

How to Configure Intel[®] X520 Ethernet Server Adapter Based Virtual Functions on Citrix* XenServer 6.0*

Technical Brief v1.0

December 2011



INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL® PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER, AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT. Intel products are not intended for use in medical, life saving, life sustaining, critical control or safety systems, or in nuclear facility applications.

Intel may make changes to specifications and product descriptions at any time, without notice.

Intel Corporation may have patents or pending patent applications, trademarks, copyrights, or other intellectual property rights that relate to the presented subject matter. The furnishing of documents and other materials and information does not provide any license, express or implied, by estoppel or otherwise, to any such patents, trademarks, copyrights, or other intellectual property rights.

IMPORTANT - PLEASE READ BEFORE INSTALLING OR USING INTEL® PRE-RELEASE PRODUCTS.

Please review the terms at http://www.intel.com/netcomm/prerelease_terms.htm carefully before using any Intel® pre-release product, including any evaluation, development or reference hardware and/or software product (collectively, "Pre-Release Product"). By using the Pre-Release Product, you indicate your acceptance of these terms, which constitute the agreement (the "Agreement") between you and Intel Corporation ("Intel"). In the event that you do not agree with any of these terms and conditions, do not use or install the Pre-Release Product and promptly return it unused to Intel.

Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined." Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them.

Products may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Code names presented in this document are only for use by Intel to identify products, platforms, programs, services, etc. ("products") in development by Intel that have not been made commercially available to the public, i.e., announced, launched or shipped. They are never to be used as "commercial" names for products. Also, they are not intended to function as trademarks.

Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order.

Copies of documents which have an order number and are referenced in this document, or other Intel literature may be obtained by calling 1-800-548-4725 or by visiting Intel's website at <http://www.intel.com>.

Intel and the Intel logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

*Other names and brands may be claimed as the property of others.

Copyright © 2011, Intel Corporation. All Rights Reserved.



Contents

1.0	Introduction	5
1.1	Hardware Requirements	5
1.2	Software Requirements	5
2.0	Installation and Configuration	5
2.1	Enabling I/O Memory Management Unit (IOMMU)	5
2.2	Intel® X520 Ethernet Server Adapter Driver Configuration	7
2.3	Preparing Virtual Functions for Assignment	8
2.4	Virtual Machine Installation and Configuration	10
2.5	Enabling Hot-Plug Device Support.....	14
3.0	Summary	17
4.0	Customer Support	17
5.0	For Product Information	17



Revision History

Date	Revision	Description
December 2011	1.0	Initial release



1.0 Introduction

Processor and platform advances, along with progress in operating systems and applications, have spurred data centers to quickly adopt 10 Gigabit Ethernet (10 GbE) as the standard for interconnects. The Intel® Ethernet Server Adapter X520 family of adapters introduces numerous industry-leading features that are helping data center administrators implement innovative solutions for difficult and challenging connectivity problems. I/O Virtualization is one of the fastest growing usage models within the data center. The X520 family of adapters provides the ability to create Virtual Functions (VFs) that are identical instantiations of the Physical Functions (PFs). VFs are capable of providing 10 GbE connectivity to Virtual Machines (VMs) within a virtualized operating system framework. This document shows how to make use of VFs using Citrix XenServer* 6.0.

1.1 Hardware Requirements

- An Intel® Ethernet X520 Server Adapter
- A server platform that supports Intel® Virtualization Technology for Directed I/O (VT-d) and the PCI-SIG* Single Root I/O Virtualization and Sharing (SR-IOV) specification¹
- A server platform with an available PCI Express*:X8 5.0 Gb/s (Gen2) slot

1.2 Software Requirements

- Citrix XenServer* 6.0

2.0 Installation and Configuration

1. Install the Intel® Ethernet X520 server adapter in an available PCI-Express x8 slot. (Ensure that the x8 slot is electrically connected as a x8; some slots are physically x8 but electrically support only x4. Verify this with your server manufacturer or system documentation.)
2. Power up the server.
3. Enter the server's BIOS setup and make sure the virtualization technology and Intel® VT-d features are enabled.
4. Install Citrix XenServer 6.0.
5. Log in to the newly-installed Citrix XenServer operating system using the Citrix XenCenter application.

2.1 Enabling I/O Memory Management Unit (IOMMU)

I/O Memory Management Unit (IOMMU) support is not enabled by default in Citrix XenServer 6.0. IOMMU support is required for a VF to function properly when assigned to a VM. The following kernel boot parameters are required to enable IOMMU support for the Citrix XenServer 6.0 kernel:

```
iommu=1
```

