Dell EMC XC Series Appliance and XC Core System Remote Direct Memory Access Deployment Guide

August 2019

Dell EMC Reference Architecture

Revisions

Date	Description
May 2019	Initial release
August 2019	Updated the following sections: • Supported platforms • Supported hypervisors • Supported Nutanix packages • Enabling RDMA using Prism

Table of contents

1 Introduction

This document describes the deployment of the Remote Direct Memory Access (RDMA) solution.

1.1 Remote Direct Memory Access

RDMA increases bandwidth while lowering latency and CPU utilization. It sidesteps the system software stack components that process network traffic.

1.1.1 RDMA benefits

This section describes RDMA benefits.

1.1.1.1 Zero-Copy

Applications can perform data transfers without the involvement of the network software stack. Data is sent and received directly to the network card buffers without being copied between the network layers, bypassing TCP, which means that there is zero work to be done by the CPU; no caching, no context switching, resulting in lower overall load on the system as this frees the CPU(s) for other workloads.

1.1.1.2 Kernel bypass

Applications can perform data transfers directly from user-space without kernel involvement. Kernel packets are offloaded for the transport layer to the NIC; packet drops are addressed by using a lossless network using Priority-based Flow Control (PFC).

Applications access remote memory without consuming any CPU time in the remote server. The remote memory server (process) is read without any intervention from the remote processor. Likewise, the remote CPU cache(s) are not filled with the accessed memory content.

Supported platforms

The following are supported platforms:

- XC740-24 with NVMe
  - Intel rNDC and 2xCX-4 PCIe Adapters.
- XC940-24 with NVMe
  - Intel rNDC and 2xCX-4 PCIe Adapters

NOTE:
Prior to Foundation 4.4.1, the following applies:

- Two CX4 PCIe adaptors are required.
- CX4 rNDC cannot be mixed with a CX4 adapter.
- Use the Intel rNDC slot to populate the rNDC slot.
  rNDC cannot be used for RDMA.

1.1.2 Supported hypervisors

The following is the supported hypervisor:

- AHV version 20170830.184
- ESXi 6.7 U2 Build #13006603

1.1.3 Supported Nutanix packages

The following are supported Nutanix packages:

- AOS 5.9.2 and later
- Foundation 4.4.1 and later (supports rNDC and PCI) Absolutely minimum version of 4.3.2.

NOTE: Prior to Foundation 4.4.1, see section Troubleshooting 1.1.8, which requires no rNDC.

1.1.4 RDMA validate switches

The following switches are used to validate the RDMA Solution.

- Dell EMC S4048-ON
- Arista 7050QX, Arista 7050qx-32
- Mellanox SN2100

Contact the switch vendor to ensure compatibility with RDMA.

1.1.5 Minimum switch requirements

This section describes RDMA switch configuration.

To configure the switch:

1. Connect the RDMA supported NIC port to switch and configure as per RDMA pre-request.
2. Configure VLAN for corresponding trunk-based switch port.
3. Ensure that the General Flow control is off for send/receive mechanism config.
4. Enable priority flow control mode (PFC) with priority 3 or priority 4.
5. Switches are DCBX capable and enabled.
6. Ports mapped are configured with DCBX.

Refer to the Switch User Guide to enable these settings or contact the switch vendor for support.

