

SRIOV ON NFV - Version 6.0



Contents

Trademarks.....	3
Notes, Cautions, and Warnings.....	4
SRIOV Architecture in NFV 6.0.....	5
Compute Node reference topology.....	5
SRIOV script operational notes.....	5
Preparation for SRIOV enablement.....	7
Hardware Requirements.....	7
Hardware Deployment and Wiring.....	7
Network.....	8
Software.....	9
How to get the software.....	10
SRIOV script cnode pass.....	11
Pre-Requisite.....	11
Script parameter:.....	11
Steps to execute cnode pass.....	12
Validate cnode Pass.....	13
SRIOV script instance pass.....	14
Pre requisite.....	14
Steps to execute instance pass.....	14
Validate instance pass.....	14

Trademarks

© 2014-2016 Dell Inc. All rights reserved. Reproduction of this material in any manner whatsoever without the express written permission of Dell Inc. is prohibited. For more information, contact Dell.

Trademarks used in this text: Dell™, the DELL logo, PowerEdge™, and Dell Networking™ are trademarks of Dell Inc. Intel® and Xeon® are registered trademarks of Intel Corporation in the U.S. and other countries. Microsoft® and Windows® are registered trademarks of Microsoft Corporation in the United States and/or other countries.


Red Hat®, Red Hat Enterprise Linux®, and Ceph are trademarks or registered trademarks of Red Hat, Inc., registered in the U.S. and other countries. Linux® is the registered trademark of Linus Torvalds in the U.S. and other countries. Oracle® and Java® are registered trademarks of Oracle Corporation and/or its affiliates.


DISCLAIMER: The OpenStack® Word Mark and OpenStack Logo are either registered trademarks/service marks or trademarks/service marks of the OpenStack Foundation, in the United States and other countries, and are used with the OpenStack Foundation's permission. We are not affiliated with, endorsed or sponsored by the OpenStack Foundation or the OpenStack community.

The Midokura® name and logo, as well as the MidoNet® name and logo, are registered trademarks of Midokura SARL.

Other trademarks and trade names may be used in this publication to refer to either the entities claiming the marks and names or their products. Dell Inc. disclaims any proprietary interest in trademarks and trade names other than its own.

Notes, Cautions, and Warnings

 A **Note** indicates important information that helps you make better use of your system.

 A **Caution** indicates potential damage to hardware or loss of data if instructions are not followed.

 A **Warning** indicates a potential for property damage, personal injury, or death.

This document is for informational purposes only and may contain typographical errors and technical inaccuracies. The content is provided as is, without express or implied warranties of any kind.

SRIOV Architecture in NFV 6.0

Dell NFV SRIOV architecture provide a low latency and high throughput network performance in a Dell-RedHat NFV solution. Dell NFV SRIOV architecture provides wireline speed through Intel®X520 NIC adaptors.

Dell SRIOV solution consists of set of scripts that will automate the enablement of SRIOV virtual functions in Dell RedHat NFV solution 6.0. Dell SRIOV scripts provides full redundancy at each layer, i.e. VM level, NIC level and physical switch level redundancy.

Compute Node reference topology

Following is the compute node reference topology while enabling SRIOV in NFV 6.0.

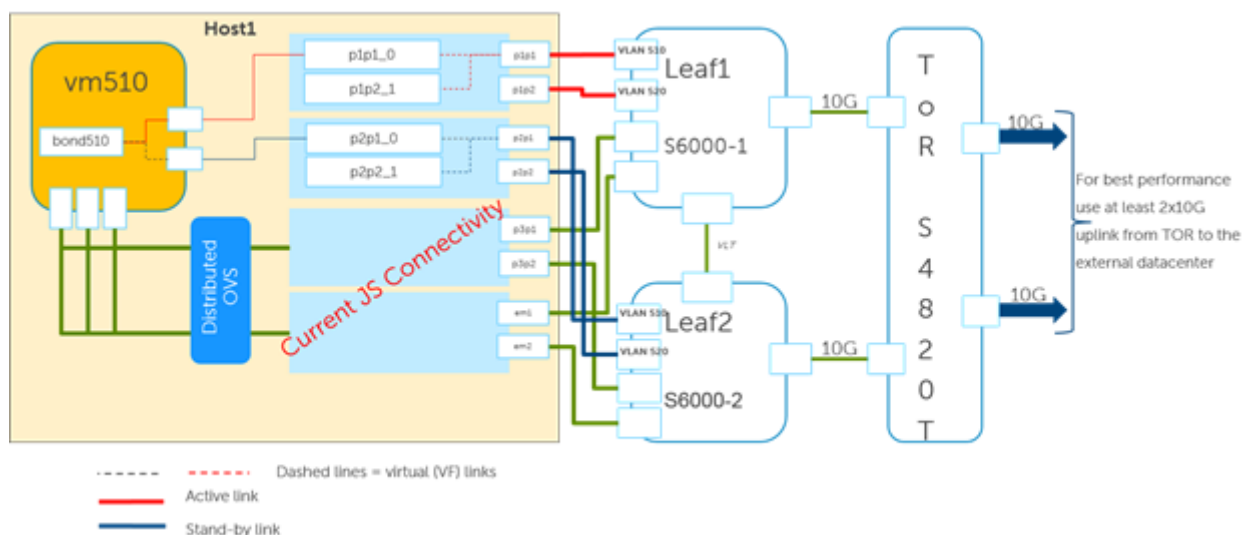


Figure 1: Reference Architecture for SRIOV using NFV 6.0

Notes :