This parameter can be appended to the kernel boot entry by executing the following command at the XenServer console. See Figure 1.

```
/opt/xensource/libexec/xen-cmdline --set-xen iommu=1
```

```
root@xenserver6:~  
  
label xe-serial  
# XenServer (Serial)  
kernel mboot.c32  
append /boot/xen.gz com1=115200,8n1 console=com1,vga dom0_mem=752M lowmem_emergency_po  
ol=1M crashkernel=64M@32M iommu=1 --- /boot/vmlinuz-2.6-xen root=LABEL=root-udfgquzg ro  
console=tty0 xencons=hvc console=hvc0 --- /boot/initrd-2.6-xen.img  
  
label safe  
# XenServer in Safe Mode  
kernel mboot.c32  
append /boot/xen.gz nosmp noreboot noirqbalance acpi=off noapic dom0_mem=752M com1=115  
200,8n1 console=com1,vga --- /boot/vmlinuz-2.6-xen nousb root=LABEL=root-udfgquzg ro con  
sole=tty0 xencons=hvc console=hvc0 --- /boot/initrd-2.6-xen.img  
  
label fallback  
# XenServer (Xen 4.1.1 / Linux 2.6.32.12-0.7.1.xs6.0.0.529.170661xen)  
kernel mboot.c32  
append /boot/xen-4.1.1.gz dom0_mem=752M lowmem_emergency_pool=1M crashkernel=64M@32M i  
ommu=1 --- /boot/vmlinuz-2.6.32.12-0.7.1.xs6.0.0.529.170661xen root=LABEL=root-udfgquzg  
ro xencons=hvc console=hvc0 console=tty0 --- /boot/initrd-2.6.32.12-0.7.1.xs6.0.0.529.17  
0661xen.img  
  
label fallback-serial  
# XenServer (Serial, Xen 4.1.1 / Linux 2.6.32.12-0.7.1.xs6.0.0.529.170661xen)  
kernel mboot.c32  
append /boot/xen-4.1.1.gz com1=115200,8n1 console=com1,vga dom0_mem=752M lowmem_emerge  
ncy_pool=1M crashkernel=64M@32M iommu=1 --- /boot/vmlinuz-2.6.32.12-0.7.1.xs6.0.0.529.17  
0661xen root=LABEL=root-udfgquzg ro console=tty0 xencons=hvc console=hvc0 --- /boot/init  
rd-2.6.32.12-0.7.1.xs6.0.0.529.170661xen.img  
[root@xenserver6 ~]#
```

Figure 1. Enabling “iommu” by Appending It to the Boot Kernel Entry

Figure 1 shows the contents of the /boot/extlinux.conf file. Reboot the server for the iommu change to take effect (required).

Upon a successful server reboot, use the dmesg command at the XenServer console to confirm iommu is successfully initialized. See Figure 2.



```

root@xenserver6/
(XEN) Initiating memory sharing.
(XEN) Intel VT-d Snoop Control not enabled.
(XEN) Intel VT-d Dom0 DMA Passthrough not enabled.
(XEN) Intel VT-d Queued Invalidation enabled.
(XEN) Intel VT-d Interrupt Remapping enabled.
(XEN) Intel VT-d Shared EPT tables not enabled.
(XEN) [VT-D]iommu.c:1993: ioapic_to_iommu: ioapic 0x3 (id: 0x4) is NULL! Will not try to enable Interrupt Remapping

(XEN) I/O virtualisation enabled
(XEN) - Dom0 mode: Relaxed
(XEN) ENABLING IO-APIC IROs
(XEN) -> Using new ACK method
(XEN) Platform timer is 14.318MHz HPET
(XEN) Allocated console ring of 32 KiB.
(XEN) VMX: Supported advanced features:
(XEN) - APIC MMIO access virtualisation
(XEN) - APIC TPR shadow
(XEN) - Extended Page Tables (EPT)
(XEN) - Virtual-Processor Identifiers (VPID)
(XEN) - Virtual NMI
(XEN) - MSR direct-access bitmap
(XEN) EPT supports ZME super page.
(XEN) HVM: ASIDs enabled.
(XEN) HVM: VMX enabled
(XEN) HVM: Hardware Assisted Paging detected and enabled.
(XEN) Brought up 16 CPUs
(XEN) Testing NMI watchdog --- CPU#0 okay, CPU#1 okay, CPU#2 okay, CPU#3 okay, CPU#4 okay, CPU#5 okay, CPU#6 okay,
CPU#7
okay, CPU#8 okay, CPU#9 okay, CPU#10 okay, CPU#11 okay, CPU#12 okay, CPU#13 okay, CPU#14 okay, CPU#15 okay.
(XEN) *** LOADING DOMAIN 0 ***
(XEN) Xen kernel: 64-bit, lsb, compat32
(XEN) Dom0 kernel: 32-bit, PAE, lsb, paddr 0x100000 -> 0x5f5000
(XEN) PHYSICAL MEMORY ARRANGEMENT:
(XEN) Dom0 alloc.: 00000000d0000000->00000000be000000 (186712 pages to be allocated)
(XEN) Init. ramdisk: 0000000033f958000->0000000033ffffc00
(XEN) VIRTUAL MEMORY ARRANGEMENT:
(XEN) Loaded kernel: 00000000c0100000->00000000c05f5000
(XEN) Init. ramdisk: 00000000c05f5000->00000000c0c9cc00
(XEN) Phys-Mach map: 00000000c0c9d000->00000000c0d59000
(XEN) Start info: 00000000c0d59000->00000000c0d594b4
(XEN) Page tables: 00000000c0d5a000->00000000c0d68000
(XEN) Boot stack: 00000000c0d68000->00000000c0d69000
(XEN) TOTAL: 00000000c0000000->00000000c1000000
(XEN) ENTRY ADDRESS: 00000000c0100000
(XEN) Dom0 has maximum 16 VCPUs
(XEN) Scrubbing Free RAM: .....done.
(XEN) Xen trace buffers: disabled
(XEN) Std. LogLevel: Errors and warnings
(XEN) Guest LogLevel: Nothing (Rate-limited: Errors and warnings)
(XEN) *** Serial input -> DOM0 (type 'CTRL-a' three times to switch input to Xen)
(XEN) Freed 216kB init memory.
(XEN) _csched_vcpu_acct_start: setting dom 0 as the privileged domain
root@xenserver6 /]#
    
```

