Cisco Expo
2012

Deploying UCS
Hints & tips from the field

T-DC3 / L2

Damien GOUJU
Specialized System Engineer DC France
Solutions and Technologies Practice – STeP

dgouju@cisco.com

What is this session about?

- It's an advanced session
  - Basic UCS knowledge is a prerequisite
- A collection of hints and tips gathered from UCS deployments
- My challenge is 45 minutes, 7 subjects ☺
  - Address pools management
  - Boot from SAN
  - Service profiles: networking
  - UCS 2.0: Upgrade
  - UCS 2.0: Disjoint L2 consideration
  - UCS 2.0: iSCSI boot
  - UCS 2.0: VMFEX

- I'll be completely available for questions after the session

A few words of introduction

What do we see in France?

- Majority of deployments (80%+) run a mix of Hypervisors and bare-metal
  - “Standalone”, vBlock, Flex Pod
- Hypervisor is for the large part (90%) VMware’s vSphere ESXi
  - ESXi 4.1 Update 2 primarily, a little bit of ESXi 5.0
  - Microsoft’s Hyper-V comes second (and growing!)
- Bare-metal deployments consist of Windows 2008 R2 server and RHEL 5/6
- All types of LAN & SAN networking / Storage infrastructures

UCS Service Profiles: A Quick Recap

Service Profiles and Templates

- Physical blades inherit a desired configuration through a Service Profile (SP)
  - One SP can be either unassociated or associated to one blade
- Service Profiles specify
  - Desired boot order
  - Number of vNICs with VLANs and vHBAs
  - Where to get addresses for those vNICs/vHBAs
  - Various policies (BIOS configuration, IPMI, etc.)
- SP Templates are master templates from which SPs can be spawned
  - Either manually (from 1 to N with auto-naming)
  - Associated to a pool of servers

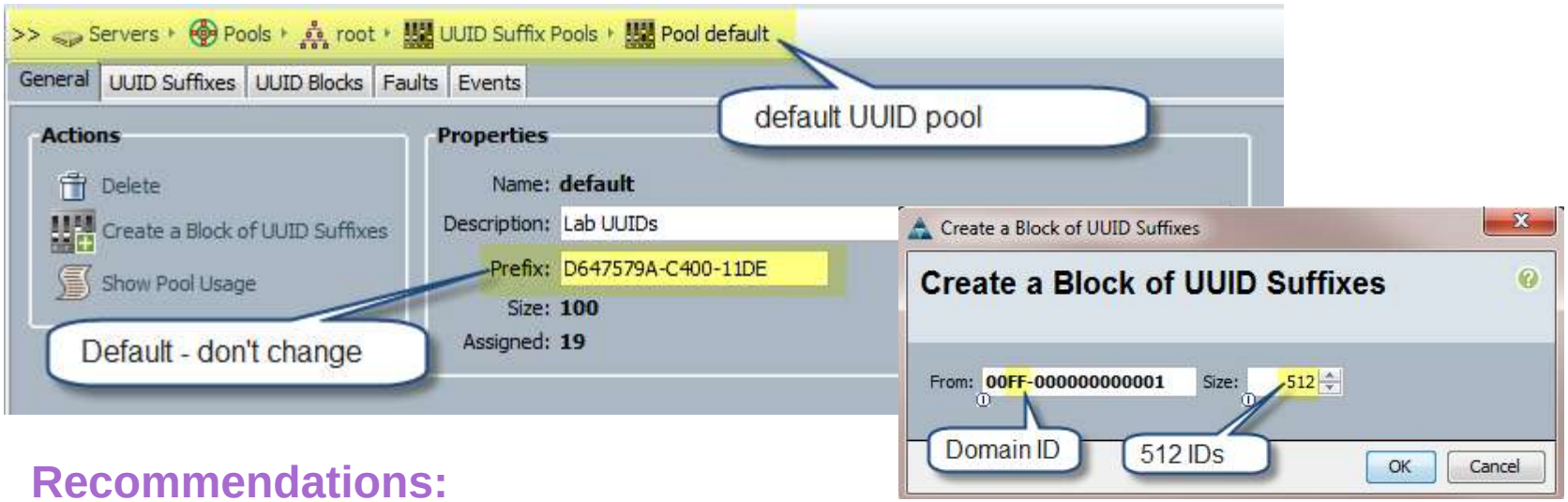
Address Pools: Best practices

UUID Pools

- UUID: global ID that is unique to a given server

  Composed of Prefix and Suffix

  Best practice: don't modify the prefix



- **Recommendations:**
  - Use root "default" pool as the global default pool for all Service Profiles
  - Populate the default pool with a block of 512 IDs
  - Don't alter original Prefix, this is unique to this UCS
  - Optional: choose a "Domain ID" for this UCS - used later in other ID pools

Assigning addresses to adapters

MAC Address Pools

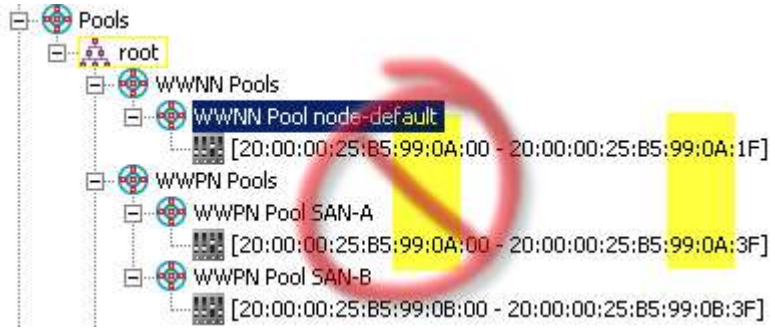
- Prefer pools to burnt-in values whenever possible
  - L2/MAC troubleshooting is a full time job… Please help Network team ☺
  - Pools let you control the exact allocation following your naming convention
  - Makes it easy to identify a given blade or OS type on switches
- On the VIC (M81KR, 1240…) adapter this is a must
  - VIC can instantiate N interfaces → there are no burnt-in addresses → "Insufficient resources"
- Best practice
  - MAC pools: create pools that are multiple of 64
  - Encode Domain/Site ID and OS Type

MAC Pool					256 MACs
OUI			Extension ID		
00	25	B5	Domain ID	OS Type	##

Assigning addresses to adapters

World Wide Name Pools

- Using pools lets you communicate WWNs to SAN team ahead of deployment
  - Pre-provision LUNs for boot-from-SAN
  - Proactively perform zoning and LUN masking configuration
- One blade uses one Node WWN and as many Port WWN as there are vHBAs
- Node pool best practice: create one large pool that's a multiple of 16
  - Create the pool at the Root organization (you can use the default pool)
  - Zoning and masking does not use Node WWN
- Ensure node pools and port pools do not overlap



Instead, use "FF"/"00" in the 2nd high-order byte to identify node names

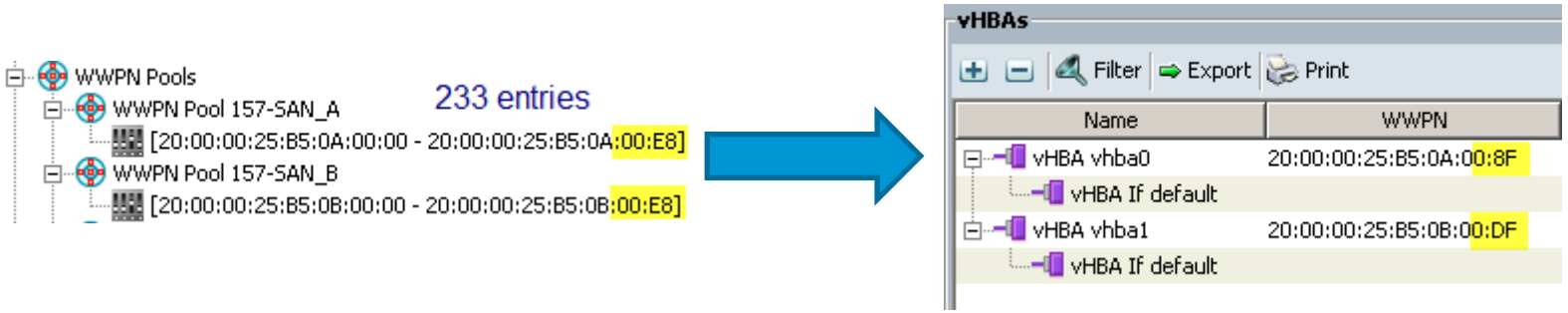
Port WWN pools

Best Practices

- Always create pools that are multiple of 16 and contain less than 128 entries

  This ensures vHBA0 (SAN A) and vHBA1 (SAN B) have the same low-order byte

- Counter-example using 233-entries pools



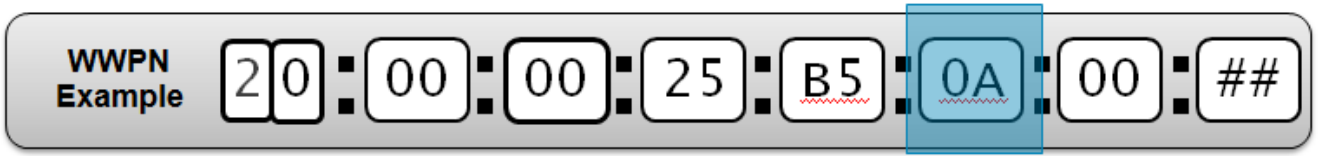
- Much better for both vHBAs to have the same low-order byte and a unique SAN Fabric identifier

  Presence of “0A” or “0B” in the port WWN indicates SAN Fabric

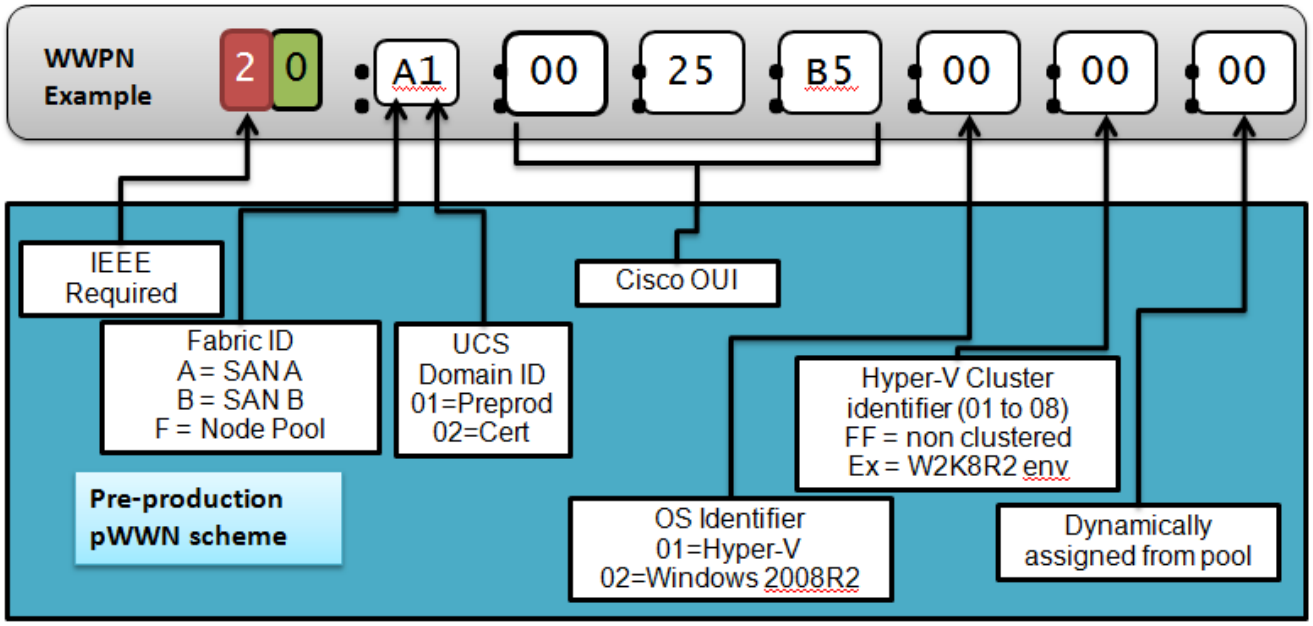
Port WWN pools

Suggested Design Patterns

- At the very minimum, create pools that identify the SAN Fabric

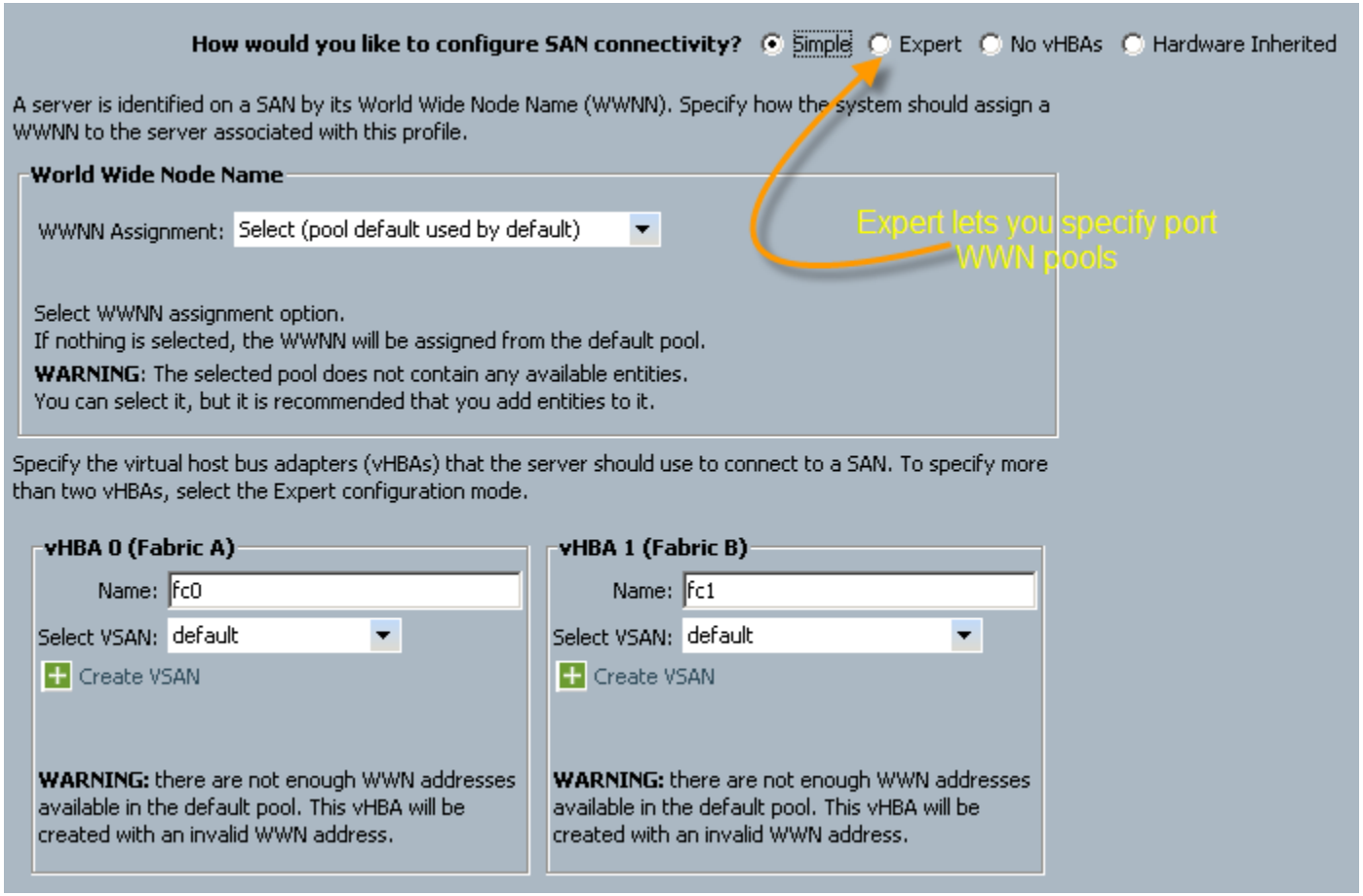


- For deployments with several OS types and multiple UCS, you could do this:



