consult your fabric policy for details.  **Set-NetFirewallProfile -All -Enabled False**  **Get-NetAdapterAdvancedProperty -Name "Test-40G-2"**  Name DisplayName DisplayValue RegistryKeyword RegistryValue  ---- ----------- ------------ --------------- -------------  .  .  Test-40G-2 VLAN ID 102 VlanID {102}  **Get-NetAdapter**  Name InterfaceDescription ifIndex Status MacAddress LinkSpeed  ---- -------------------- ------- ------ ---------- ---------  Test-40G-2 Mellanox ConnectX-3 Pro Ethernet A...#3 3 Up E4-1D-2D-07-43-D1 40 Gbps  **Get-NetAdapterRdma**  Name InterfaceDescription Enabled  ---- -------------------- -------  Test-40G-2 Mellanox ConnectX-3 Pro Ethernet Adap... True  **C:\TEST\Test-RDMA.PS1 -IfIndex 3 -IsRoCE $true -RemoteIpAddress 192.168.2.111 -PathToDiskspd C:\TEST\Diskspd-v2.0.17\amd64fre\**  VERBOSE: Diskspd.exe found at C:\TEST\Diskspd-v2.0.17\amd64fre\\diskspd.exe  VERBOSE: The adapter Test-40G-2 is a physical adapter  VERBOSE: Underlying adapter is RoCE. Checking if QoS/DCB/PFC is configured on each physical adapter(s)  VERBOSE: QoS/DCB/PFC configuration is correct.  VERBOSE: RDMA configuration is correct.  VERBOSE: Checking if remote IP address, 192.168.2.111, is reachable.  VERBOSE: Remote IP 192.168.2.111 is reachable.  VERBOSE: Disabling RDMA on adapters that are not part of this test. RDMA will be enabled on them later.  VERBOSE: Testing RDMA traffic now for. Traffic will be sent in a parallel job. Job details:  VERBOSE: 34251744 RDMA bytes sent per second  VERBOSE: 967346308 RDMA bytes written per second  VERBOSE: 35698177 RDMA bytes sent per second  VERBOSE: 976601842 RDMA bytes written per second  VERBOSE: Enabling RDMA on adapters that are not part of this test. RDMA was disabled on them prior to sending RDMA traffic.  VERBOSE: RDMA traffic test SUCCESSFUL: RDMA traffic was sent to 192.168.2.111  **C:\TEST\Test-RDMA.PS1 -IfIndex 3 -IsRoCE $true -RemoteIpAddress 192.168.2.222 -PathToDiskspd C:\TEST\Diskspd-v2.0.17\amd64fre\**  VERBOSE: Diskspd.exe found at C:\TEST\Diskspd-v2.0.17\amd64fre\\diskspd.exe  VERBOSE: The adapter Test-40G-2 is a physical adapter  VERBOSE: Underlying adapter is RoCE. Checking if QoS/DCB/PFC is configured on each physical adapter(s)  VERBOSE: QoS/DCB/PFC configuration is correct.  VERBOSE: RDMA configuration is correct.  VERBOSE: Checking if remote IP address, 192.168.2.222, is reachable.  VERBOSE: Remote IP 192.168.2.222 is reachable.  VERBOSE: Disabling RDMA on adapters that are not part of this test. RDMA will be enabled on them later.  VERBOSE: Testing RDMA traffic now for. Traffic will be sent in a parallel job. Job details:  VERBOSE: 0 RDMA bytes written per second  VERBOSE: 0 RDMA bytes sent per second  VERBOSE: 485137693 RDMA bytes written per second  VERBOSE: 35200268 RDMA bytes sent per second  VERBOSE: 939044611 RDMA bytes written per second  VERBOSE: 34880901 RDMA bytes sent per second  VERBOSE: Enabling RDMA on adapters that are not part of this test. RDMA was disabled on them prior to sending RDMA traffic.  VERBOSE: RDMA traffic test SUCCESSFUL: RDMA traffic was sent to 192.168.2.222 |