Figure 2. XenServer Console Display Showing Successful "iommu" Initialization

2.2 Intel® X520 Ethernet Server Adapter Driver Configuration

1. Citrix XenServer 6.0 includes the X520 Server Adapter driver. XenServer loads the driver at boot time and creates 40 virtual functions (VFs) per port by default. The X520 Server Adapter supports up to 63 VFs per port.
2. The number of VFs created can be adjusted by using the `ixgbe` driver load-time parameter called `"max_vfs."` This parameter is located in the driver configuration file called `"ixgbe."` The `"ixgbe"` file is located in the `/etc/modprobe.d/` folder in Citrix XenServer 6.0. Figure 3 shows the default `ixgbe` configuration file contents:

```

root@xenserver6:/etc/modprobe.d
[root@xenserver6 ~]# cd /etc/modprobe.d/
[root@xenserver6 modprobe.d]# ls
blacklist          blacklist-ide          bonding             modprobe.conf.dist
blacklist-be2iscsi blacklist-ipmi         disable-ipv6       qisioctl
blacklist-bridge   blacklist-vfunc-drivers igb
blacklist-compat   bnx2x                 ixgbe
[root@xenserver6 modprobe.d]#
[root@xenserver6 modprobe.d]#
[root@xenserver6 modprobe.d]# cat ixgbe
options ixgbe max_vfs=40,40
[root@xenserver6 modprobe.d]#

```

Figure 3. Default Contents of the “ixgbe” Configuration File

Any warnings, errors and information messages during the `ixgbe` driver load are logged in the `/var/log/messages` file. It is a good practice to review this file to confirm that the driver loaded successfully without warnings or errors.

Use the `lspci` command to confirm that the VF was successfully created. An example entry:

```
#lspci | grep 82599
```

Figure 4 shows the results of this command.

```

root@xenserver6/
[root@xenserver6 ~]# lspci | grep 82599
03:00.0 Ethernet controller: Intel Corporation 82599EB 10-Gigabit SFI/SFP+ Network Connection (rev 01)
03:00.1 Ethernet controller: Intel Corporation 82599EB 10-Gigabit SFI/SFP+ Network Connection (rev 01)
03:10.0 Ethernet controller: Intel Corporation 82599 Ethernet Controller Virtual Function (rev 01)
03:10.1 Ethernet controller: Intel Corporation 82599 Ethernet Controller Virtual Function (rev 01)
03:10.2 Ethernet controller: Intel Corporation 82599 Ethernet Controller Virtual Function (rev 01)
03:10.3 Ethernet controller: Intel Corporation 82599 Ethernet Controller Virtual Function (rev 01)
03:10.4 Ethernet controller: Intel Corporation 82599 Ethernet Controller Virtual Function (rev 01)
03:10.5 Ethernet controller: Intel Corporation 82599 Ethernet Controller Virtual Function (rev 01)
03:10.6 Ethernet controller: Intel Corporation 82599 Ethernet Controller Virtual Function (rev 01)
03:10.7 Ethernet controller: Intel Corporation 82599 Ethernet Controller Virtual Function (rev 01)
03:11.0 Ethernet controller: Intel Corporation 82599 Ethernet Controller Virtual Function (rev 01)
03:11.1 Ethernet controller: Intel Corporation 82599 Ethernet Controller Virtual Function (rev 01)
03:11.2 Ethernet controller: Intel Corporation 82599 Ethernet Controller Virtual Function (rev 01)
03:11.3 Ethernet controller: Intel Corporation 82599 Ethernet Controller Virtual Function (rev 01)
03:11.4 Ethernet controller: Intel Corporation 82599 Ethernet Controller Virtual Function (rev 01)
03:11.5 Ethernet controller: Intel Corporation 82599 Ethernet Controller Virtual Function (rev 01)
03:11.6 Ethernet controller: Intel Corporation 82599 Ethernet Controller Virtual Function (rev 01)
03:11.7 Ethernet controller: Intel Corporation 82599 Ethernet Controller Virtual Function (rev 01)
03:12.0 Ethernet controller: Intel Corporation 82599 Ethernet Controller Virtual Function (rev 01)
03:12.1 Ethernet controller: Intel Corporation 82599 Ethernet Controller Virtual Function (rev 01)

```

Figure 4. VFs Per Port: Bus #, Device #, and Function #

Figure 4 shows the VFs per port for the X520 Server Adapter. Each VF is identified by a unique bus, device, and function number. In the example, the first VF is assigned Bus #3, Device #10, Function #0. VFs ending with an even function number belong to PF 0, that is, port 0. VFs ending with an odd function number belong to PF 1, that is, port 1.