Port pools

Use Expert setting when creating vNICs / vHBAs

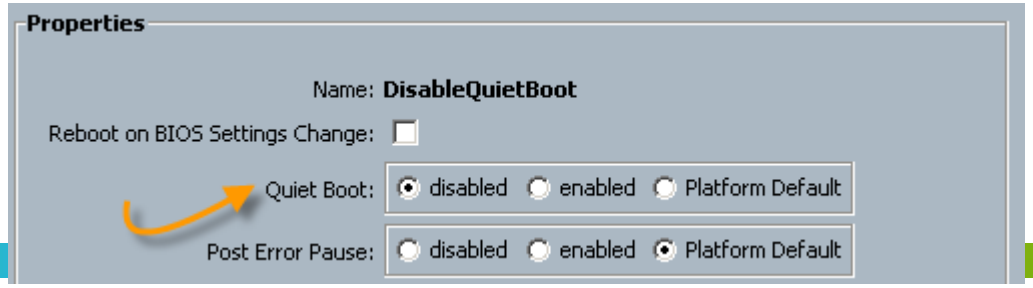


Boot from SAN

Tune your BIOS policy

Let the server speak up

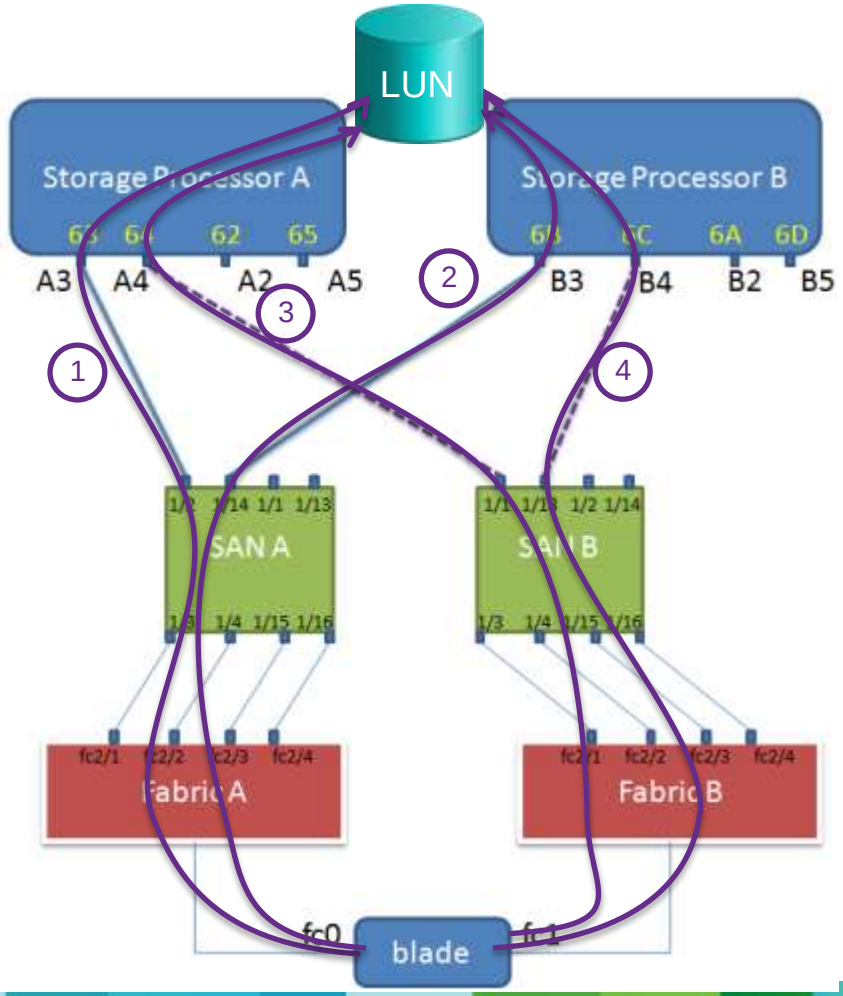
- Boot from SAN involves several key components working hand in hand
  - Correct UCSM boot-from-SAN policy with the right target port WWNs
  - Correct SAN zoning and LUN masking are imperative
  - SAN array must present a LUN (storage groups, initiator groups, etc.)
- During your first trial a component won't work the way it's supposed to
- UCSM lets you create BIOS policies that you can attach to the Service Profile
- Best Practice: for Boot-from-SAN you always want Quiet Boot disabled



Physical topology

What yours should look like

- SAN network designers always want SAN A / SAN B isolation



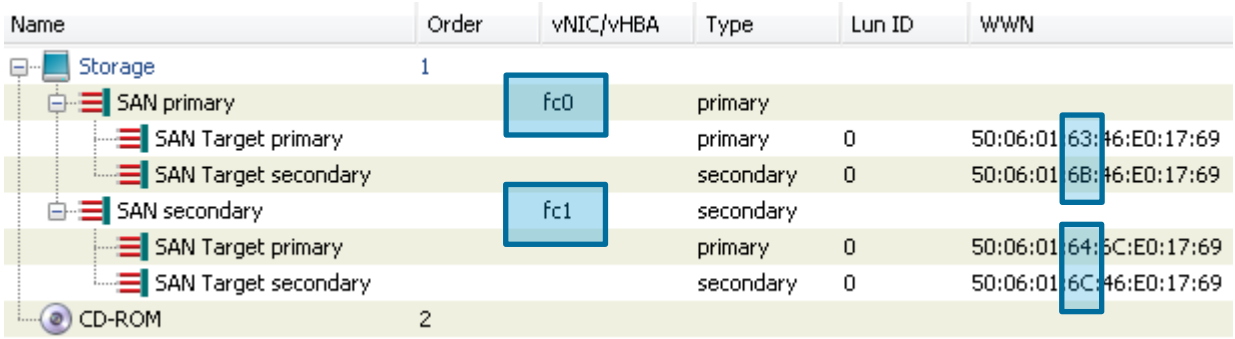
Write down the array's port WWNs before configuring your Service Profile!

1. SAN Primary, target primary
2. SAN Primary, target secondary
3. SAN Secondary, target primary
4. SAN Secondary, target secondary

Build your boot policy

One path works, but if resiliency matters …

- UCS can boot from 4 different paths
  - You can boot with just a single target boot policy, but not ideal for resiliency
- Typically, you'll want a boot policy that goes like this:

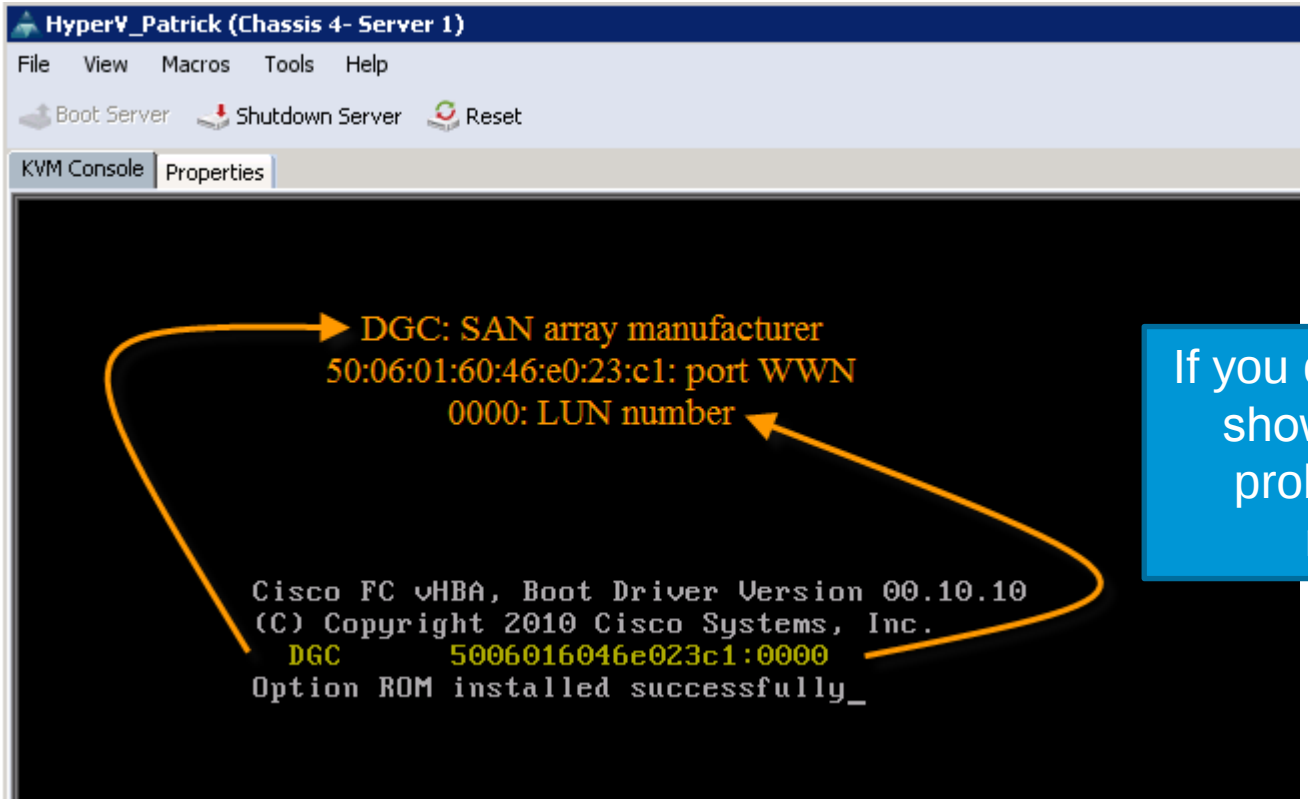


- That policy says:
  - First try vHBA fc0 pWWN "63" via fc0 → Storage Processor A, port A3
  - Then try vHBA fc0 pWWN "6B" via fc0 → Storage Processor B, port B3
  - If those fail, then try fc1 (first pWWN "64" on SP A; then pWWN "6C" on SP B)
- Windows: Don't forget to append CD-ROM or PXE after the SAN targets

Let's boot the server

Keep an eye out

- Associate the boot policy you just defined then boot the server
- With a M81KR adapter, this is what you'll see for each vHBA



If you do not see the array show up here, there's probably a zoning or masking error

Note: the 4th byte here is "60" because we are booting from a different array than slide 18

Troubleshoot

Attach VIC firmware

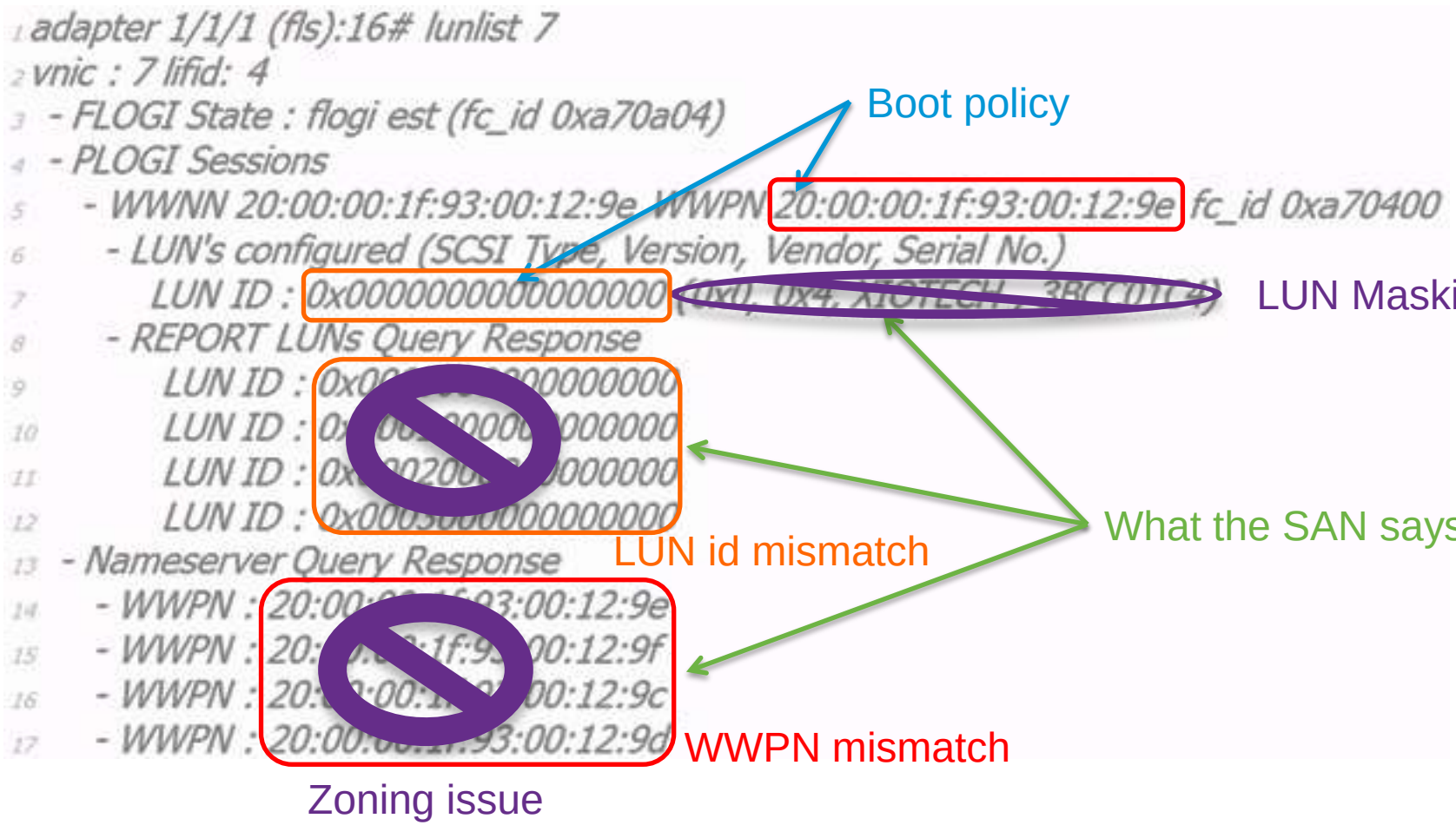
Chassis / Server / Adaptor

```
UCS-6100-A# connect adapter 1/1/1
adapter 1/1/1 # connect
adapter 1/1/1 (top):1# attach-fls
adapter 1/1/1 (fls):1# vnic
---- ---- ---- ------- -------
vnic ecpu type state   lif
---- ---- ---- ------- -------
7    1    fc   active  3
8    2    fc   active  4
```