1. P1P1, P1P2, P2P1 and P2P2 are the physical functions available on Host1 and Host2.
2. All the Virtual Functions and Physical Functions are available on compute nodes.
3. For best performance numbers, use at least 2x10Gbps uplinks from TOR to external datacenter.

SRIOV script operational notes

Followings are the operational recommendations for SRIOV scripts:

1. SRIOV scripts has 2 passes, cnode pass and instance pass.
2. "cnode" pass provides the fully automated enablement of SRIOV infrastructure on top of Dell Redhat NFV solution 6.0. This includes creating virtual functions, assigning bandwidth to each virtual function.
3. "instance" pass provides the fully automated method of attaching virtual functions to VMs and creating bonds inside the virtual machines for redundancy.
4. SRIOV scripts require a pair of two Physical Functions. PF in each pair should be part of different X520. For Example: In pair 1, first PF is P1P1 then the second PF should belong to slot P2, in this case, P2P1. Such pairing mechanism is used to provide NIC level redundancy, in case if one of the NIC goes down.

5. While assigning the VFs to a VM, script will assign first VF from the first PF in the pair, and second VF from the second PF in the pair.
6. SRIOV cnode pass also needs to have compute node name as input to the script in the settings file.
7. SRIOV instance pass needs to have instance name as input to the script.

Preparation for SRIOV enablement

Following are the preliminary list of pre-conditions for SRIOV to be installed:

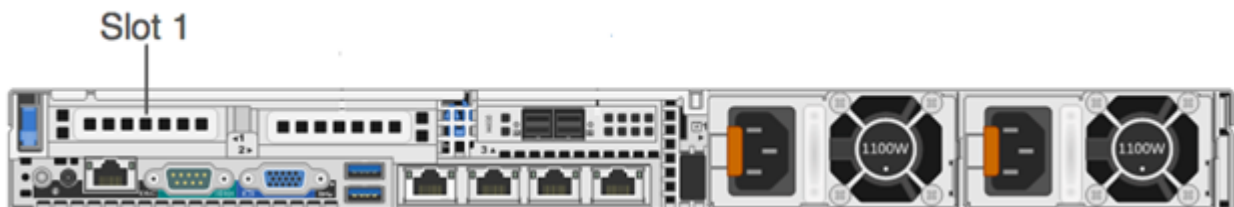
Hardware Requirements

1. NFV 6.0 full deployment, see the following document for NFV 6.0 deployment. Note: Assuming the compute node will consist of R630 or R730. This script is currently validated with Dell S6000 switches.
2. In addition to standard NFV 6.0 Hardware following additional Hardware is required:
 - a. 6 x Intel® X520 (2 X520 for each compute node)
 - a. 2 x Intel ® X520 will be installed in each compute node at following slots respectively:
 - P1
 - P2
 - b. 4 x QSFP connectors
 - a. 2 x QSFP connector will be connected to S6000-1
 - b. 2 x QSFP connector will be connected to S6000-1
 - c. 12 x SFP connectors
 - a. SFP connectors will be connected to each Intel ® X520 NIC ports.
 - d. 4 x QSFP to SFP breakout fiber cables, following is the cabling reference:

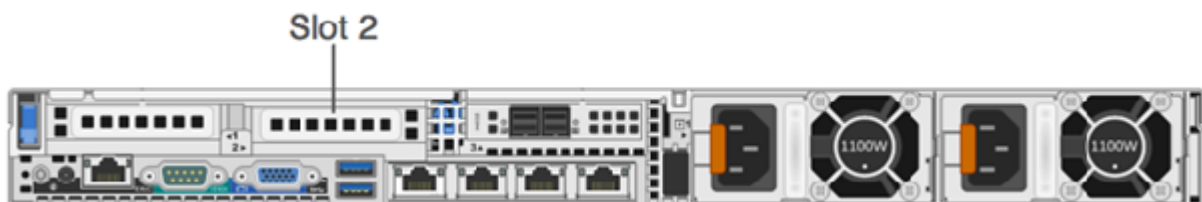
Hardware Deployment and Wiring

Please follow the steps below to deploy hardware and wiring reference:

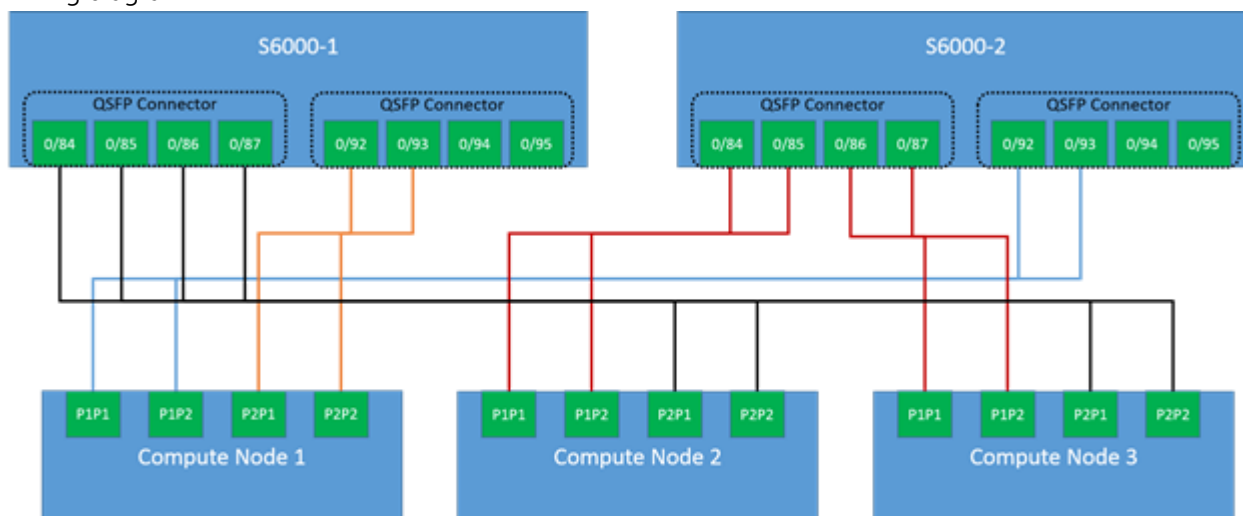
1. Check BIOS setting of each compute node, and make sure that "SRIOV Global Enable" is "Enables".
2. In BIOS, under "Processor Settings", "Virtualization Technology" flag should be enabled as well.
3. Insert Intel ® X520 in P1 slot of compute node 1, as shown below



4. Insert Intel ® X520 in P2 slot of compute node 1, as shown below



5. Repeat the steps (3) and (4) for compute 2 and 3 as well.
6. Now wire the newly added X520s in compute nodes to S6000 leaf switch, following is the reference wiring diagram:



- Each color represents a separate Fiber Break Out cable.
- Total of 4 Break Out cables are being used.

Figure 2: Wiring Topology between compute nodes and S6000 Leaf switches

Additional Wiring Notes

- a. 1x breakout cable will be connected to p1 of Compute Node 1
- b. 1x breakout cable will be connected to p2 of Compute Node 1
- c. 1x breakout cable will be connected to p1 of Compute Node 2 and 3
- d. 1x breakout cable will be connected to p2 of Compute Node 2 and 3
7. Double check the QSFP port coloring for each vendor and document that the solution is only validated against the particular one.
8. Connect S4820T switch uplink to external datacenter. It is highly recommended to allocate a dedicated 2x10Gbps link between S4820T and external datacenter.

Network

Following are the network requirements for the SRIOV enablement using NFV 6.0

1. Network Topology
 - a. Add a network diagram
 - b. VLAN topology
2. Switch configurations
 - a. Port config,
 - b. VLANs
 - c. Corresponding commands
3. PFs that are used for SRIOV should be connected to a physical switch.
 - a. Wiring Diagram
4. An Access VLAN should be configured on the corresponding switch port on which the PF is connected.

Software

1. Make sure ixgbe driver is loaded in compute host. If you are using RHEL 7.2, it will be loaded out of the box. If you are using some other version, make sure IXGBE is loaded.
2. The script will insert "intel_iommu=on" in the grub menu.
3. After all pre-condition, steps to execute scripts on director.
4. You need to have OSP admin privileges.
5. Access of RC files of under cloud and over cloud.
6. List of compute node names.
7. Which instance should you apply the script to? Instance Name is required.
8. VM should have one tenant network assigned to it, prior to script execution.