2.3 Preparing Virtual Functions for Assignment

Citrix XenServer 6.0 requires SR-IOV-compliant, X520 Server Adapter-based Virtual Devices (VFs) not to be claimed by the Xen Domain-0 if the VFs are intended to be assigned to Virtual Machines. VF bus, device and function numbers are required for



preparing VFs for Virtual Machine assignment. Execute the following two commands at the XenServer console for each VF that is intended for Virtual Machine assignment. See Figure 5.

```
echo 0000:03:10.2 > /sys/bus/pci/drivers/pciback/new_slot
echo 0000:03:10.2 > /sys/bus/pci/drivers/pciback/bind
```

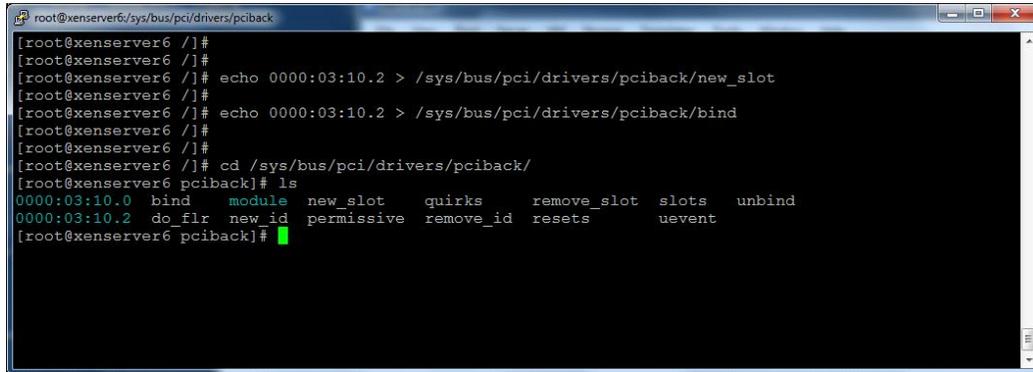


Figure 5. Results of new_slot and bind Commands With Link Node Information for the VF Identified by Bus #3, Device #10, and Function #2

Figure 5 also shows the link node for the VF identified by Bus # 3, Device # 10 and Function #2 that was created as a result of the commands. Citrix XenServer 6.0 will create these nodes for all the VFs that are intended for VM assignment

Note: Intel strongly recommends creating a script with the two commands above for all the VFs that are intended for VM assignment. This script should run every time XenServer is rebooted.

Use the following command to see available VFs and their identification attributes for assignment to a Virtual Machine.

```
#xl pci-list-assignable
```

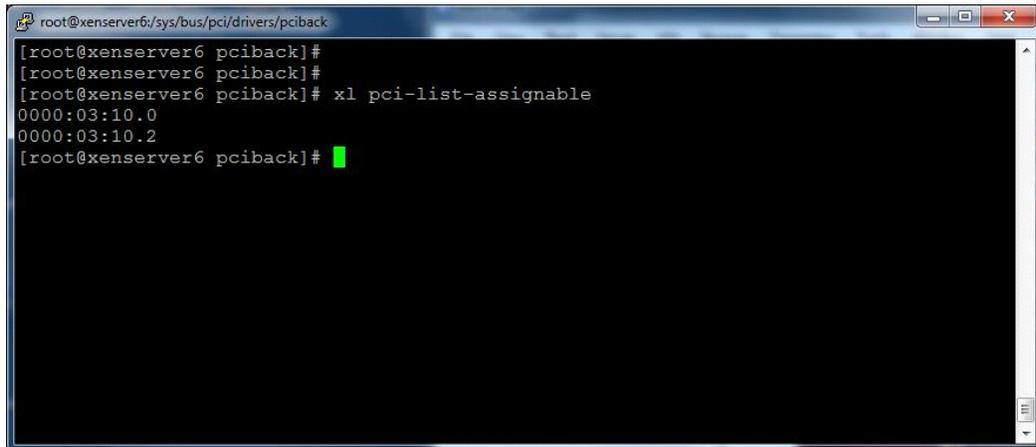


Figure 6. Display Showing Available VFs and Their ID Attributes

Figure 6 above shows the list of VFs that are available for assignment to Virtual Machines and their respective bus, device and function numbers.

2.4 Virtual Machine Installation and Configuration

Citrix XenServer 6.0 includes tools for creating and managing VMs. These tools offer both Command Line Interface (CLI) and Graphical User Interface (GUI). Citrix XenCenter is a GUI tool for creating and managing VMs.