1.1.6 Limitations

1.1.6.1 Update path

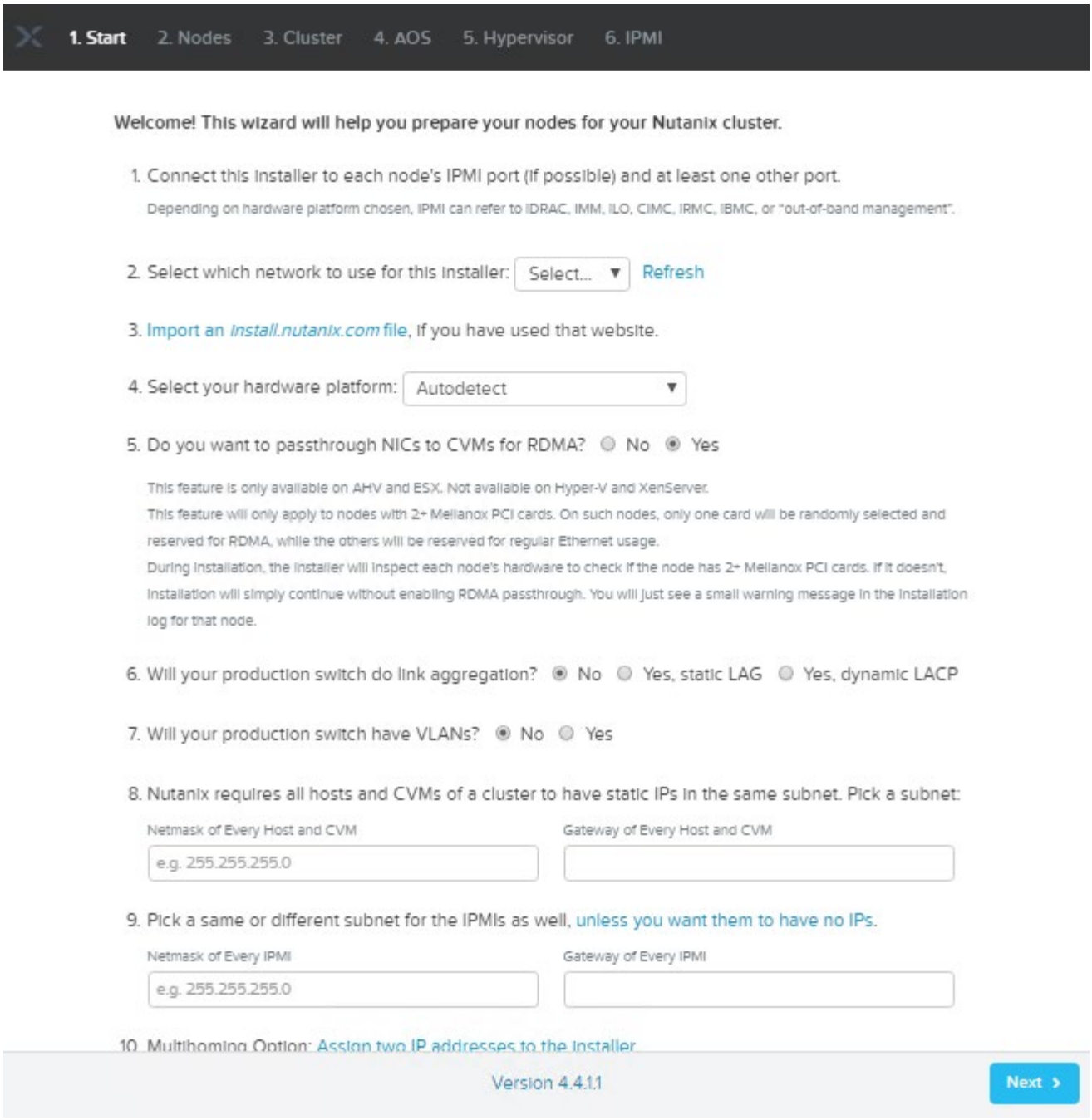
The cluster must have been created with Foundation 4.3.2 or later and AOS 5.9.2 or later. Otherwise, the interfaces are not created. Dell EMC recommends that you deploy with Foundation 4.4.1 or later.

1.1.6.2 Enabling RDMA using Prism

This section describes how to enable RDMA using Prism.

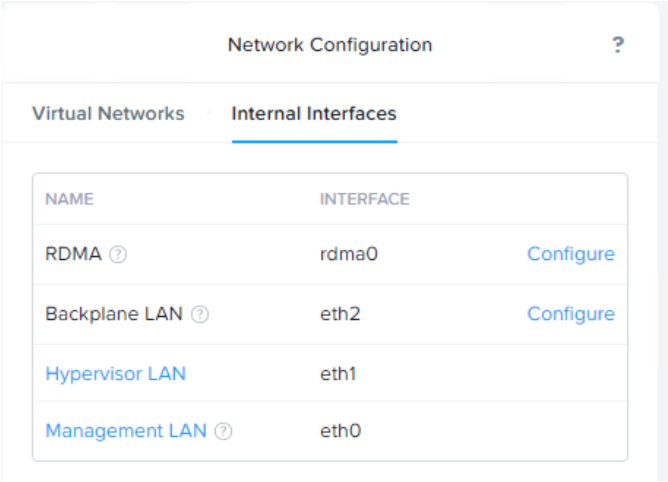
After the cluster is created with the supported hardware and software, use the follow these steps to enable RDMA.