| **15** |  |  | # Finally, test for RDMA traffic from the **local system** to the remote system.  **Get-NetAdapter | ft –AutoSize**  Name InterfaceDescription ifIndex Status MacAddress LinkSpeed  ---- -------------------- ------- ------ ---------- ---------  vEthernet (SMB2) Hyper-V Virtual Ethernet Adapter #4 45 Up 00-15-5D-30-AA-03 80 Gbps  vEthernet (SMB1) Hyper-V Virtual Ethernet Adapter #3 41 Up 00-15-5D-30-AA-02 80 Gbps  **C:\TEST\Test-RDMA.PS1 -IfIndex 41 -IsRoCE $true -RemoteIpAddress 192.168.2.5 -PathToDiskspd C:\TEST\Diskspd-v2.0.17\amd64fre\**  VERBOSE: Diskspd.exe found at C:\TEST\Diskspd-v2.0.17\amd64fre\\diskspd.exe  VERBOSE: The adapter vEthernet (SMB1) is a virtual adapter  VERBOSE: Retrieving vSwitch bound to the virtual adapter  VERBOSE: Found vSwitch: VMSTEST  VERBOSE: Found the following physical adapter(s) bound to vSwitch: TEST-40G-1, TEST-40G-2  VERBOSE: Underlying adapter is RoCE. Checking if QoS/DCB/PFC is configured on each physical adapter(s)  VERBOSE: QoS/DCB/PFC configuration is correct.  VERBOSE: RDMA configuration is correct.  VERBOSE: Checking if remote IP address, 192.168.2.5, is reachable.  VERBOSE: Remote IP 192.168.2.5 is reachable.  VERBOSE: Disabling RDMA on adapters that are not part of this test. RDMA will be enabled on them later.  VERBOSE: Testing RDMA traffic now for. Traffic will be sent in a parallel job. Job details:  VERBOSE: 0 RDMA bytes written per second  VERBOSE: 0 RDMA bytes sent per second  VERBOSE: 0 RDMA bytes written per second  VERBOSE: 15250197 RDMA bytes sent per second  VERBOSE: 896320913 RDMA bytes written per second  VERBOSE: 33947559 RDMA bytes sent per second  VERBOSE: 912160540 RDMA bytes written per second  VERBOSE: 34091930 RDMA bytes sent per second  VERBOSE: Enabling RDMA on adapters that are not part of this test. RDMA was disabled on them prior to sending RDMA traffic.  VERBOSE: RDMA traffic test SUCCESSFUL: RDMA traffic was sent to 192.168.2.5  **C:\TEST\Test-RDMA.PS1 -IfIndex 45 -IsRoCE $true -RemoteIpAddress 192.168.2.5 -PathToDiskspd C:\TEST\Diskspd-v2.0.17\amd64fre\**  VERBOSE: Diskspd.exe found at C:\TEST\Diskspd-v2.0.17\amd64fre\\diskspd.exe  VERBOSE: The adapter vEthernet (SMB2) is a virtual adapter  VERBOSE: Retrieving vSwitch bound to the virtual adapter  VERBOSE: Found vSwitch: VMSTEST  VERBOSE: Found the following physical adapter(s) bound to vSwitch: TEST-40G-1, TEST-40G-2  VERBOSE: Underlying adapter is RoCE. Checking if QoS/DCB/PFC is configured on each physical adapter(s)  VERBOSE: QoS/DCB/PFC configuration is correct.  VERBOSE: RDMA configuration is correct.  VERBOSE: Checking if remote IP address, 192.168.2.5, is reachable.  VERBOSE: Remote IP 192.168.2.5 is reachable.  VERBOSE: Disabling RDMA on adapters that are not part of this test. RDMA will be enabled on them later.  VERBOSE: Testing RDMA traffic now for. Traffic will be sent in a parallel job. Job details:  VERBOSE: 0 RDMA bytes written per second  VERBOSE: 0 RDMA bytes sent per second  VERBOSE: 385169487 RDMA bytes written per second  VERBOSE: 33902277 RDMA bytes sent per second  VERBOSE: 907354685 RDMA bytes written per second  VERBOSE: 33923662 RDMA bytes sent per second  VERBOSE: Enabling RDMA on adapters that are not part of this test. RDMA was disabled on them prior to sending RDMA traffic.  VERBOSE: RDMA traffic test SUCCESSFUL: RDMA traffic was sent to 192.168.2.5 |
|  |  |  |  |