1. Use Citrix XenCenter to create a VM.
2. Install the operating system of choice on the newly created VM. For the purposes of this document, Red Hat Enterprise Linux* was installed in the VM. See Figure 7 for an example:

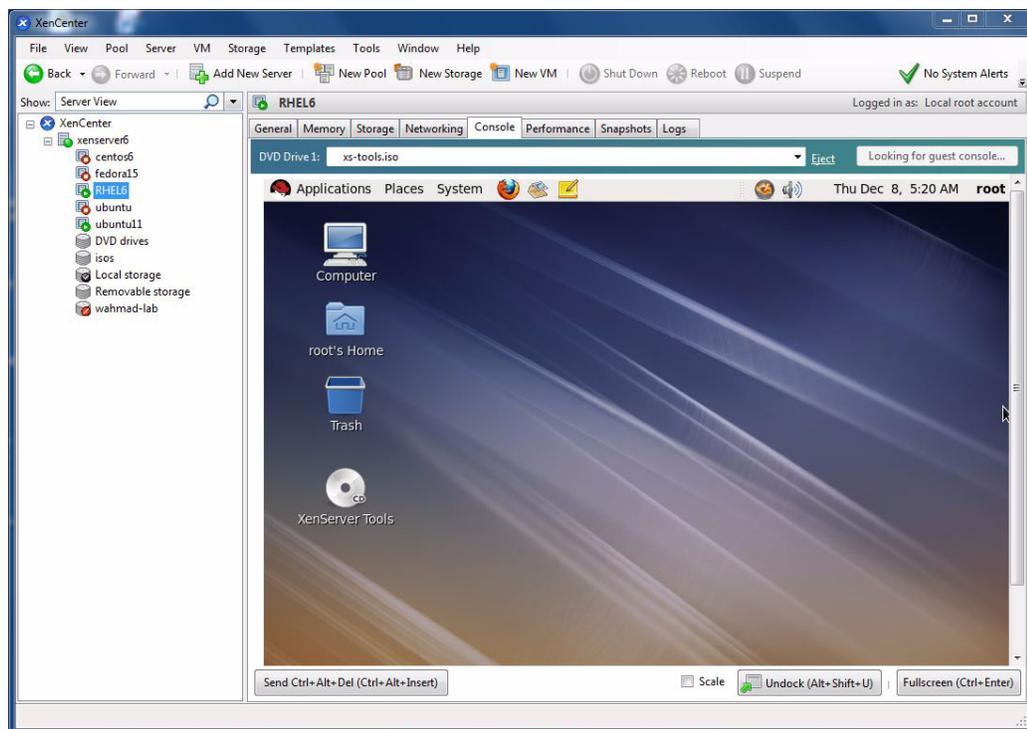


Figure 7. Red Hat Enterprise Linux* Installation Example

3. Citrix XenCenter doesn't provide a method via the GUI to assign PCI devices to Virtual Machines. Use the XenServer console commands below to list all available Virtual Machines and to assign VFs to a VM.

```
#xl list-vm
```

```
#xe vm-param-add other-config:pci=0/0000:03:10.0 uuid=bd5...
```



```

root@xenserver6:~#
[root@xenserver6 ~]# xl list-vm
      UUID                                ID   name
e880725c-895b-eb83-cbd8-b6494956152b    2   RHEL6
bdf5a0f2-4c60-9f85-ef11-b9a49a13c836    5   ubuntu
[root@xenserver6 ~]#
[root@xenserver6 ~]#
[root@xenserver6 ~]# xl pci-list-assignable-devices
0000:03:10.0
0000:03:10.2
0000:03:10.4
0000:03:10.6
[root@xenserver6 ~]#
[root@xenserver6 ~]#
[root@xenserver6 ~]# xe vm-param-set other-config:pci=0/0000:03:10.0 uuid=bdf5a0f2-4c60-9f85-ef11-b9a49a13c836
[root@xenserver6 ~]#

```

Figure 8. List of VMs, VFs Available, and Assignment to VM "RHEL6"

Figure 8 above shows the list of Virtual Machines configured on the server, the list of VFs available for assignment and subsequent VF assignment to Virtual Machine named "RHEL6."

4. Power up the RHEL6 VM. Log into the VM using the credentials created during the VM installation process. The VM must be powered down and powered back up in case the VF assignment took place while the VM was running.
5. Open the Linux Terminal. Use the Linux `lspci` utility to confirm that the assigned VF is visible within the VM's PCI-Express hierarchy. See Figure 9:

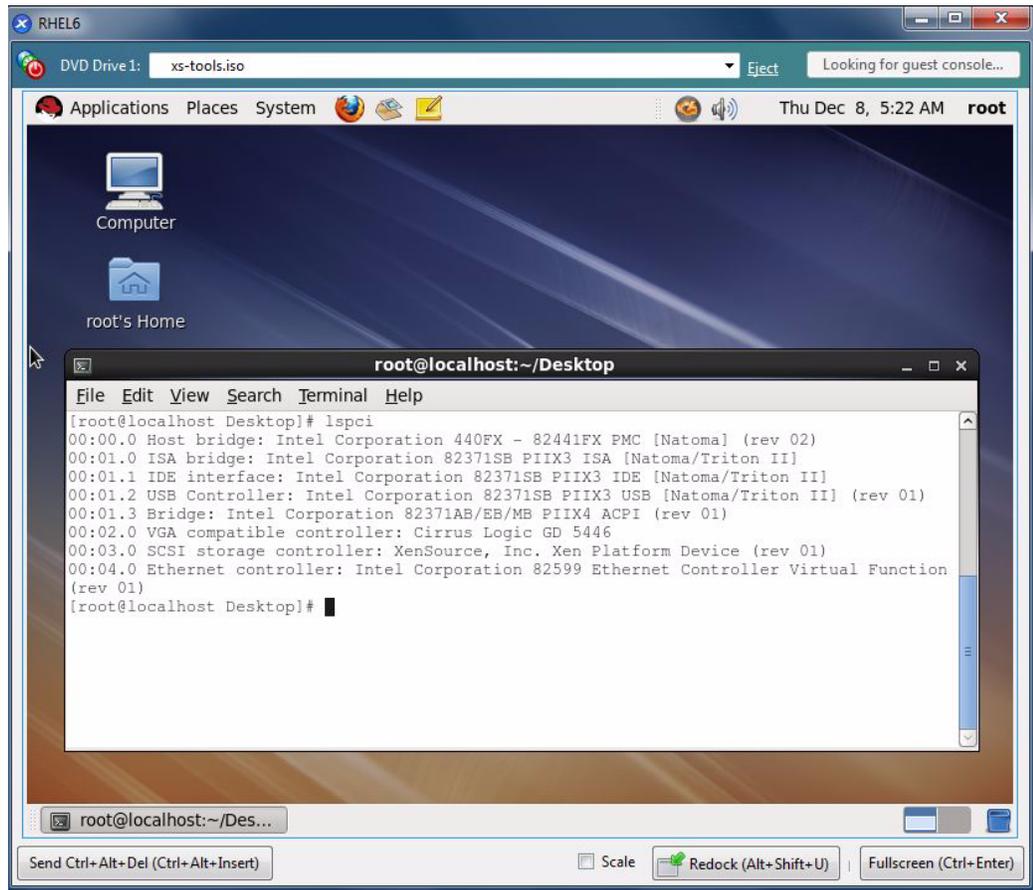


Figure 9. Linux Terminal Running the “lspci” Utility

6. Use the Linux `lsmod` utility to confirm that the driver for the VF has loaded successfully, as shown in Figure 10:

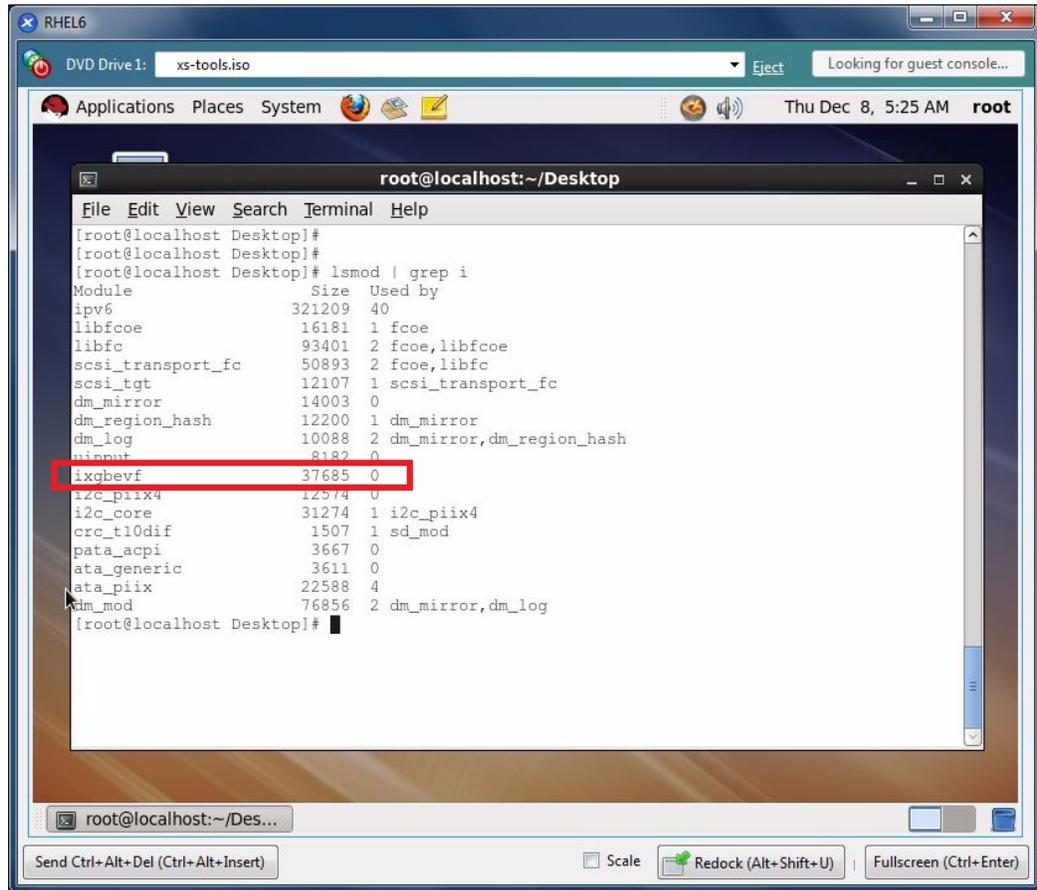


Figure 10. Linux Terminal Using “lsmod” Utility to Show Successful Driver Load

7. Use the Linux ifconfig utility to confirm that the newly assigned VF is ready for use (Figure 11).

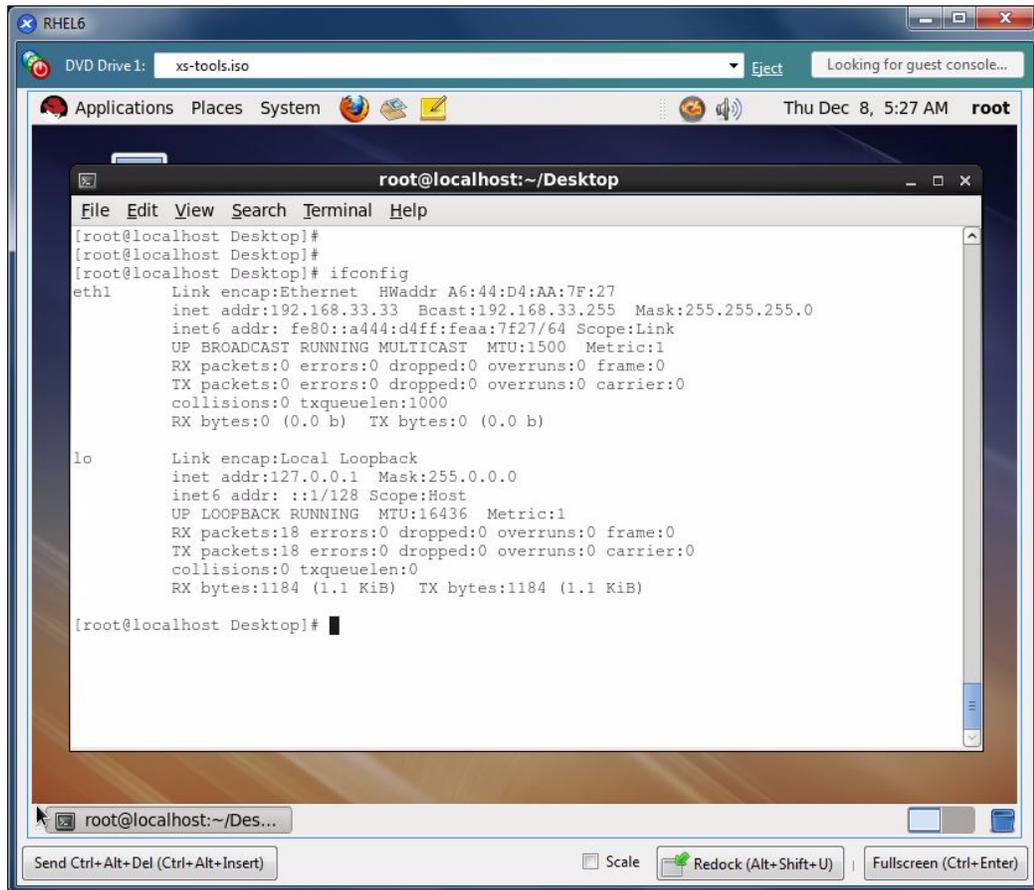


Figure 11. Linux Terminal Using "ifconfig" Utility Confirming Assigned VF Is Ready

The VF can be configured for DHCP or static IP address assignment. The VF is ready to communicate once it has an IP address assigned.

2.5 Enabling Hot-Plug Device Support

Hot plug device assignment to Virtual Machines is not enabled by default in Citrix XenServer 6.0. It is possible to enable hot pluggable device assignment support to Virtual Machines in XenServer 6.0. Citrix doesn't support this feature at this time; use this feature at your own risk.