1. Make sure to image the nodes with CX4 adapter interface connected to Mellanox switch – RDMA traffic.
2. Make sure to have one adapter port connected to management switch – Node imaging.
3. Make sure foundation version is 4.3.x and above.



rdma0 should display among the listed interfaces.

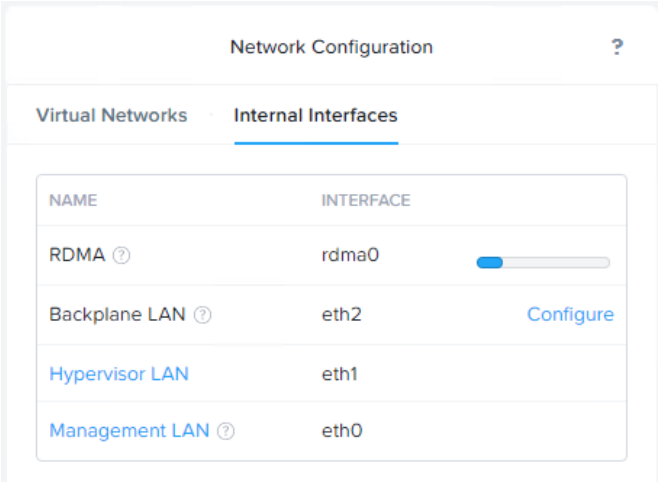
Alternatively, log in to the cluster prism UI navigate to **Settings>Network Configuration**.



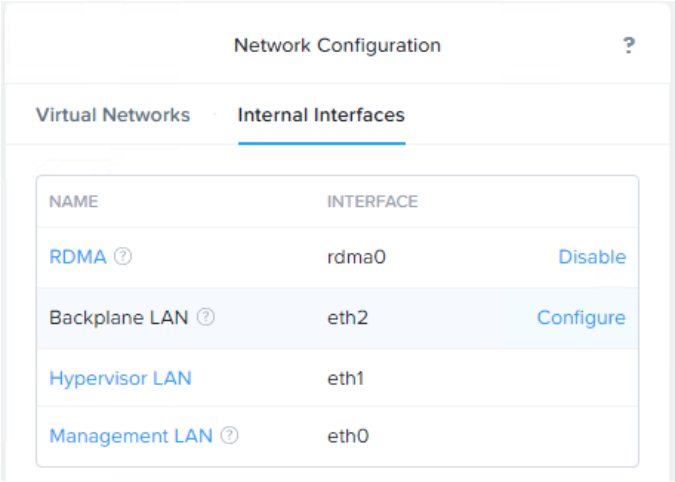
4. Enable RDMA on the cluster.
5. Click **Configure** rdma0 interface.
6. Input values as appropriate:

   - IP: xxx.xx.x.x (for example:172.16.0.0)
     NOTE: Do not use 192.168.5.0
   - Netmask: xxx.xx.xxx.x (for example: 255.255.252.0)
   - Vlan: xx (Vlan = 100)
   - PFC: 0 – 7 (for example 3 or 4)

7. Click to verify.



Successful RDMA configuration displays as *Disable* – meaning its enabled as shown below.