Appendix 1: Physical Switch Configuration

|  |  |
| --- | --- |
| **Arista** switch (dcs-7050s-64, EOS-4.13.7M):  (These are only commands and their uses; admins need to determine which ports the NICs are connected to.  Ensure that VLAN and no-drop policy is set for the priority over which SMB is configured.) | 1. en (go to admin mode, usually asks for a password) 2. config (to enter into configuration mode) 3. show run (shows current running configuration) 4. find out switch ports to which your NICs are connected to. In these example, they are 14/1,15/1,16/1,17/1. 5. int eth 14/1,15/1,16/1,17/1 (enter into config mode for these ports) 6. dcbx mode ieee 7. priority-flow-control mode on 8. switchport trunk native vlan 225 9. switchport trunk allowed vlan 100-225 10. switchport mode trunk 11. priority-flow-control priority 3 no-drop 12. qos trust cos 13. show run (verify that configuration is setup correctly on the ports) 14. wr (to make the settings persists across switch reboot)   **Tips:**   1. No #command# negates a command 2. How to add a new VLAN: int vlan 100 (If storage network is on VLAN 100) 3. How to check existing VLANs : show vlan 4. For more information on configuring Arista Switch, search online for: Arista EOS Manual 5. Use this command to verify PFC settings: show priority-flow-control counters detail |
| **Dell** switch (S4810, FTOS 9.9 (0.0)) | !  dcb enable  ! put pfc control on qos class 3  configure  dcb-map dcb-smb  priority group 0 bandwidth 90 pfc on  priority group 1 bandwidth 10 pfc off  priority-pgid 1 1 1 0 1 1 1 1  exit  ! apply map to ports 0-31  configure  interface range ten 0/0-31  dcb-map dcb-smb  exit |
| **Cisco** switch (Nexus 3132, version 6.0(2)U6(1)) | **Global:**  class-map type qos match-all RDMA  match cos 3  class-map type queuing RDMA  match qos-group 3  policy-map type qos QOS\_MARKING  class RDMA  set qos-group 3  class class-default  policy-map type queuing QOS\_QUEUEING  class type queuing RDMA  bandwidth percent 50  class type queuing class-default  bandwidth percent 50  class-map type network-qos RDMA  match qos-group 3  policy-map type network-qos QOS\_NETWORK  class type network-qos RDMA  mtu 2240  pause no-drop  class type network-qos class-default  mtu 9216  system qos  service-policy type qos input QOS\_MARKING  service-policy type queuing output QOS\_QUEUEING  service-policy type network-qos QOS\_NETWORK  **Port specific:**  switchport mode trunk  switchport trunk native vlan 99  switchport trunk allowed vlan 99,2000,2050   çuse VLANs that already exists  spanning-tree port type edge  flowcontrol receive on (not supported with PFC in Cisco NX-OS)  flowcontrol send on (not supported with PFC in Cisco NX-OS)  no shutdown  priority-flow-control mode on |

Appendix 2: Troubleshooting/Verification steps.