There is the script, 2 passes,


Pass 1 is Cnode pass (enable SRIOV on Cnode)

Pass 2 is Instance Pass (when ready to have SRIOV enable on Cnode, now execute Instance pass)

How to get the software

Following are 2 options:

1. Clone from Dell NFV repository
 - a. Tell them how to clone the scripts
2. Ask your Dell representative

 **Note** : The script should be available on Director Node.

SRIOV script cnode pass

Pre-Requisite

Followings are the items that needs to be completed before we start executing Pass-1 of the scripts:

1. Make sure, Dell Redhat NFV solution 6.0 is deployed successfully and all the OpenStack services are up and running.
2. Make sure Compute Nodes are equipped with additional hardware as mentioned in Hardware reference section.
3. Make sure the wiring of additional NIC adaptors to physical Dell S6000 switches are according to the wiring reference mentioned above.
4. You must be logged in as "osp_admin" to execute the scripts.
5. Scripts must be cloned from the upstream repo of Dell NFV.

For more detail, refer to "How to get Software" section.

Best Practice:

- a. After modifying the .ini files, make sure to name the setting files according to the date/time/hardware_details etc. So it can be referenced easily for debugging.
- b. It is helpful to verify that SSH is working b/w director and control/compute. You can go to director node and run the following command from there.

```
ssh heat-admin@ compute-node-ip>
```

This will validate the SSH connectivity between director node and compute node.

Script parameter:

Followings are the script parameters that will be input in the sriov_settings_pass-cnode_R141-compute-1.ini file.

No.	Parameter	Supported values	Description/Example
1	settingsVersion	1	This parameter reflects the version of the settings file that is being used. Only supported value is 1. Example: settingsVersion=1
2	scriptName	enable_sriov	This parameter represents the script name that is being run. Only supported value is enable_sriov. Example: scriptName=enable_sriov
3	scriptVersion	0.01	This parameter is the version of the script that is being executed. Only supported value is '0.01' Example: scriptVersion=0.01
4	RHOSPVersion	9	This parameter is the RedHat OpenStack version on which scripts are being executed. Currently, only supported version is 9 Example: RHOSPVersion=9

No.	Parameter	Supported values	Description/Example
5	JSVersion	6.0	This specifies the Jet Stream version on which script are being executed. Only supported JetStream version is 6.0. Example: JSVersion=6.0
6	scriptPass	"cnode" "instance_pass"	This parameter is the argument to the script to run on compute node or VM instance. First pass to the script should be "cnode" pass, this will create virtual functions and enable compute nodes for SRIOV. Example: scriptPass=cnode
7	scriptPassMode	"ephemeral" "persistent"	This parameter describes weather the SRIOV and virtual functions will be ephemeral or persistent across multiple reboots of compute nodes. The default value is "ephemeral". Example for ephemeral: scriptPassMode=ephemeral Example for persistent: scriptPassMode=persistent
8	num_vfs_per_pf	'2'-'64'	This parameter reflects the number of virtual functions created per physical functions. The minimum value is 2 and maximum can be 64. However, these number can vary depending on the NIC adaptor being used. Default value in the script is 4 Example: num_vfs_per_pf=4
9	PFPair1pf1	interface name	This parameter is to define the logical pairs of physical functions that script creates in order to provide NIC level redundancy. Two PF pairs are created by script. Each pair consists of 2 PF for redundancy. Example PF Pair 1 PFPair1pf1=p1p1 PFPair1pf2=p2p1 Similarly, example for PF Pair 2: PFPair2pf1=p1p2 PFPair2pf2=p2p2
10	bw_available_per_pf	10	This parameter describes the bandwidth (in Gbps) available for each physical function. The scripts has been only tested for Intel®X520 dual port adaptors. Therefore, the only supported value is 10. Example: bw_available_per_pf=10
11	bw_strategy	equal	This parameter reflects the distribution among virtual functions. Example: bw_strategy=equal
12	cnodeName	Compute Node Name	This parameter contains the name of the compute node on which the SRIOV enablement is being triggered. Example: cnodeName=R141-compute-1

Steps to execute cnode pass

Followings are the steps that will be followed in the order to execute cnode pass scripts:

1. cnode pass the script requires following arguments described below:
 - a. **--settings_file:** This argument requires an input of settings file name. The file contains all the parameters described in the parameter section in above section. **Where to find this file:** This file comes with script software package cloned from github repository. The file contains default values for standard Dell RedHat NFV solution 6.0.
 - b. **--ucrc:** Under Cloud RC files, it contains the authentication URL and credentials for Under Cloud Director node. **Where to find this file:**

SRIOV script instance pass

Pre requisite

1. SRIOV script pass-1 must be successfully completed.
2. Virtual functions are successfully created and assigned to Virtual Machines.
3. Following is the explanation of the attributes of the ".properties" file that will be used during this pass.

Steps to execute instance pass

1. Run the script with following arguments to execute instance pass

Validate instance pass

1. Make sure the virtual functions bonds are successfully created inside the virtual machines.
2. VM is able to send and receive 2.5GB/VF to physical network.
3. Verify SRIOV HA by pulling the cable of one of PF from the physical switch and make sure VM is still able to send/receive traffic.