Useful (see next slide)

The cool command: VIC lunlist

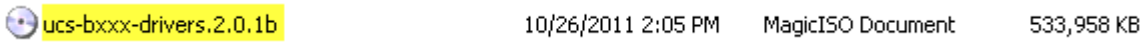
Before driver loading



Microsoft Windows Boot from SAN

Checklist

- Best practice: ask the SAN team to present a single target pWWN for now
- Don’t proceed unless you’ve seen the target’s port WWN on the previous screen!
  - Your SAN team can pre-register hosts; you know the server’s pWWN!
- Make sure you have the latest Cisco drivers iso image handy!

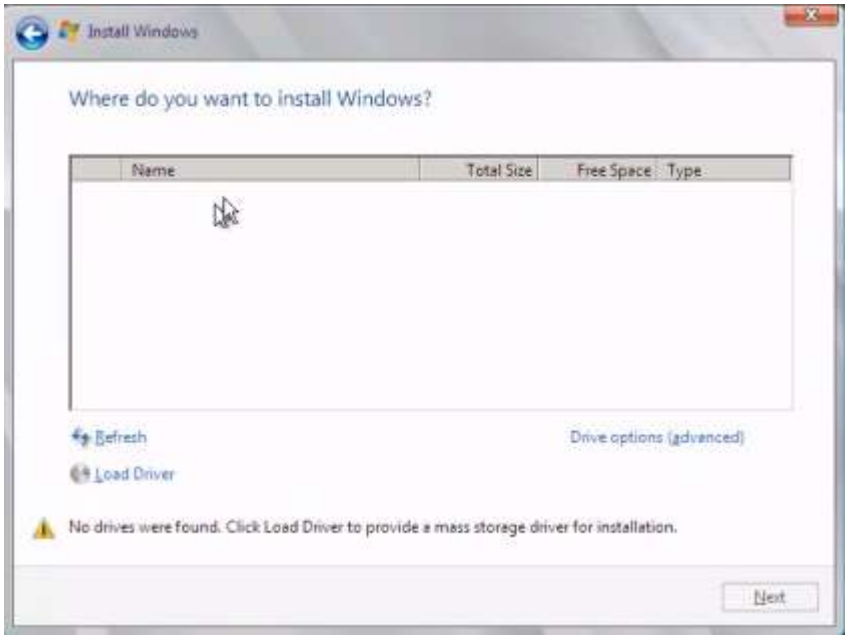


- Your boot policy must have the SAN targets first before PXE or CD-ROM
  - If you put PXE or CD-ROM first, the installer will refuse to install on the LUN!
- Mount the virtual media (or ensure PXE is working) and boot the server
- Do not press F6 to force select boot options, UCSM has taken care of that
- After a few click-next (accept EULA; select Custom Installation), you’ll see a screen asking you where to install Windows

Windows: nowhere to install!

Drivers to the rescue

- You should now see this:



- Windows does not ship with built-in drivers for Cisco CNAs
  - Unmap the installation ISO image and map the Cisco drivers image
  - Then click “Load Driver”

Drivers loaded

LUN appears

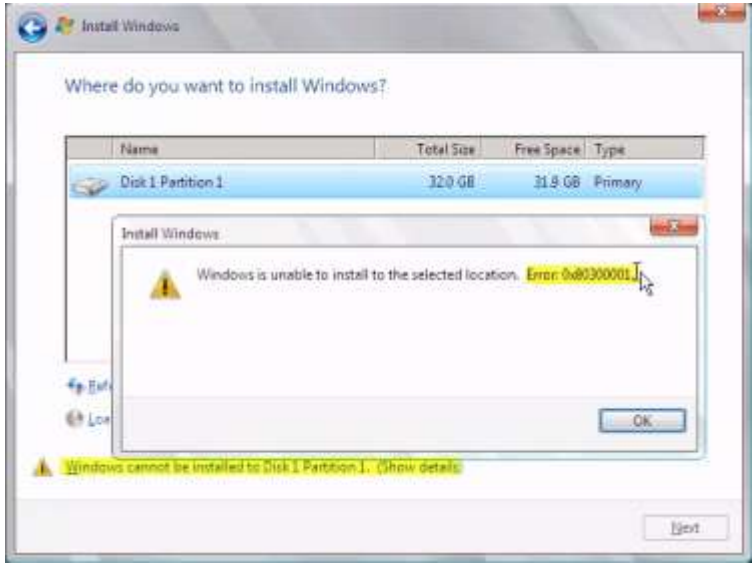
- Load the driver then wait a bit and the LUN will appear

- If the LUN does not appear, check FLOGI database on SAN switches
- If FLOGI is present, check zoning
  - If it is not present, double check the VSAN membership of the vHBA
  - Don’t forget to activate the new zoning configuration
- If zoning is correct, verify the host pWWN is/are registered on the array
  - Present a boot LUN (for Windows 2008 R2 typically 40GB) only to that host

I am getting an error!

What did I do wrong?

- You're very likely going to run into this error:



- At this point, unmap the Cisco drivers ISO image and remap the Windows installation media
  - You may need to click Refresh
- The installer should now proceed

After the installation, enable MPIO (register your array with the MPIO driver) then present all available SAN paths

VMware ESXi Boot from SAN

Easy as pie

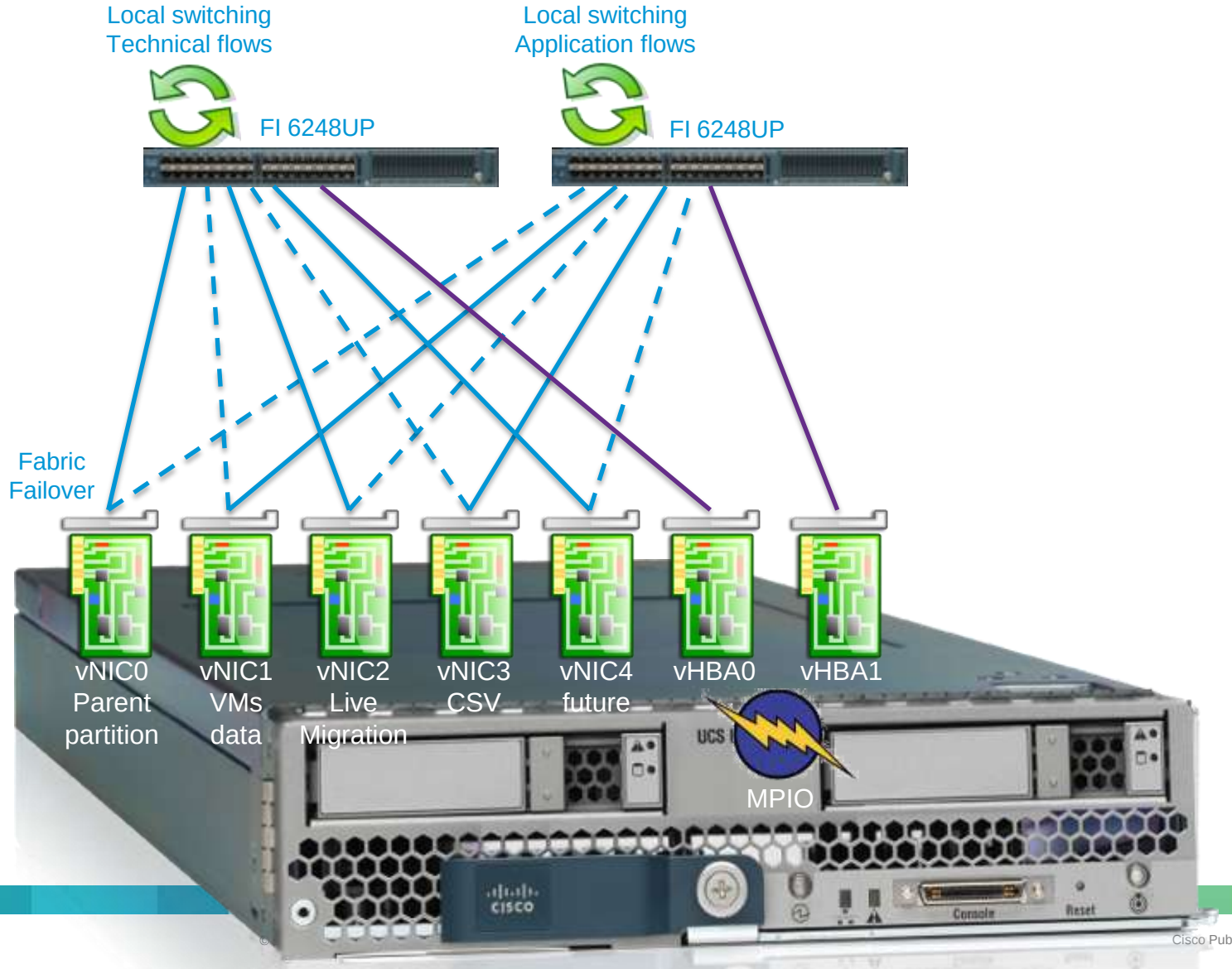
- All ESXi versions ship with Cisco CNA drivers built-in
  - But check Cisco & VMware matrix (inbox vs async)
  - Cisco OEM image for ESXi 5.0 soon
- The installer is more forgiving than with Windows
  - You could place the virtual CD-ROM or PXE before the SAN targets and it will work
- The installer is multipath-aware
  - It is okay to present all available paths at installation time
- Best practice: through LUN masking only present the boot LUN while installing
  - If you data is important, do not present VMFS datastores just yet!
- Before ESXi 5.0 you only get one chance
  - If LUN masking is incorrect, the installer won't refresh and you'll have to reboot
  - The ESXi 5.0 installer introduces a proper refresh function which will save you a lot of time

Recommended Service Profile Network Designs

What's a Service Profile design?

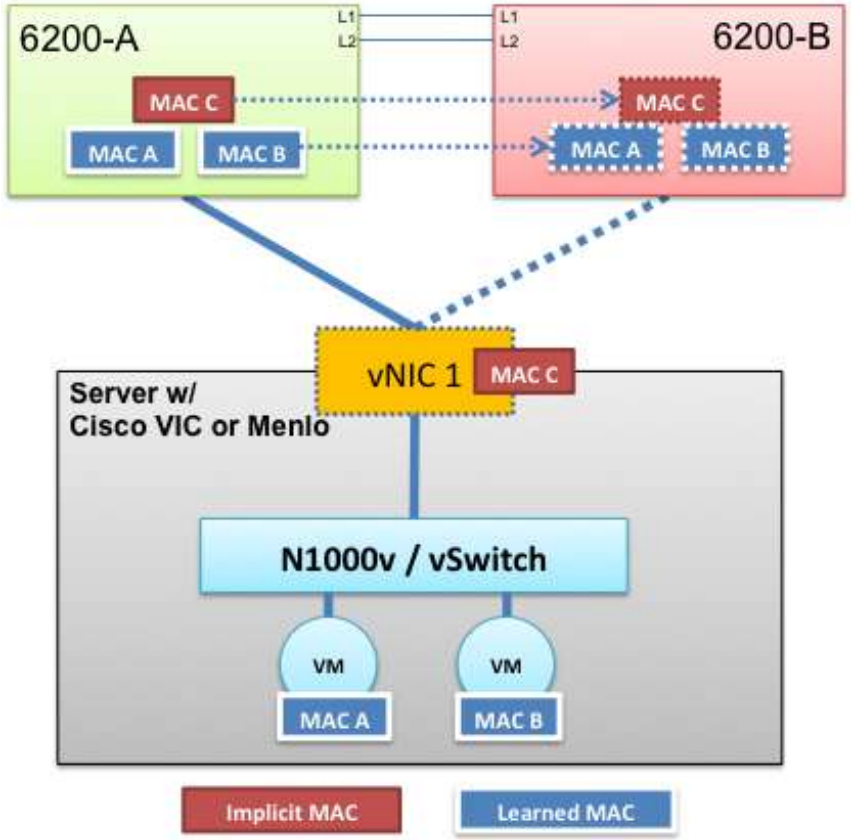
- Question: "how many vHBAs and vNICs should I assign to my Windows profile? How about ESXi?"
  - If you don't have a Cisco VIC adapter, the answer is simple: 2 plus 2!
  - We're going to focus on VIC use cases
- For bare-metal deployments: the answer is "it depends"
  - Consult the application owner(s)
  - Empirical observation: typically 4 to 6 vNICs; 0 to 2 vHBas
- For ESXi: most of the times 6-8 vNICs and 2 vHBAs
  - Depending on Failover mode
- For Hyper-V: most of the times 5 vNICs and 2 vHBAs
  - Details in the next slides