Script to check if RDMA configuration is correct on the host:

<https://github.com/Microsoft/SDN/blob/master/Diagnostics/Test-Rdma.ps1>

|  |  |  |
| --- | --- | --- |
| **Cmdlet** | **Expected** | **What to do if unexpected is encountered?** |
| Get-NetAdapterRdma | fl \* | Host vNIC and the physical NIC show non-zero RDMA capabilities.    cid:image001.jpg@01D240CB.E79DC160 | * 1. Make sure the Mlnx miniport and Mlnx bus drivers are latest. For Mellanox, use at least drop 42.   2. Verify that Mlnx miniport and bus drivers match by checking the driver version through Device Manager. The bus driver can be found in System Devices. The name should start with Mellanox Connect-X 3 PRO VPI.   3. Machine generated alternative text:      Mellanox ConnectX-3 Pro Ethernet Adapter Properties       Mellanox ConnectX-3 PRO "PI (MTC41C'3) Network Adapter Prope.       General Port Protocol Driver Details Events Resources       Mellanox ConnectX-3 PRO VPI (MT04103) Network Adapter       Driver Pm vider Mellanox Technologies Ltd       x       Information       Adapter Homation       Power Management       Performance Driver       Me llanox       Information       Divar Version       Firmware Version       Port Number       Bus Type       Link Speed       Part Number       Device Id       Revision Id       Currant MAC Address       Parmanant MAC Address       Network Status       Adapter Friendly Name       IPv4 Address       Adapter user Name       Adapter P Key       PCI-E80Gbpsx8       40 0 Gbps/FuII Duplex       MCX314A-8CCT       F4-52-14-74-E5-41       F4-52-14-74-E5-41       Sava To Fila       Dnver Date       Driver Vers,on       Digital Signer       Driver Details       Roll Back Divar       8/6/2015       495 0       Mellanox Technologies LTD       To view details about the driver files       To update the driver software for this device       f the device fails after updating tha driver roll       back to the previously installed driver       Disables the selected device       To uninstall the driver (Advanced)   4. Make sure Network Direct (RDMA) is enabled on both, physical NIC and host vNIC.   5. Make sure vSwitch is created over the right physical adapter by checking its RDMA capabilities.   6. Check EventViewer System log and filter by source “Hyper-V-VmSwitch”. |
|  | * 1. The host vNIC should appear as RDMA capable from SMB’s perspective as well.      * 1. cid:image003.jpg@01D240CB.E79DC160 | * 1. Perform the same steps as above.   2. Check EventViewer logs for “SMB Client” in Application And Services-> Microsoft-> Windows. |
| Get-NetAdapterQos | Priorities and traffic classes should be displayed as we configured them above in Step 1.  cid:image004.jpg@01D240CB.E79DC160  vst | * 1. Make sure the physical adapter supports DCB/QoS   2. Make sure the adapter’s drivers are up to date. |
| Get-SmbMultiChannelConnection | Remote node’s IP address is shown as RDMA capable.    cid:image005.jpg@01D240CB.E79DC160 | * 1. Make sure ping works both ways.   2. Make sure firewall is not blocking SMB connection initiation. Specifically, enable the firewall rule for SMB direct port 5445. |
| Get-SmbClientNetworkInterface | Virtual NIC that was enabled for RDMA must be seen as RDMA capable by SMB.    Machine generated alternative text: ps C: Win  s system32> Get-Sm I entNetwor Inter ace  Interface Index RSS Capable  RWA Capable Speed  False  False  False  False  False  : floe: 9a6a:  True  True  True  False  False  False  False  False  False  True  True  O  O  O  O  10  20  20  20  bps  bps  bps  bps  Gbps  Gbps  Gbps  Gb s  I pAddr es s es  {fe80: :7cde : 94cf : 20ef : c051}  {fe80: :3994 : 6957}  {fe80: : 618e :4643 :9047 : 29bc}  {fe80: :9ef : 92df : 685 :2979}  {fe80: : 7046, 169. 254. 7.181}  fe80:  {10.10.1.1}  10.10. 2.1  Friendly Name  NICI  NIC2  NIC3  NIC4  Local Area Connection* 2  6530:  66ce,  10. 248.138.  227}  vet her net  vet her net  vet her net  (NS-oo)  (vNIC2)  vNIC4 | * 1. Make sure ping works both ways.   2. Make sure firewall is not blocking SMB connection initiation. |
| vstat  (Mellanox specific) | RoCE version on both nodes must be the same. This is also a good way to verify that the firmware version on both nodes is latest.    cid:image007.jpg@01D240CB.E79DC160 | * 1. Set correct RoCE version using  Set-MlnxDriverCoreSetting   2. Install the latest firmware from Mellanox website. |
| Perfmon Counters | Perfmon counters for RDMA should be non-zero  cid:image008.jpg@01D240CB.E79DC160 |  |

Appendix 3: How to enable Guest RDMA.