- Hot plug device assignment can be enabled by editing the "qemu-dm-wrapper" file that is located in the /opt/xensource.libexec/ folder. Add the following line in "def main(argv) :" function, as shown in Figure 12:



```
xenserver6 server console

from resource import getrlimit, RLIMIT_CORE, setrlimit

limit = 64 * 1024 * 1024
oldlimits = getrlimit(RLIMIT_CORE)
setrlimit(RLIMIT_CORE, (limit, oldlimits[1]))
return limit

def main(argv):
    import os

    qemu_env = os.environ
    qemu_dm = '/usr/lib/xen/bin/qemu-dm'
    domid = int(argv[1])
    qemu_args = ['qemu-dm-%d'%domid] + argv[2:]

    qemu_args.append("-priv")

    if is_sdk() is True:
        return fake_dm(domid)

    print "qemu-dm-wrapper in python:"
    print "Using domid: %d" % domid
    print "Arguments: %s" % " ".join(argv[1:])

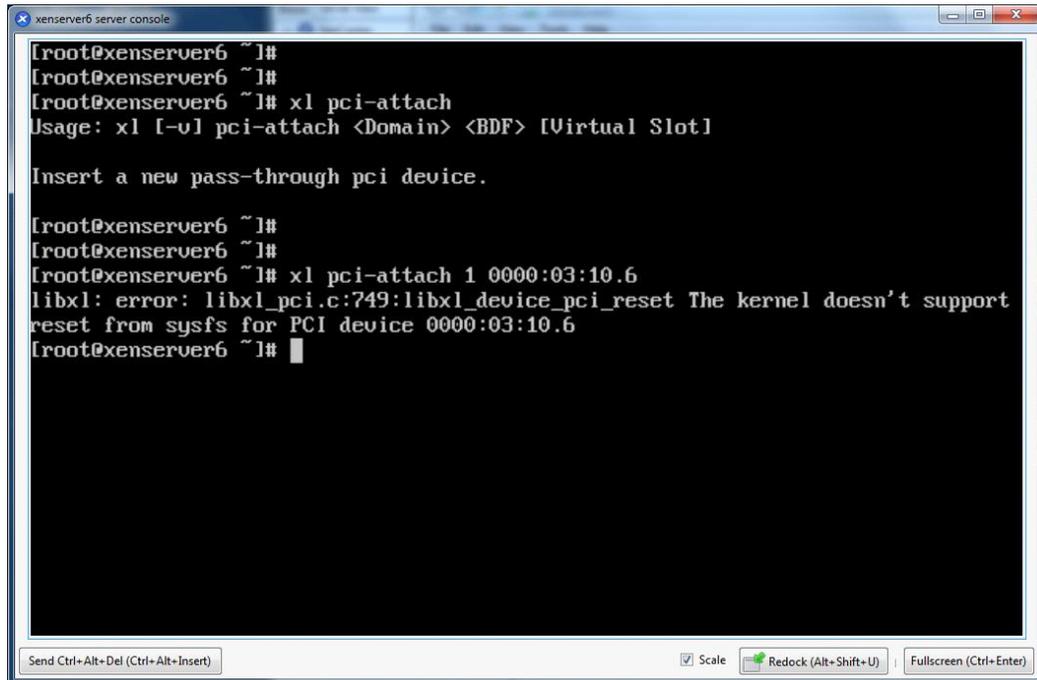
-- INSERT --
```

Figure 12. Amending the "def main(argv):" Function

2. Use the XenServer console commands below to list all available Virtual Machines and to assign desired VF to a VM of choice.

```
xl list-vm

xl pci-attach 1 0000:03:10.6
```



```
xenserver6 server console
[root@xenserver6 ~]#
[root@xenserver6 ~]#
[root@xenserver6 ~]# xl pci-attach
Usage: xl [-u] pci-attach <Domain> <BDF> [Virtual Slot]

Insert a new pass-through pci device.

[root@xenserver6 ~]#
[root@xenserver6 ~]#
[root@xenserver6 ~]# xl pci-attach 1 0000:03:10.6
libxl: error: libxl_pci.c:749:libxl_device_pci_reset The kernel doesn't support
reset from sysfs for PCI device 0000:03:10.6
[root@xenserver6 ~]#
```

Figure 13. VF (Bus #3, Device #10, and Function #6) Assigned to VM Domain ID #1

Figure 13 shows VF represented by Bus#3, Device#10 and Function#6 assigned to VM domain ID#1. Error message "libxl: error:" can be safely ignored. This error doesn't affect VM or VF's functionality.

3. Use the XenServer console command below to verify VF assignment.

```
xl pci-list 1
```



```
xenserver6 server console
[root@xenserver6 ~]#
[root@xenserver6 ~]#
[root@xenserver6 ~]#
[root@xenserver6 ~]# xl list
Name                               ID   Mem  VCPUs   State   Time(s)
Domain-0                           0   477    4   r----- 142.4
centos6                             1  1023    1   -b----- 37.0
[root@xenserver6 ~]#
[root@xenserver6 ~]#
[root@xenserver6 ~]# xl pci-list 1
Vdev Device
04.0 0000:03:10.6
[root@xenserver6 ~]#
```

Figure 14. VF Assignment to VM "centos6" with Domain ID #1

3.0 Summary

Intel's best-of-breed 10 GbE solutions are now available with I/O Virtualization capabilities. Customers get world-class Ethernet support along with I/O virtualization support in mainstream Linux distributions--all in a single adapter.

4.0 Customer Support

Intel® Customer Support Services offers a wide selection of programs, including phone support and warranty service. For more information, contact us at:

support.intel.com/support/go/network_adapter/home.htm

(Service and availability may vary by country.)

5.0 For Product Information

To speak to a customer service representative regarding Intel products, please call 1-800-538-3373 (U.S. and Canada) or visit

support.intel.com/support/go/network/contact.htm

for the telephone number in your area.

§ §