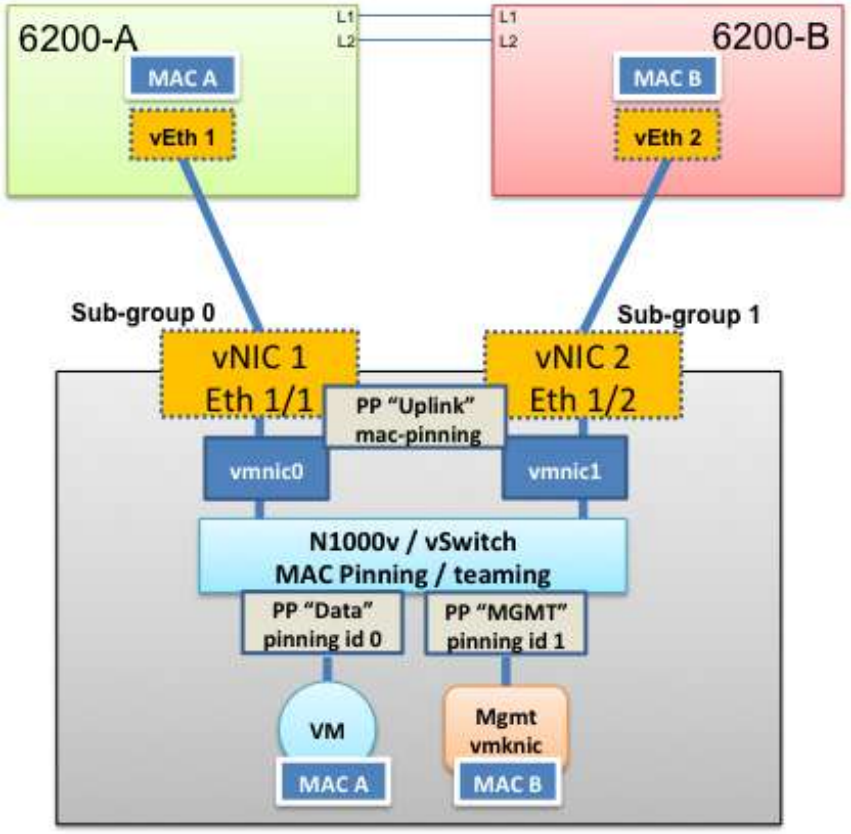
Sample HyperV design



The dilemna

Fabric failover or vSwitch mac-pinning?





+ Less vNIC
+ HyperV support (KB 968703)
+ Local switching
- No load sharing
- Not supported in Eth Switch mode (since 2.0)

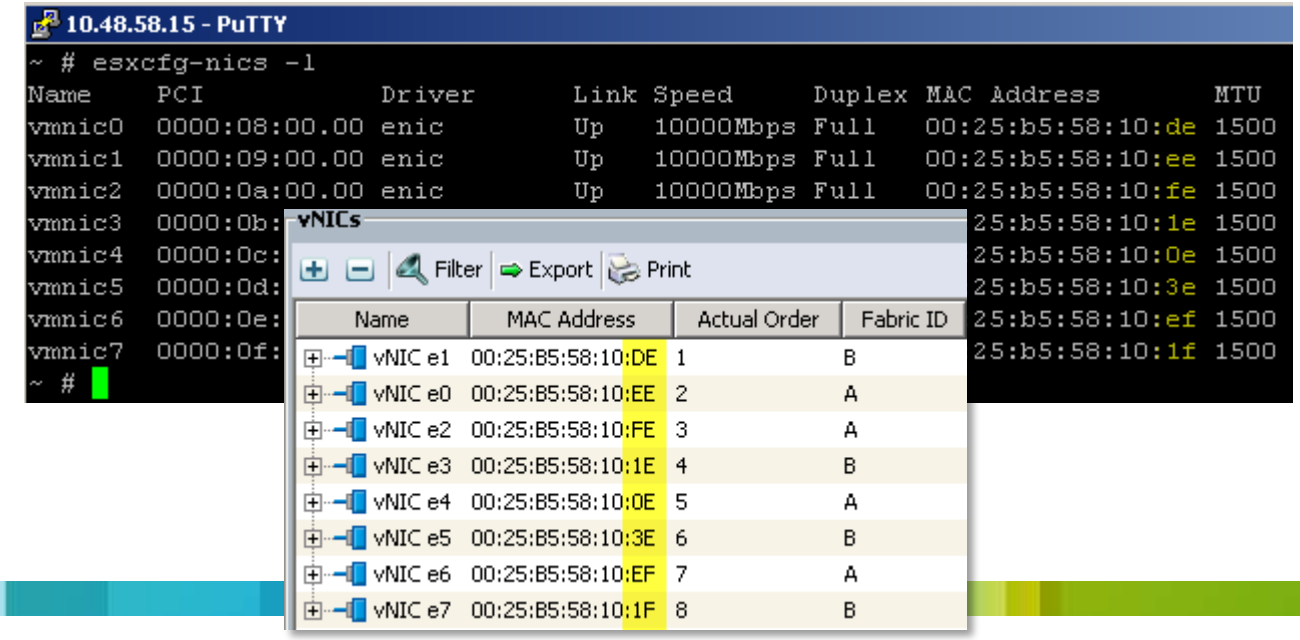
+ Load sharing across fabric
+ Supported in Eth Switch mode (teaming driver)
+ Traditional approach
- "Bad" HyperV support (KB 968703)
- East-West through north

Service Profiles: OS versus UCSM view

SP view vs UCSM view

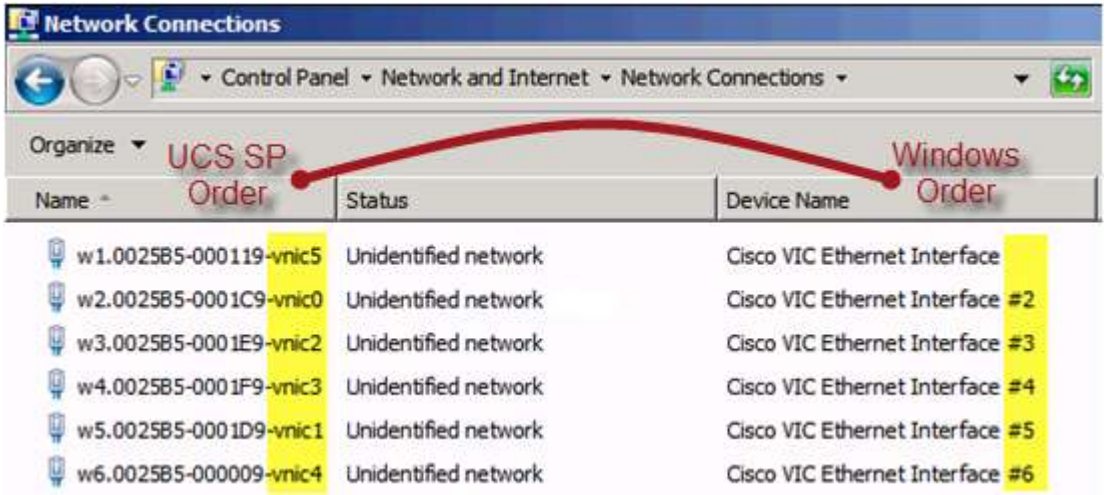
Do they disagree?

- Problem statement
  - an SP instantiates 6 vNICs and 2 vHBAs
  - Windows 2008 R2 is installed on the blade
  - NIC ordering at the OS level does not match what the SP shows
- Typically encountered with Microsoft Windows OS, usually not an issue with ESXi



What's specific to Microsoft Windows and UCS

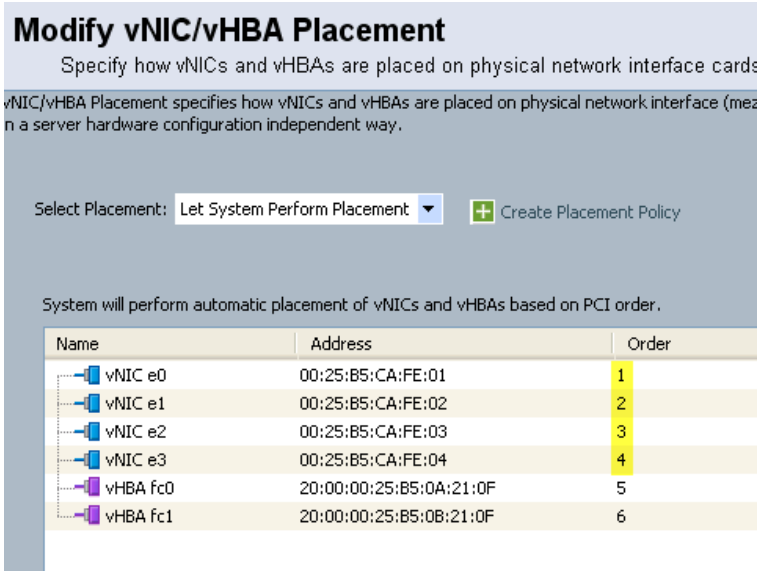
- Two variables come in the picture:
  - Windows does not ship with built-in drivers for M81KR
  - Once you load the driver the OS decides in which order to bring up the devices
  - That order does not have a direct relation with PCI-bus addressing
- You often end up with this:



How to reconcile the OS and SP view

When order matters

- First make sure the vNICs have a sequential order in the SP



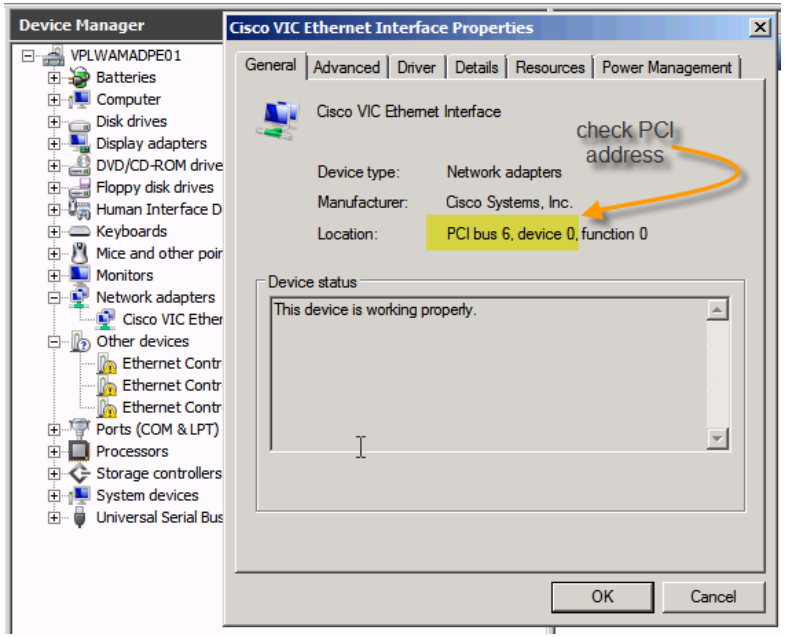
- vHBAs can appear before vNICs, it doesn't really matter
- Just ensure your vNIC order column follow the correct sequence

- Install Windows then launch the Device Manager

Device Manager View

Identify PCI bus order of each vNIC

- Before installing the device drivers, right-click on each vNIC one by one



- Write down the PCI address of each vNIC
- This is where the device order in the UCS SP comes into play – the first vNIC or vHBA is usually at PCI 06

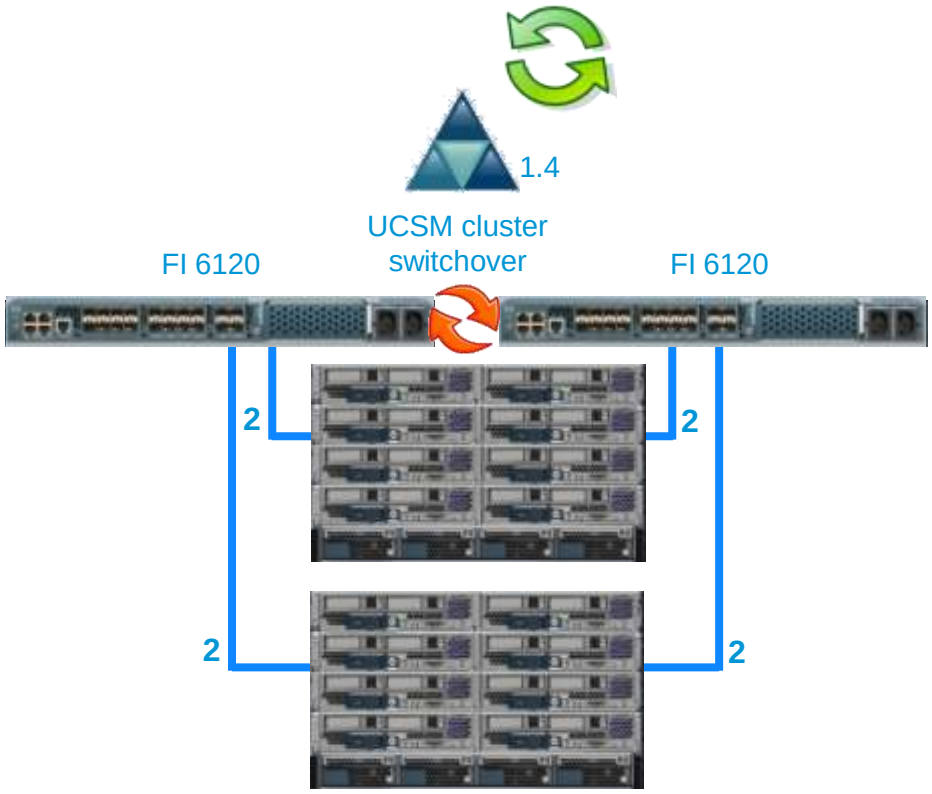
- Install the drivers sequentially following the PCI bus order you discovered