How to enable Guest RDMA

1. Set the DWORD regkey *EnableGuestRdma* on the host at KEY\_LOCAL\_MACHINE\SYSTEM\CurrentControlSet\Services\mlx4\_bus\Parameter

Machine generated alternative text:
Computar+KEY LOCAL 
v mtx4 bus 
?trameters 
mtx4eth63 
MMCSS 
bus\Paramataß 
A Nama 
(DefauH) 
Device ManagedSteenng 
LoadadDevices 
Num Of Devices 
REG SZ 
REG 
DWORD 
REG 
OWORO 
REG 
OWORO 
REG 
DWORD 
(value not set) 
0100000001 
0100000001 
0100000001 
0100000001 

While creating registry key name, *EnableGuestRdma*, make sure there is no space at the end. These types of issues are very hard to figure out.

1. Restart the bus driver on the host. It is listed under System Devices:

Machine generated alternative text:
Intel(R) Xeon(R) E7 E5 i7VCU 2F8A 
Intel(R) Xeon(R) E7 E5 i7VCU 2F8A 
Mellanox ConnectX-3 PRO VPI (MT041D3) Network Adapter 
Microsoft ACPI-Comphant Power Meter Device 
Microsoft ACPI- Compliant Syst 

1. Create an IOV enabled vSwitch
   1. Non-SET: New-VMSwitch MlnxSwitch -NetAdapterName TEST-40G-1 -EnableIov $true
   2. SET:  New-VMSwitch MlnxSetSwitch -NetAdapterName TEST-40G-1,TEST-40G-2 -EnableIov $true

1. Currently Guest RDMA is only available on Mellanox port 1. You can find out the port by checking the adapter properties in ncpa.cpl:

Machine generated alternative text:
Mellanox ConnectX-3 Pro Ethernet Adapter Properties 
Power Management 
Homation 
Performance Driver 
Me llanox 
Homation 
Driver Version 
Pod Number 
Bus Type 
Link Speed 
Part Number 
Revision 
Current MAC Address 
P amanant MAC Address 
Network Status 
Adapter Friendly Name 
Pv4 Address 
Adapter Homation 
525 126650 
40 0 Gbps/FuII Duplex 
OR3FON 
7C-FE-w938F-OO 
7C-FE-w93-8F-OO 
Save To file 

1. Verify that native and host vNIC RDMA are working according to Converged NIC Configuration and Troubleshooting (vSwitch Host vNIC)

1. For a SET vSwitch, make sure you map the vmNic to port 1 of Mellanox NIC by using Set-VMNetworkAdapterTeamMapping

Set-VMNetworkAdapterTeamMapping -VMNetworkAdapterName "Network Adapter" -VMName VM1 -PhysicalNetAdapterName TEST-40G-2

1. Get SRIOV working by following the instructions at Enable SR-IOV with Mellanox CX-3

1. Make sure that external traffic from the VM flows over SRIOV path by checking Perfmon counters for the VF adapter in the VM.

Machine generated alternative text:
Network Adapter 
Bytes Received/sec 
Bytes Sent/sec 
Bytes Total/sec 
Current Bandwidth 
Output Queue Length 
Packets Outbound Discarded 
Packets Outbound Errors 
Packets Received Discarded 
Packets Received Errors 
Packets Received Non-Unicast/sec 
Packets Received Unicast/sec 
Packets Received Unknown 
Packets Received/sec 
Packets Sent Non- Unicast/sec 
Packets Sent Unicast/sec 
Packets Sent/sec 
Pac kets/sec 
TCP Active RSC Connections 
TCP RSC Average Packet Size 
TCP RSC Coalesced Packets/sec 
TCP RSC Exceptions/sec 
Mellanox ConnectX-3 Virtual Function Ethemet Adapter 
13,61$476.107 
2109, "7,353 
40DDODDODDODDDO 
149,841 mo 
1,377,413.142 
1,380545.517 
1.016 
145,277.429 
145,278.445 
531,823.962 
Microsoft Hyper-V Network Adapter 
12290505080 
1,50DD0780 
80DDODDODDODDDO 
1,284,276.135 
1,284,276.135 
1.016 
141,307.362 
141,308.378 