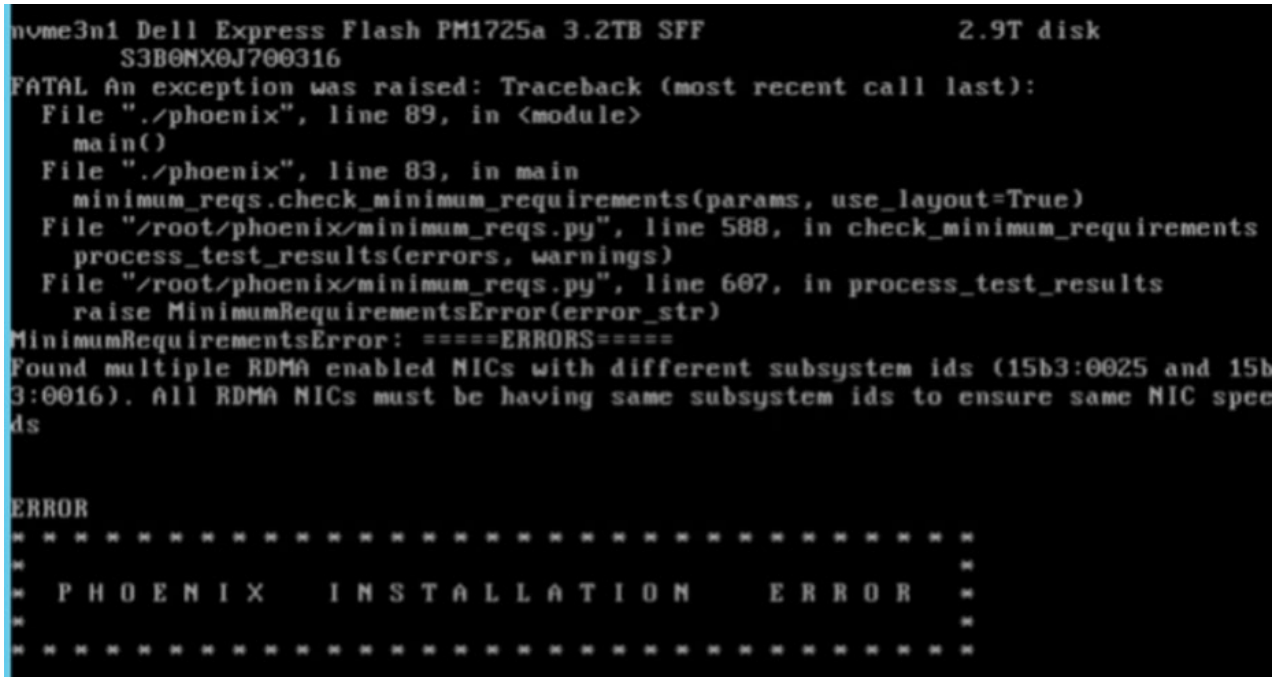
8. After the cluster is created, log in to each CVM and issue the command below to ensure the RDMA interface is created.

```
ip a |grep rdma0
```

```
nutanix@NTNX-7ZT2JL2-A-CVM:100.80.148.78:~$ allssh ip a |grep rdma0.100
10: rdma0.100@rdma0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc noqueue state UP
    inet 172.20.0.3/22 brd 172.20.3.255 scope global rdma0.100
10: rdma0.100@rdma0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc noqueue state UP
    inet 172.20.0.2/22 brd 172.20.3.255 scope global rdma0.100
10: rdma0.100@rdma0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc noqueue state UP
    inet 172.20.0.1/22 brd 172.20.3.255 scope global rdma0.100
nutanix@NTNX-7ZT2JL2-A-CVM:100.80.148.78:~$
```

1.1.7 Troubleshooting

1.1.7.1 Found multiple RDMA enabled NICs with different subsystem IDs



Solution: Make sure you have 2 CX4 Adapter on the supported hardware. Refer to the section for hardware requirements.

1. Check `ip a |grep rdma0` to make sure that interface rdma0 is created.
2. Make sure IPs match from the RDMA interfaces; `allssh ip a |grep rdma0` (check that the creation was successful and the interfaces were brought up).
3. Refer to **/home/nutanix/data/logs/genesis.out** for any failure.
4. For any reason RDMA is disabled, make sure it is enabled. See configuration steps on enabling interface.

For the RDMA architecture overview, refer to the Nutanix website.

https://portal.nutanix.com/

NOTE: Use Nutanix credentials to log in and then search for *RDMA*.

For information on *Configuring Network Segmentation on an Existing RDMA Cluster*, refer to the Nutanix KB article located here.

A Technical support and resources

Dell.com/support is focused on meeting customer needs with proven services and support.

Dell TechCenter is an online technical community where IT professionals have access to numerous resources for Dell EMC software, hardware and services.

Storage Solutions Technical Documents on Dell TechCenter provide expertise that helps to ensure customer success on Dell EMC Storage platforms.

A.1 Related resources

Provide a list of documents and other assets that are referenced in the paper; include other resources that may be helpful.