UCS 2.0
Upgrading

Upgrading to UCS Gen2

Software and Hardware

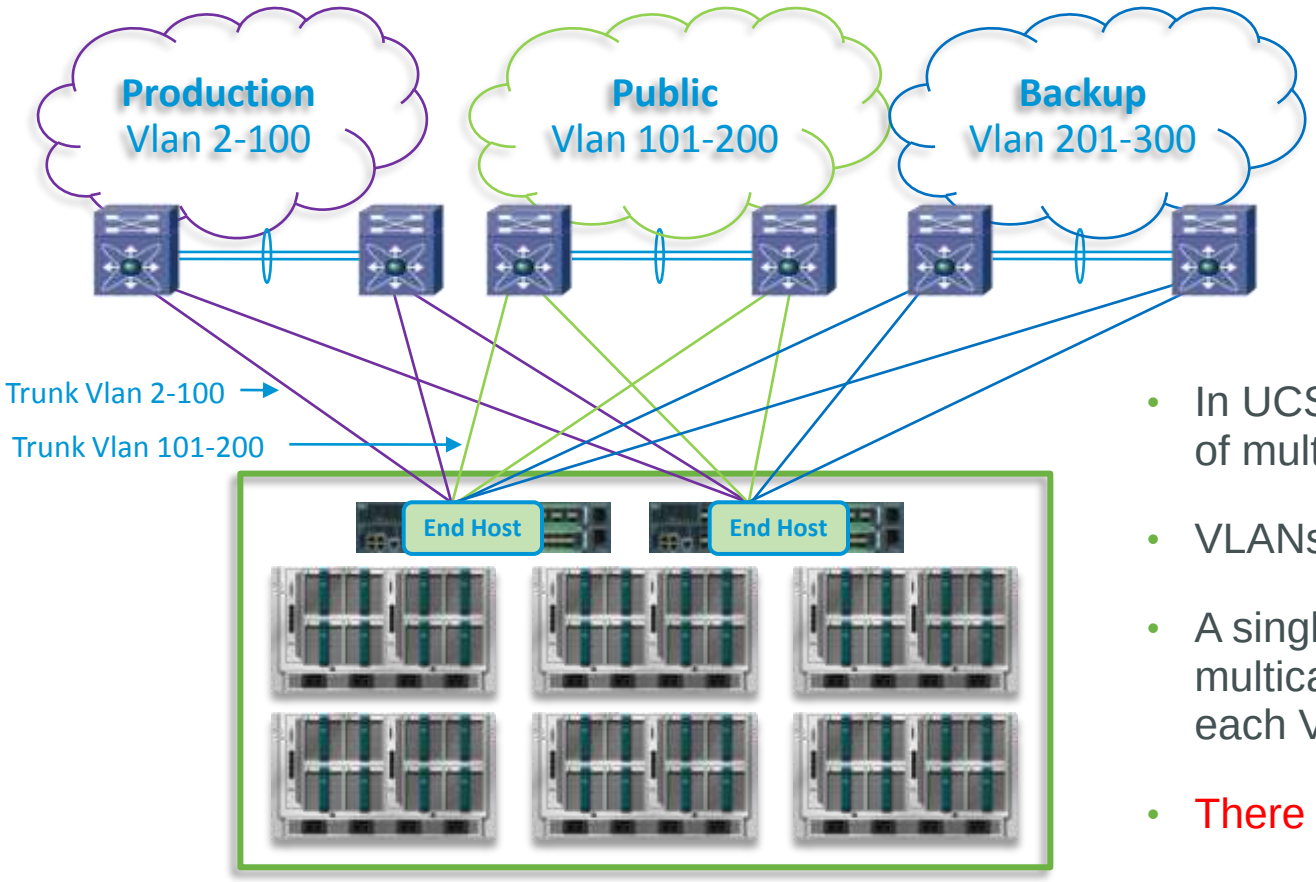
- Capitola: Major announcements
  - FI with more ports, all Unified Ports
  - More VNTags for better Adapter / VMFEX scalability
  - Software features
- Gen2 hardware ships with UCSM 2.0, but Gen1 can provide the software features
- 3 major points to consider and it will be fine
  1. Strictly apply upgrade procedure
  2. Ensure that FCoE VLANs don't overlap with Ethernet VLANs
     - FCoE VLAN set for every VSAN
  3. If you upgrade FI, ensure to have the same version on Gen1 and Gen2



Overview

UCS 2.0
Disjoint L2

Disjoint Layer 2 Support



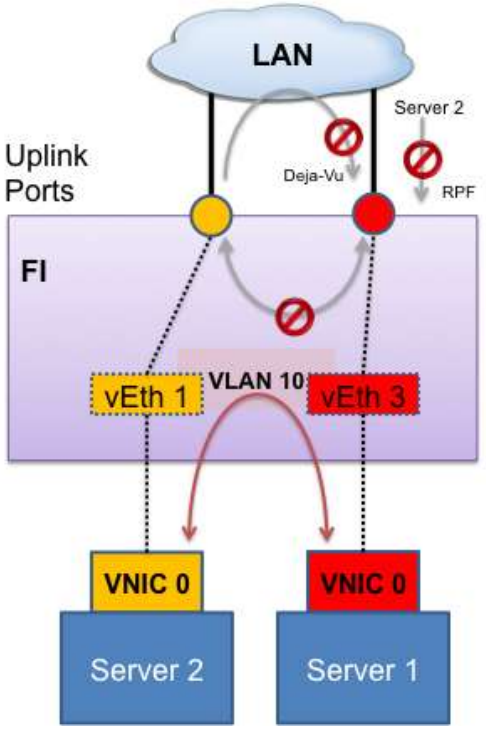
- In UCS 2.0, End host mode is aware of multiple L2 domains upstream
- VLANs can be filtered on uplinks
- A single uplink port designated as multicast and broadcast receiver on each VLAN.
- There musn't be VLANs overlapping!
- No VLAN translation

Is it secured?

- VLAN-based isolation
  - UCS since 1.4 passed EAL4+ Common Criteria certification
  - http://www.niap-ccevs.org/cc-scheme/st/vid10403/

- Hardened End Host Mode forwarding scheme
  - Disjoint L2 only add mapping & M/B receiver

- What is the risk? What is the cost?
  - Then you choose ☺
  - Most customer keep exposed / not exposed
  - physical separation

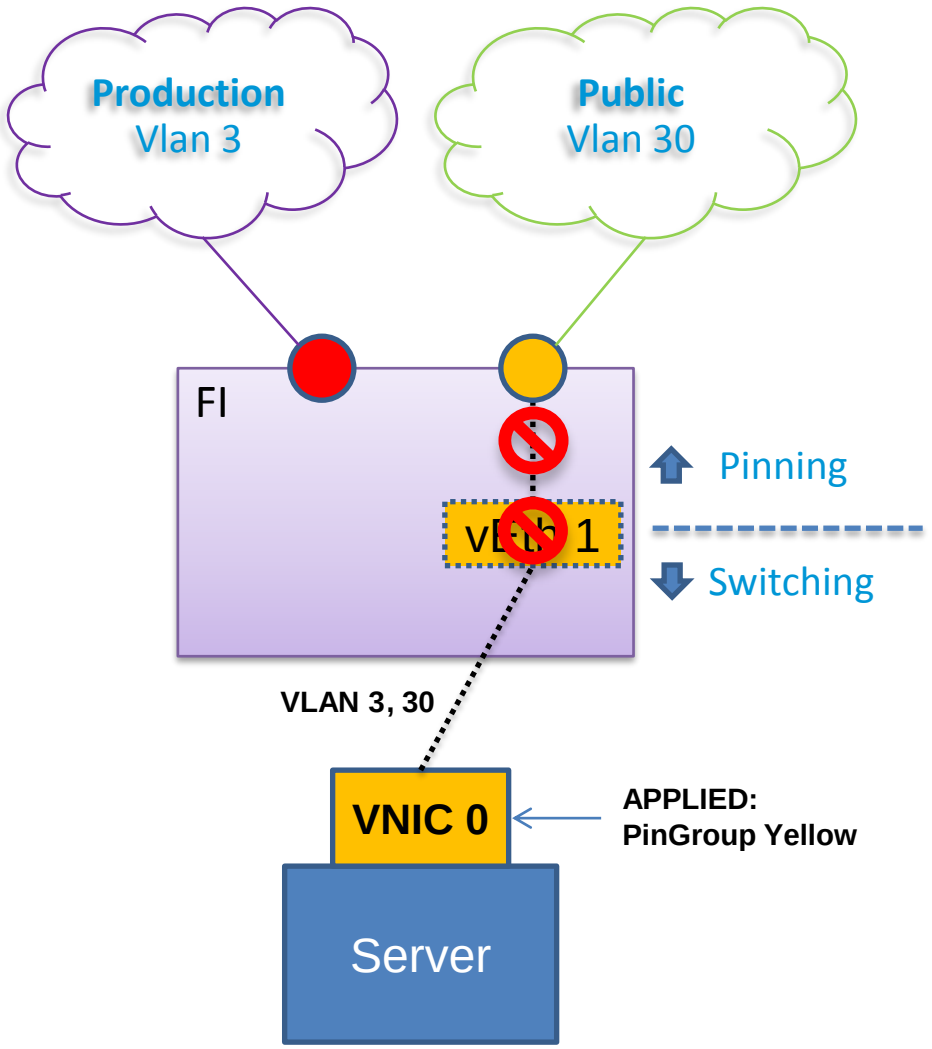


Considerations

- By default, everything is everywhere
  - Every uplink carry all VLANs
  - New uplinks will carry all VLANs
  - New VLANs will be carried on all uplinks
  - Removing VLAN from ALL its member uplinks will make it be carried by all uplinks

- vNIC is still pinned to uplinks
  - Pin groups take the lead (configuration warning)
  - vNIC can trunk VLANs from **1** domain

  So VIC is almost mandatory…



UCS 2.0
iSCSI Boot

iSCSI Boot Support in UCS Manager

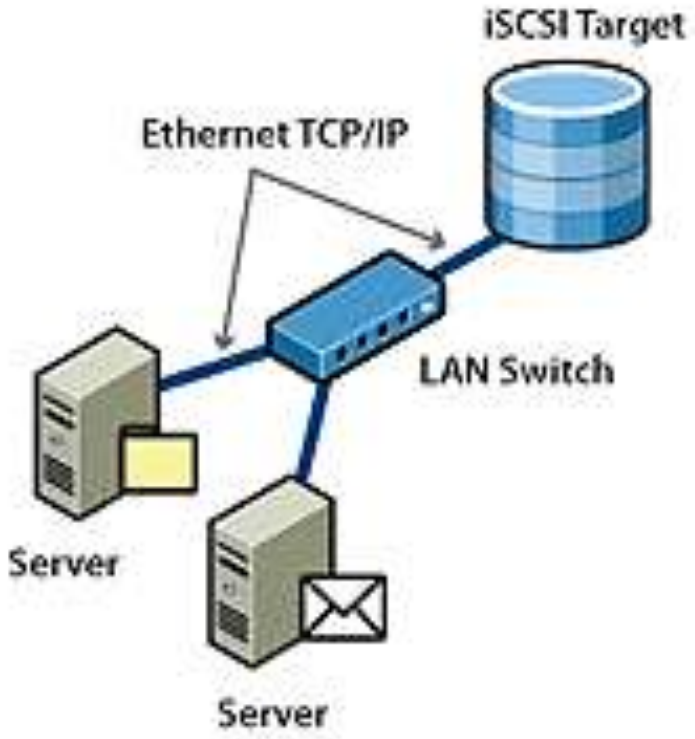
Feature details

- Adapters
  - M81KR / VIC 1280
    - iBFT, no iSCSI offloads
  - M51KR-B
    - full offload, iSCSI HBA
- Operating System support
  - VMware
  - Windows
  - Linux Red Hat

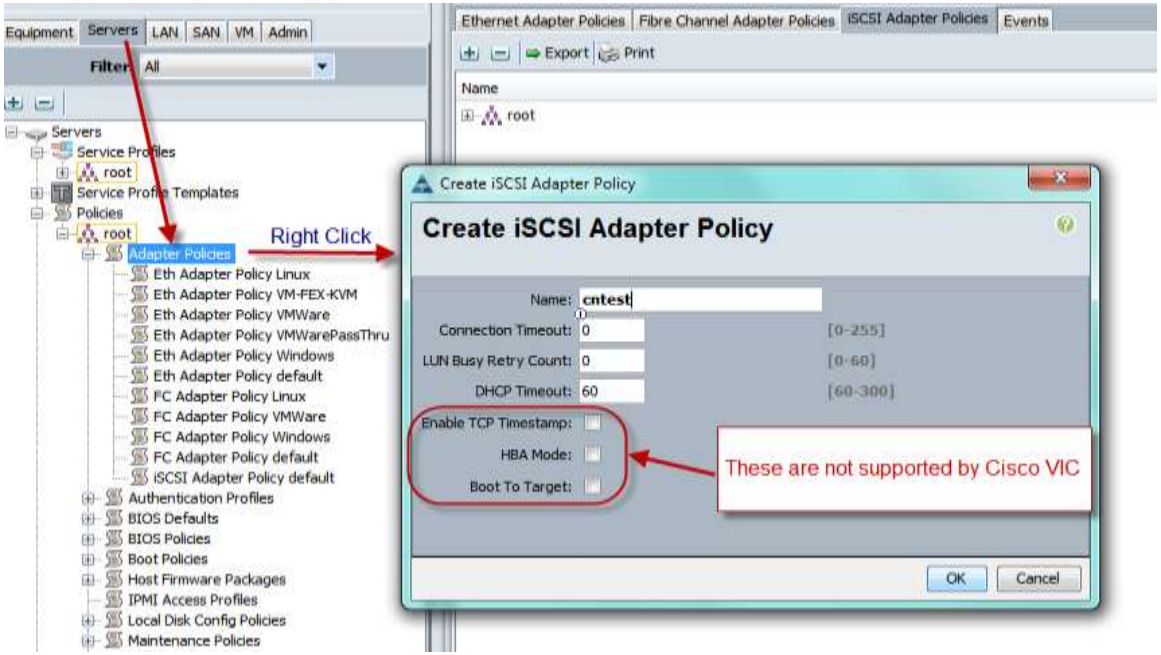
Refer release notes for specific versions supported and UCS Configuration guide:

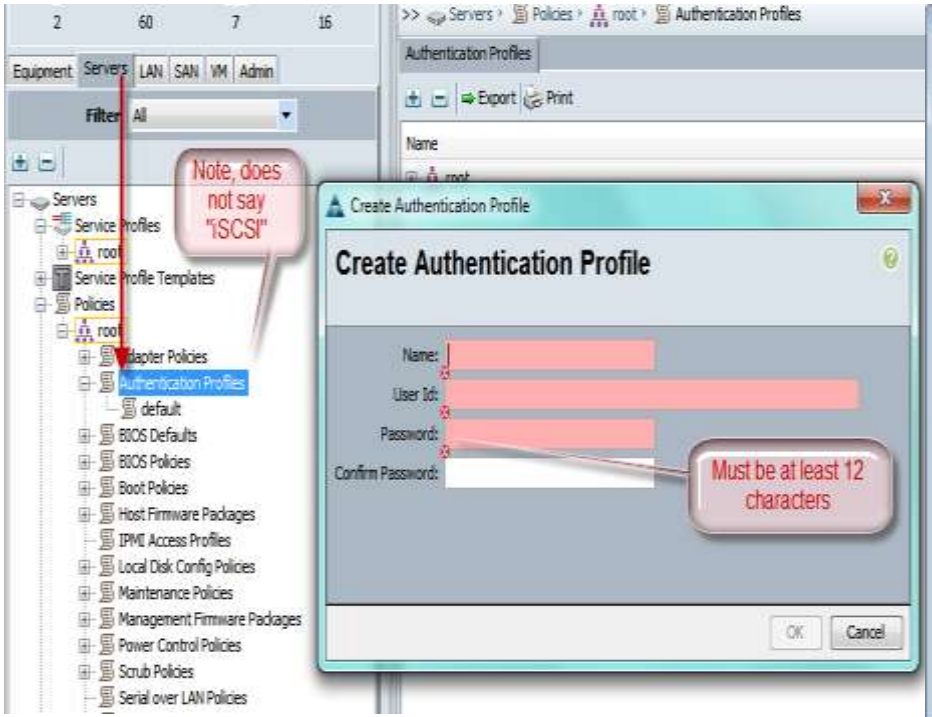
- Windows failure scenarios
- HBA mode during installation…

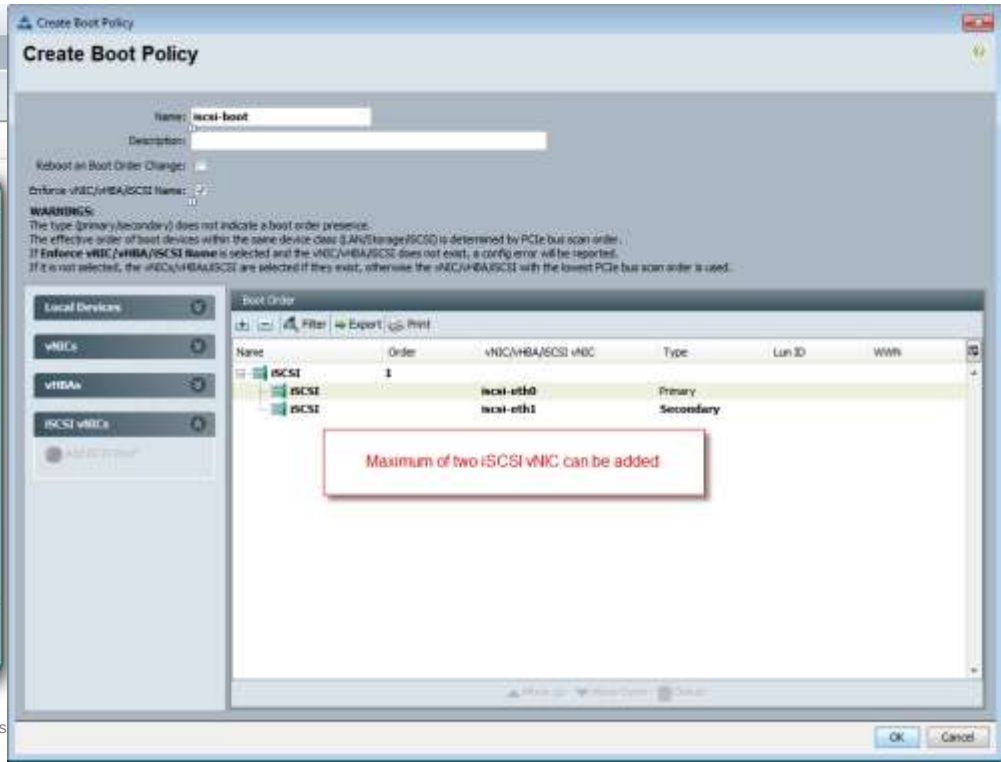
It is boot from SAN: so take care of your storage array mode of operation and failure scenarios