1. Enable guest RDMA inside the VM by doing Enable-NetAdapterRdma on the “Hyper-V Network Adapter”

Check RDMA capabilities of the “Hyper-V Network Adapter” by doing “Get-NetAdapterRdma | fl \*”. They should be non-zero.

Machine generated alternative text:
PS Enable-NetAdapterRdma Ethernet 
PS Get-NetAdapterRdma f I 
. Ethernet 
as 
Interfacul i as 
. Ethernet 
ifDesc 
. Microsoft Hyper-V Network Adapter 
apt ion 
. MSFT_NetAdapterRdmaSettingOata 'Mi crosoft Hyper-V Network Adapter ' 
Description 
. Microsoft Hyper-V Network Adapter 
El ementName 
. Microsoft Hyper-V Network Adapter 
. {IA328680-4C20-452F-9DAO-8520AOOF365C} 
InstanceID 
InterfaceDes cri pts on 
. Microsoft Hyper-V Network Adapter 
. Ethernet 
Name 
Source 
Syst emName 
: WIN-7LPCLME79SQ 
Enabled 
. True 
axCompI et i onQu eu eCou nt 
65408 
axlnboundReadL imit 
: 16776128 
axMenoryRegionCount 
. 524032 
ax" emoryn'i n donCou nt 
axOutboundReadLimit 
: 134209024 
axPr ot ect i on Domai nCount 
32764 
axQu eu epai rCount 
: 1048508 
axSharedReceiveQueueCount • 
. 65472 
RdmaAdapter I n fo 
RdmaN i s s i ngCounter Info 
PSComput er Name 
inCIass 
ROOT/ St andardC i mv2 : MSFT_NetAdapt er RdmaSett i ngDat a 
imlnstanceproperties 
{Caption, Description, ElementName, InstanceID.. 
imSyst es 
. Microsoft. Managenent. Infrastructure. CimSystenProperties 

1. Get-SmbClientNetworkInterface should now show Hyper-V Network adapter as RDMA capable.

Machine generated alternative text:
Users 
m strator 
Name 
2 Up 
Des op sq Get-Net 
InterfaceDescri ption 
Mel lanox ConnectX-3 Virtual 
apter 
iflndex 
Ethernet 2 
Eth et 
Function... 
Mi crosoft Hyper-V Network Adapter 
Stat us 
Up 
M acAddr ess 
00-15-50-32-67-00 
00-15-50-32-67-00 
Li n kSpeed 
40 Gbps 
80 Gbps 
PS Get-Sn«IientNetworkInterface 
Interface Index RSS Capable RWA Capable Speed 
IpAddr ess es 
{fe80: : cf47, 
Fri endly 
169. 254. 207. 71} Ethernet 
Name 
True 
True 
80 Gbps 

1. Traffic should now flow over RDMA over SRIOV path. Check Perfmon counters for "SMB Direct Connections" inside the VM.

Machine generated alternative text:
SMB Direct Connection 
Bytes R DMA W ritten/sec 
Bytes Received/sec 
Bytes Sent/sec 
Memory Regions 
RCQ Notification Events/sec 
R DMA Registrations/sec 
Receives/sec 
Remote Invalidations/sec 
SCQ Notification Events/sec 
Sends/sec 
Stalls (ROMA Read)/sec 
Stalls (R DMA 
Stalls (Send Credit)/sec 
Stalls (Send Queue)/sec 
3279,710444 
650681.572 
50996.144 
0095.037 
0256.55-4 
0255.563 
0255.563 
5,925.592 
0256.55-4 