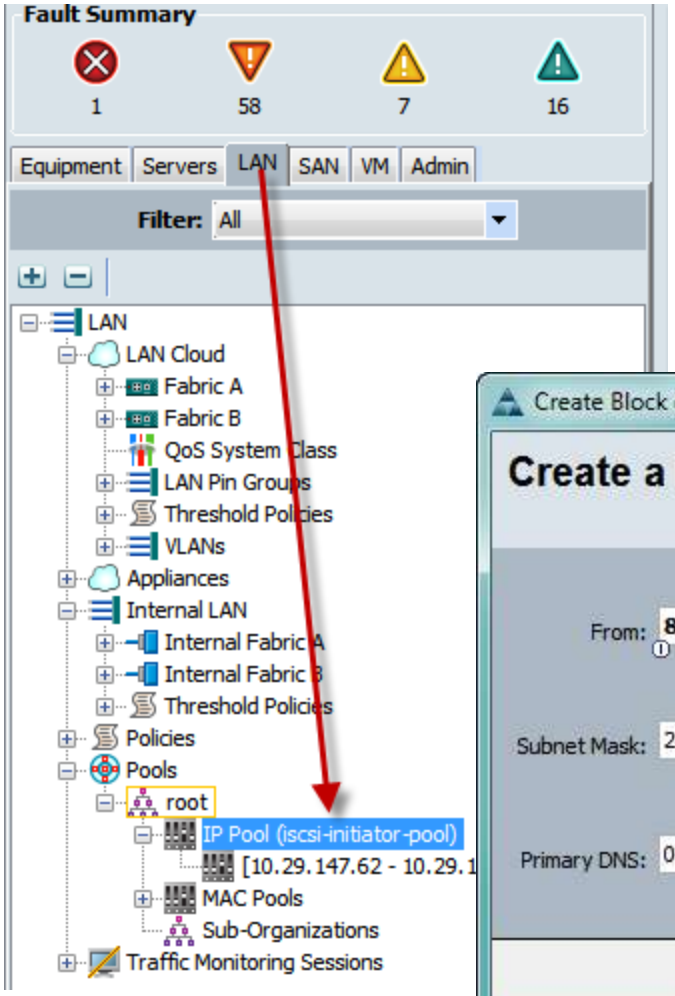
Policies

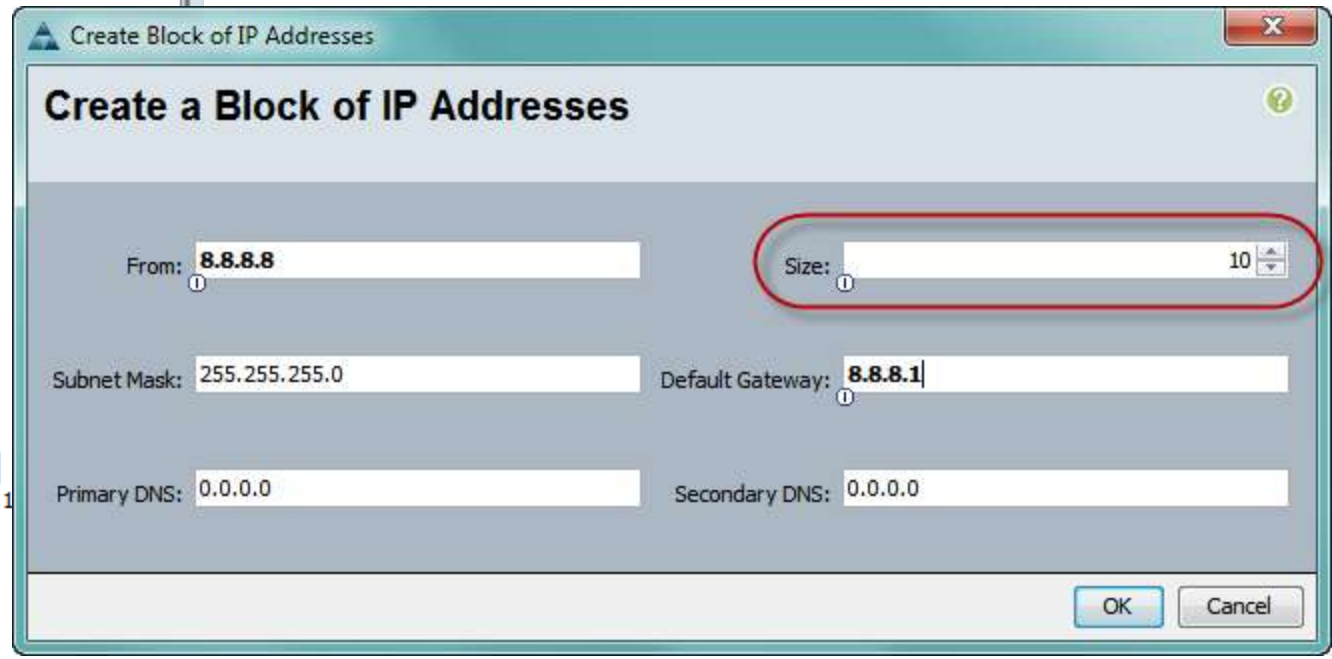




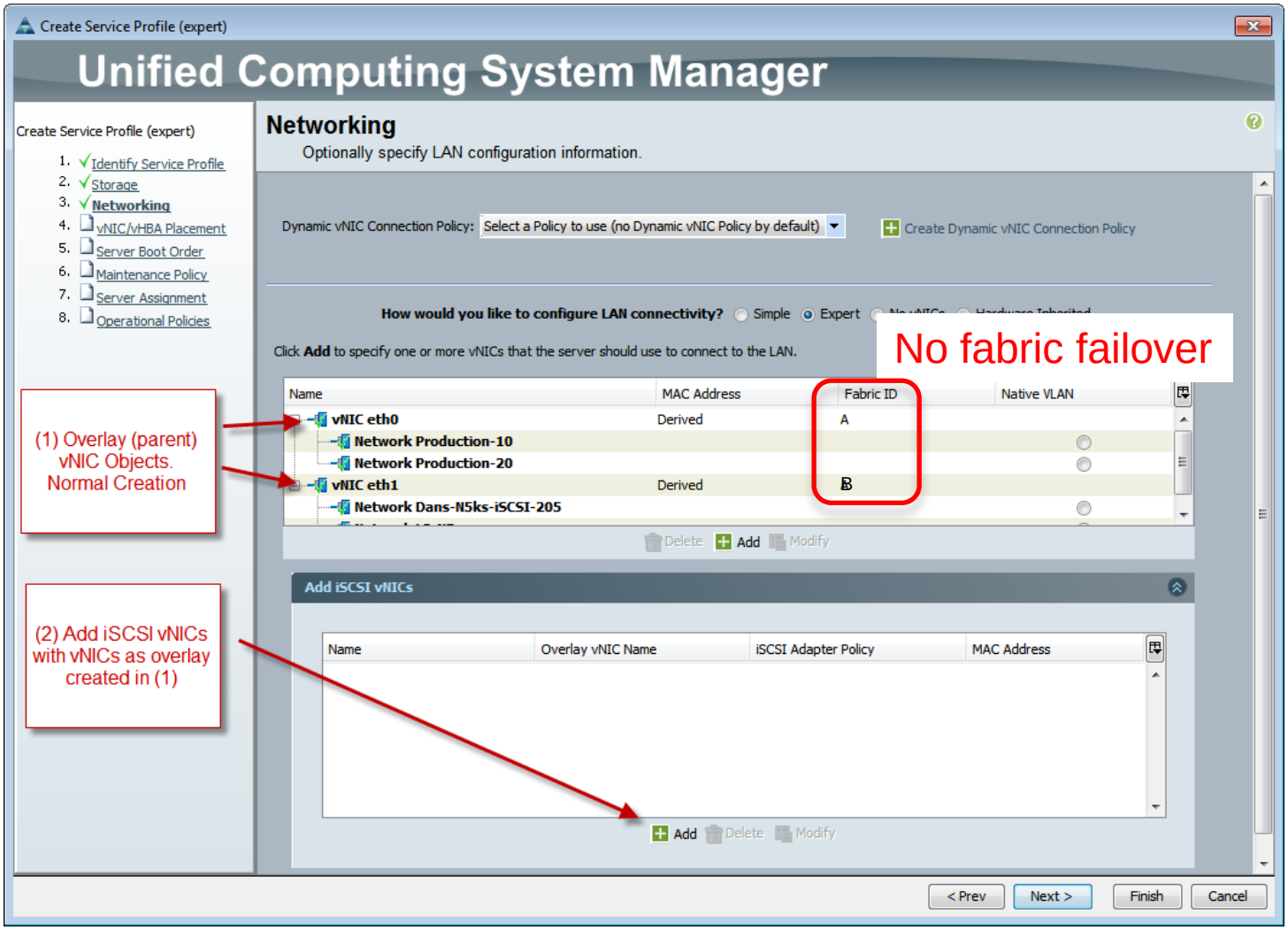


IP Address pool



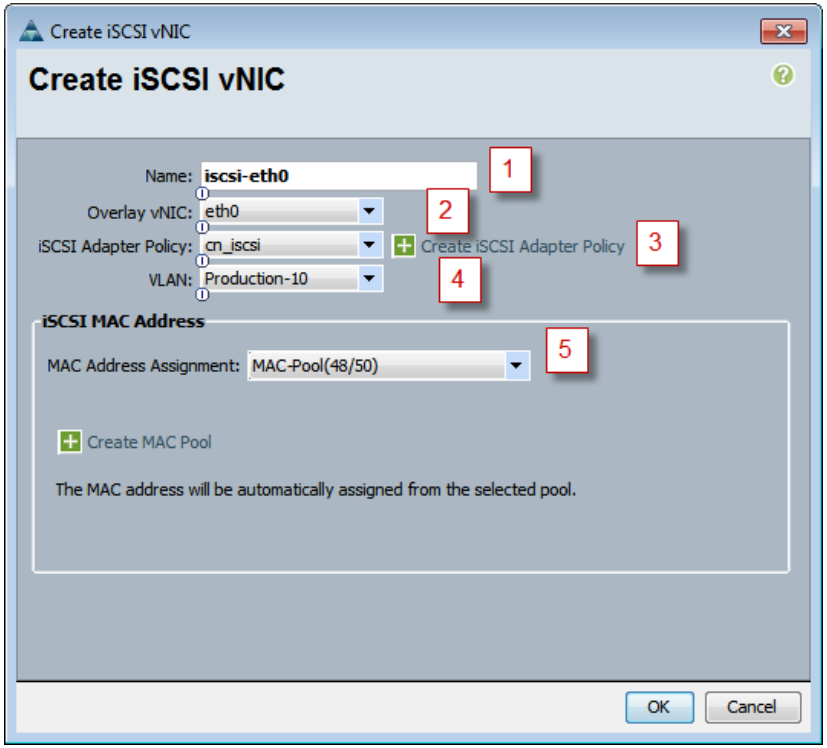


Overlay & iSCSI vNICs

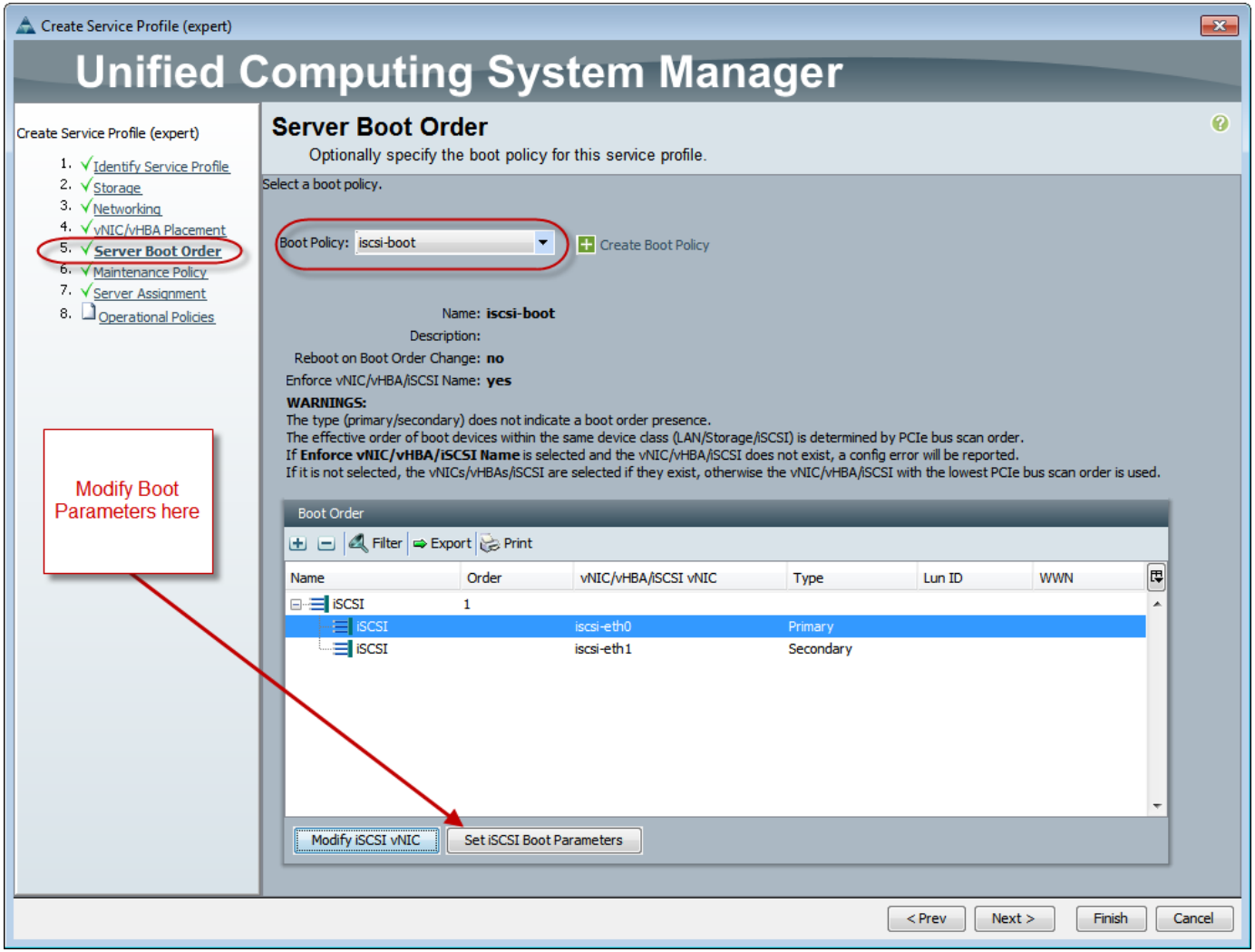


iSCSI vNIC configuration

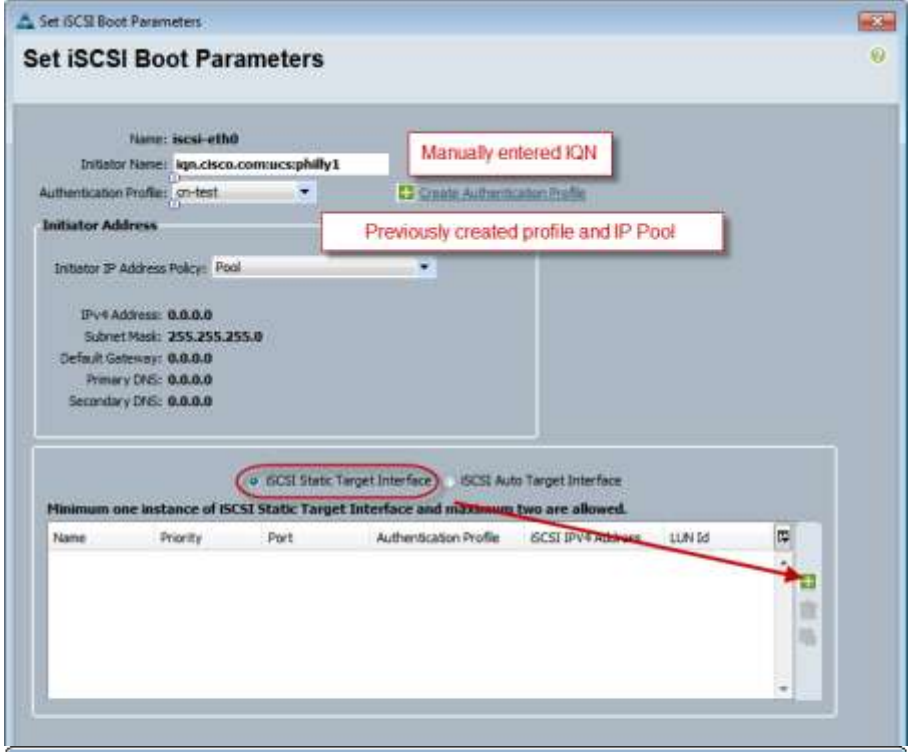
1. Object name, user preference
2. Overlay vNIC created in previous step. Child object, inherits some attributes
3. Adapter Policy previously created before SP process
4. VLAN, inherited from overlay vNIC, drop down list of available VLANs assigned to overlay vNIC. **MUST BE THE NATIVE VLAN with VIC**
5. **For VIC set to "None"** For Broadcom you can use the Standard MAC Pool
6. Max of 2 iSCSI vNICs per SP is allowed

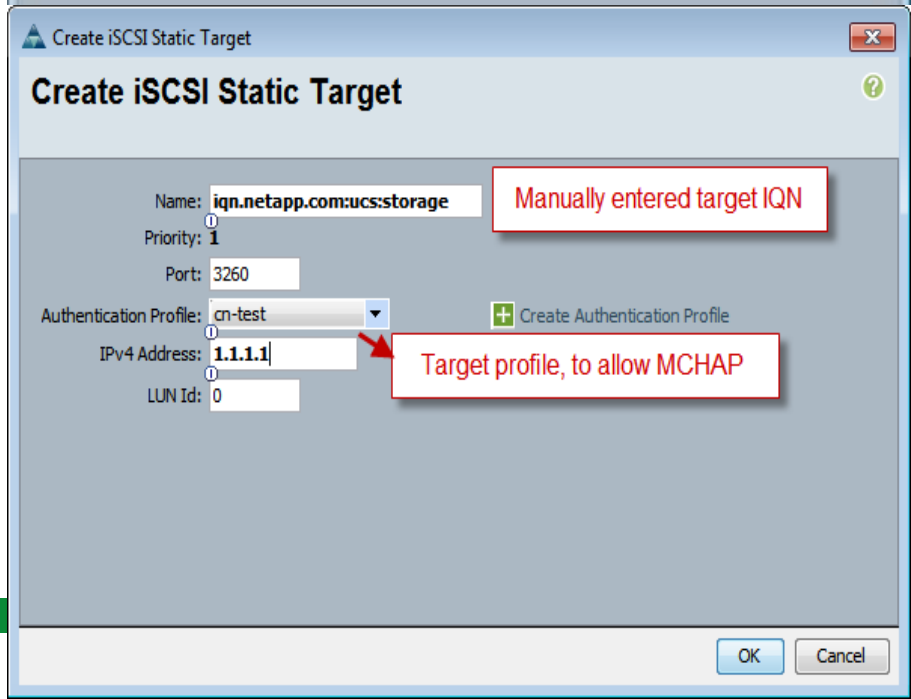


Define iSCSI boot parameters



Target modes





iSCSI boot troubleshooting on VIC

```
UCS-6100-A# connect adapter 1/1/1
adapter 1/1/1 # connect
adapter 1/1/1 (top):1# attach-mcp
adapter 1/1/1 (mcp):1# iscsi_get_config

vnic iSCSI Configuration:
--------------------------
vnic_id: 6
          link_state: Up

       Initiator Cfg:
     initiator_state: ISCSI_INITIATOR_READY
initiator_error_code: ISCSI_BOOT_NIC_NO_ERROR
                vlan: 0
         dhcp status: false
                 IQN: iqn.1991-11.com.cust:ucs-node1
             IP Addr: 10.20.13.179
         Subnet Mask: 255.255.255.0
             Gateway: 10.20.13.250

          Target Cfg:
          Target Idx: 0
               State: ISCSI_TARGET_READY
          Prev State: ISCSI_TARGET_DISABLED
        Target Error: ISCSI_TARGET_NO_ERROR
                 IQN: iqn.1992-08.com.netapp:sn.101201005
             IP Addr: 10.20.13.151
                Port: 3260
            Boot Lun: 0
          Ping Stats: Success (9.553ms)

        Session Info:
          session_id: 0
         host_number: 0
          bus_number: 0
           target_id: 0
```

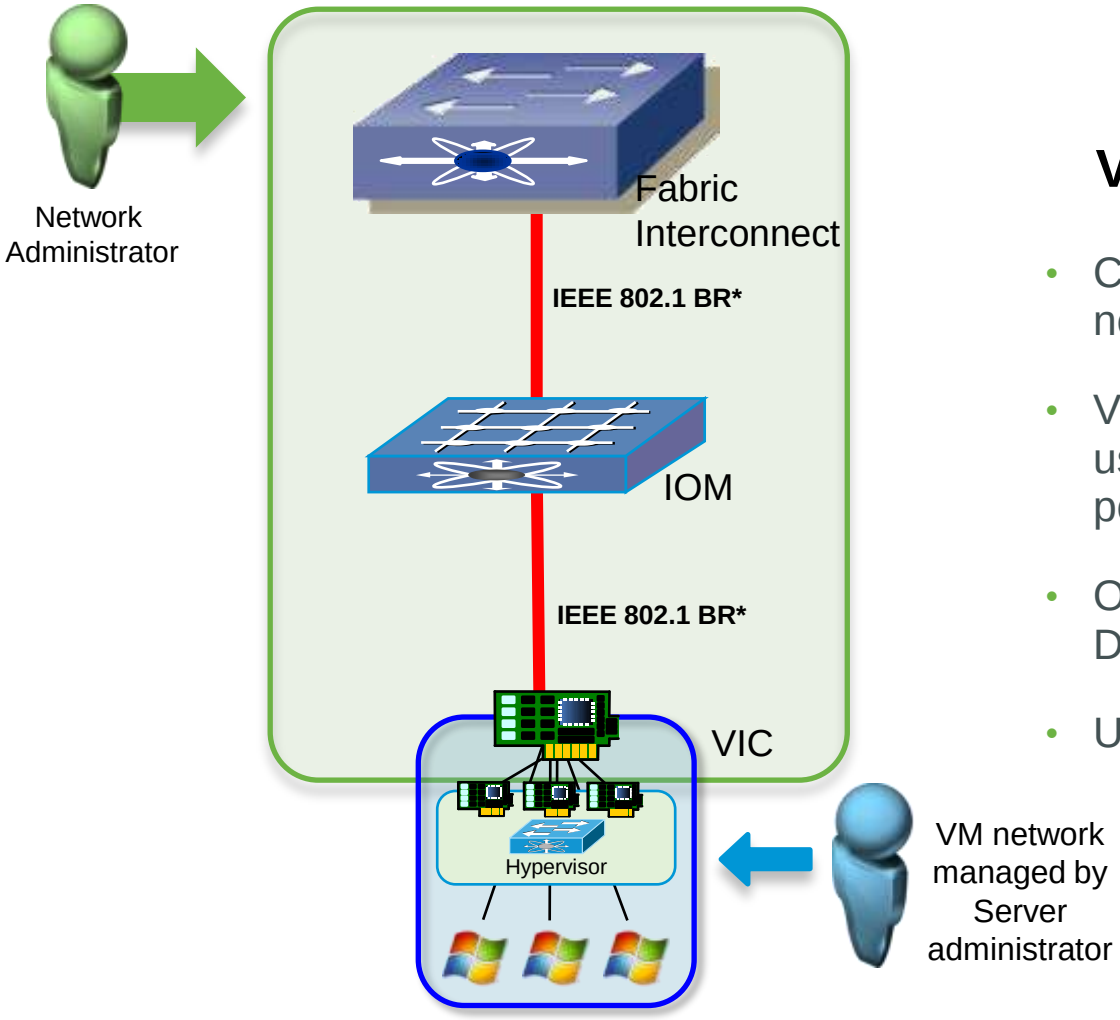
```
UCS-6100-A# connect adapter 1/1/1
adapter 1/1/1 # connect
adapter 1/1/1 (top):1# attach-mcp
adapter 1/1/1 (mcp):1# iscsi_ping
id  name            tgt          address  port tcp ping status
--- --------------- --- --------------- ----- ----------------
----------------
  6 vnic_2            0  100.65.160.181  3260 Not issued
  6 vnic_2            1  100.65.160.181  3260 Not issued
  7 vnic_3            0  100.65.160.145  3260 Not issued
  7 vnic_3            1  100.65.160.145  3260 Not issued
adapter 1/1/1 (mcp):2# iscsi_ping start

adapter 1/1/1 (mcp):3# iscsi_ping
id  name            tgt          address  port tcp ping status
--- --------------- --- --------------- ----- ----------------
----------------
  6 vnic_2            0  100.65.160.181  3260 Success (2.740ms)
  6 vnic_2            1  100.65.160.181  3260 Success (2.666ms)
  7 vnic_3            0  100.65.160.145  3260 Success (2.555ms)
  7 vnic_3            1  100.65.160.145  3260 Success (2.261ms)
```