1. Guest RDMA can co-exist with host vNIC RDMA but there can be starvation issues because all NDK traffic coming from SMBDirect has ND port 445. So, RDMA traffic from both host vNIC and the VM will be mapped to only one priority and therefore only one traffic class. Following screenshot shows host vNIC RDMA and Guest RDMA on a SET vSwitch working with a remote native RDMA node:

Machine generated alternative text:
Help 
Processor Information 
% Processor Time 
S MB Direct Connection 
Bytes R DMA Read/sec 
Bytes R DMA W ritten/sec 
Bytes Received/sec 
Bytes Sent/sec 
Memory Regions 
RCQ Notification a•ents/sec 
RDMA 
Receives/sec 
Remote Invalidations/sec 
SCQ Notification Events/sec 
Se nds/sec 
Stalls (ROMA Read)/sec 
Stalls (RDMA RegÉtrati0'B)/sec 
Stalls (Send Credit)/sec 
Stalls (Send Queue)/sec 
Administrator: Windows PowerSheII 
PS C: Users A ml n Istrator> Get-WSW1tc 
67-01 
0:1 
TEST-40G 2 
TEST-40G 2 
67-00 
"IMI on 27-3145HD81g - Virtual Machine Connection 
File Action Media View Help 
Performance Monitor 
File Action View Window Help 
0580 
406,301.194 
55180.767 
3847.686 
3915.663 
3908.665 
3908.665 
4,227.556 
3915.663 
11B6 
SwitchType NetAdapterInterfaceDescription 
6) 
Performance 
v Monitoring Tools 
Performance Monitor 
Data Collector Sets 
Reports 
\WlN-7LPCLME79SQ 
SMB Direct Connection 
Bytes R DMA Read/sec 
Bytes R DMA W ritten/sec 
Bytes Received/sec 
Bytes Sent/sec 
Memory Regions 
RCQ Notification Events/sec 
R DMA 
Receives/sec 
Remote Invalidations/sec 
SCQ Notification Events/sec 
Sends/sec 
Stalls (ROMA Read)/sec 
Stalls (R DMA Registrations)/sec 
Stalls (Send Credit)/sec 
Stalls (Send Queue)/sec 
470,477.751 
680215045 
4,443.381 
4,472.753 
4,523.396 
4,522.383 
4,945.755 
4,473.766 
Name 
MI nxSetSwitch Ext ernal Teamed-Interface 
PS Get-WSwitchTeam 
Name 
NetAdapt er Int erfaceDes cri pt ion 
Teami ngMode 
Load8aI anci ngAIgor i t hm 
MInxSetSwitch {Mel I anox ConnectX-3 Pro Ethernet Adapter 
PS Get-W 
Name State MenoryAssigned(M) Uptime 
Mel lanox ConnectX-3 Pro Ethernet Adapter} Switchlndependent Dynamic 
St at us 
Version 
WI Running 3 
1024 
PS Get-VMNetworkAdapter 
00:42:54. 
W Name Switch Name 
5760000 Operating normally 8.0 
Name 
vNIC2 
IsManagementOs 
True 
M -I nxSetSwi t ch 
M -I nxSetSwi t ch 
M acAddr es s 
001550326701 
001550326700 
VID 
101 
101 
Status IPAddresses 
{0k} 
{0k} 
{169. 254. 207. 71, 
Network Adapter False 
ers\Admi ni st r ator> 
fe80. 
• cf47} 
Name 
TEST 
TEST 
Get- NetAdapterVPort 
M a cAddr ess 
00-15-50 
00-15-50- 
Proc" ask 
FID 
-40G-1 
-40G-1 
Stat e 
Act i vat ed 
Act i vat ed 
Act i vat ed 
Act i vat ed 
Unknown 
Adapt i ve 
Unknown 
Unknown 
QPairs 
1 
1 
1 
1 