UCS 2.0
VMFex with vMotion

VM-FEX

Extending the FEX Architecture to VM's



One Network

Virtual Same As Physical

- Consolidates virtual and physical network
- VM vNIC attached to the network using VM-FEX and gets a dedicated port on switch
- Operates in Standard (Emulated) or DirectPath I/O (UPT) Mode
- Uses Pre-standard IEEE 802.1BR

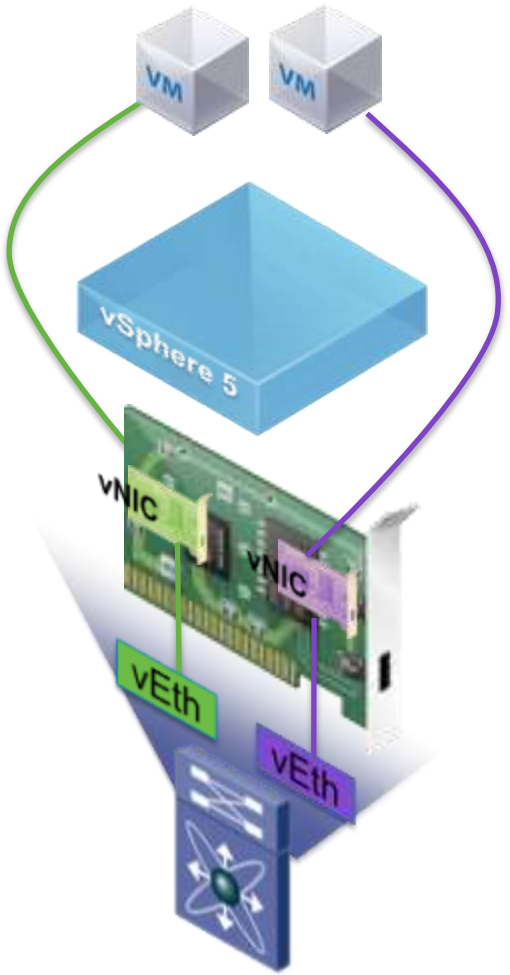
*IEEE 802.1BR pre-standard

DirectPath I/O with vMotion

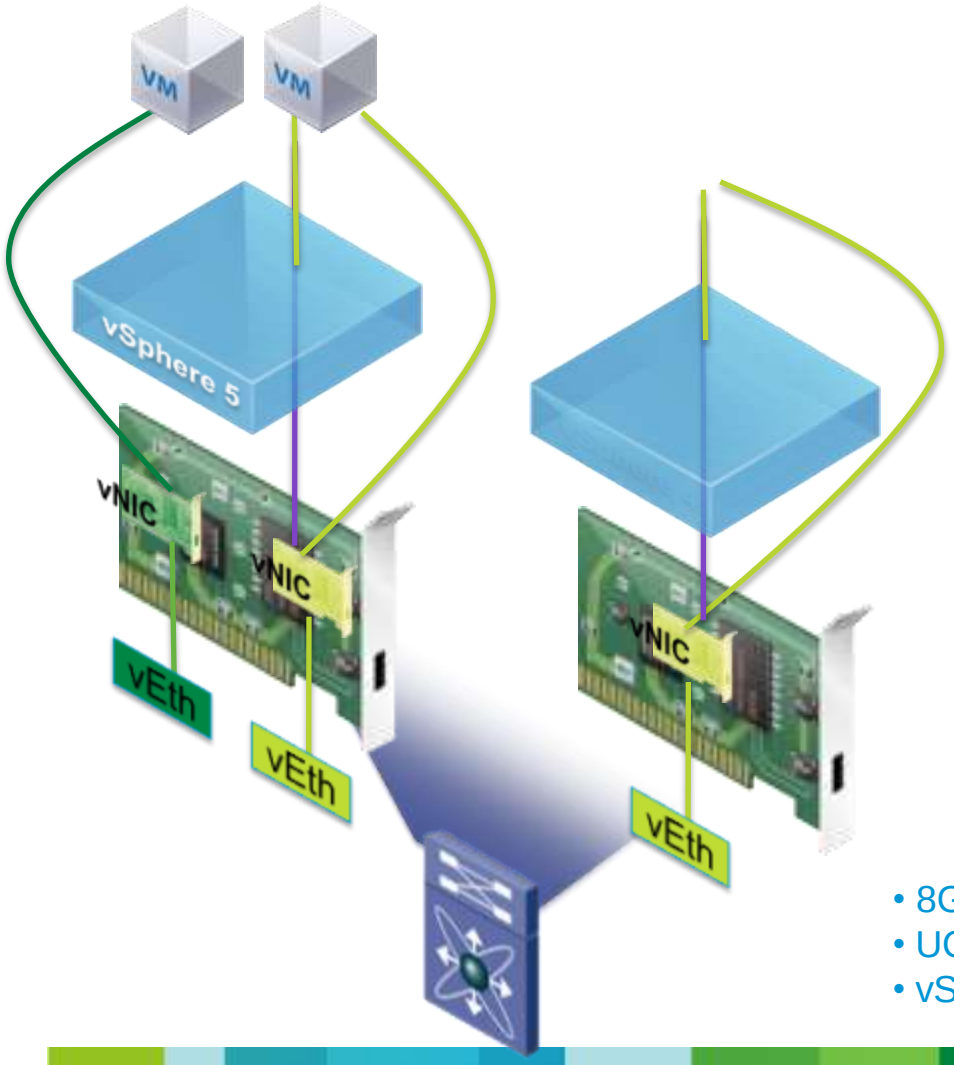
- Data traffic from VM bypasses the hypervisor
  - Take care about OS support
- Available in vSphere 5
- Other names it is known by
  - VM-FEX High Performance Mode
  - UPT
  - VMDirectpath Gen2
- Co-exists with standard mode – Cisco VIC required
- UCS specifically called out in vSphere 5 Networking Guide (Page 42)

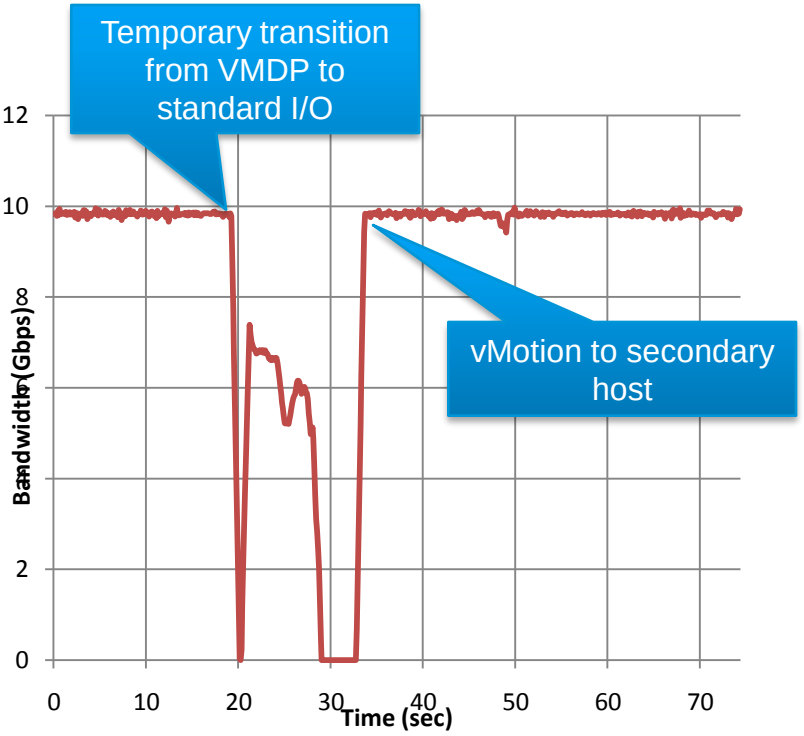
You can enable DirectPath I/O with vMotion for virtual machines in a datacenter on a Cisco UCS system that has at least one supported Cisco distributed switch.

DirectPath I/O with vMotion



VMDirectPath VMotion in action





- 8GB VM, sending UDP stream using pktgen (1500MTU)
- UCS B200 blades with UCS VIC card
- vSphere 5

DirectPath I/O

- Data traffic from VM bypasses the hypervisor
- Does not follow the vNetwork distributed switch model – PCI devices explicitly assigned
- Following features unavailable

  vMotion / DRS

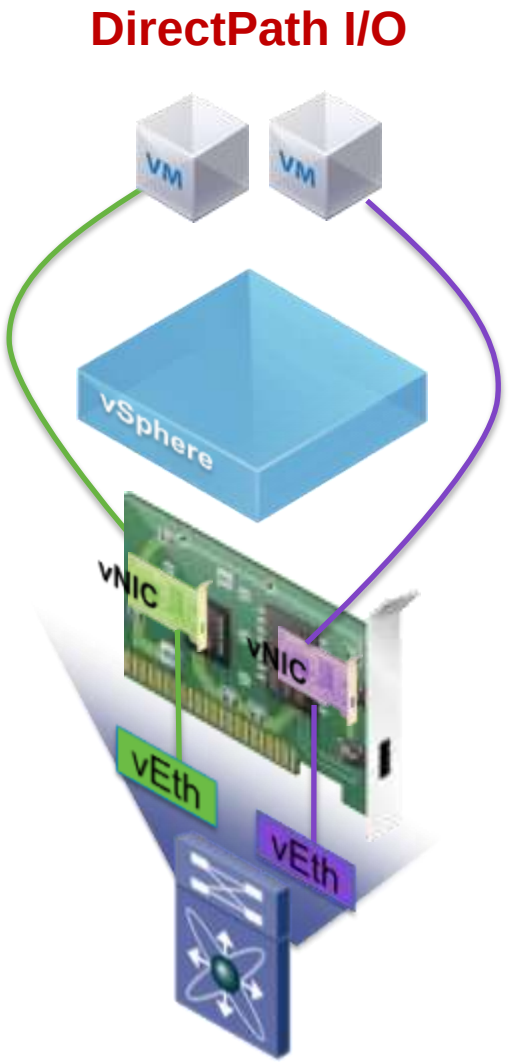
  Suspend and resume

  Record and replay

  Fault tolerance

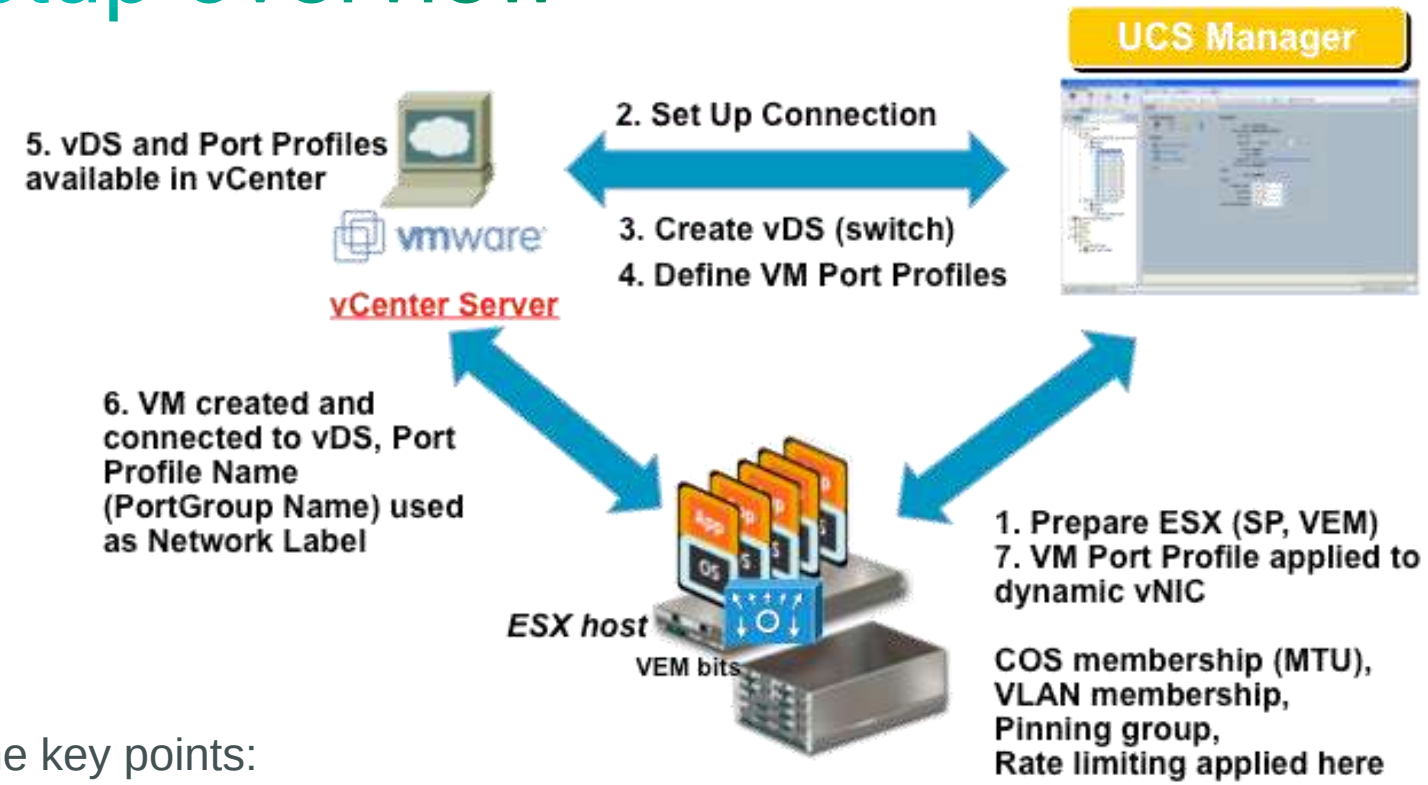
  High availability

  Snapshots



Not Recommended / Tested on Cisco VIC

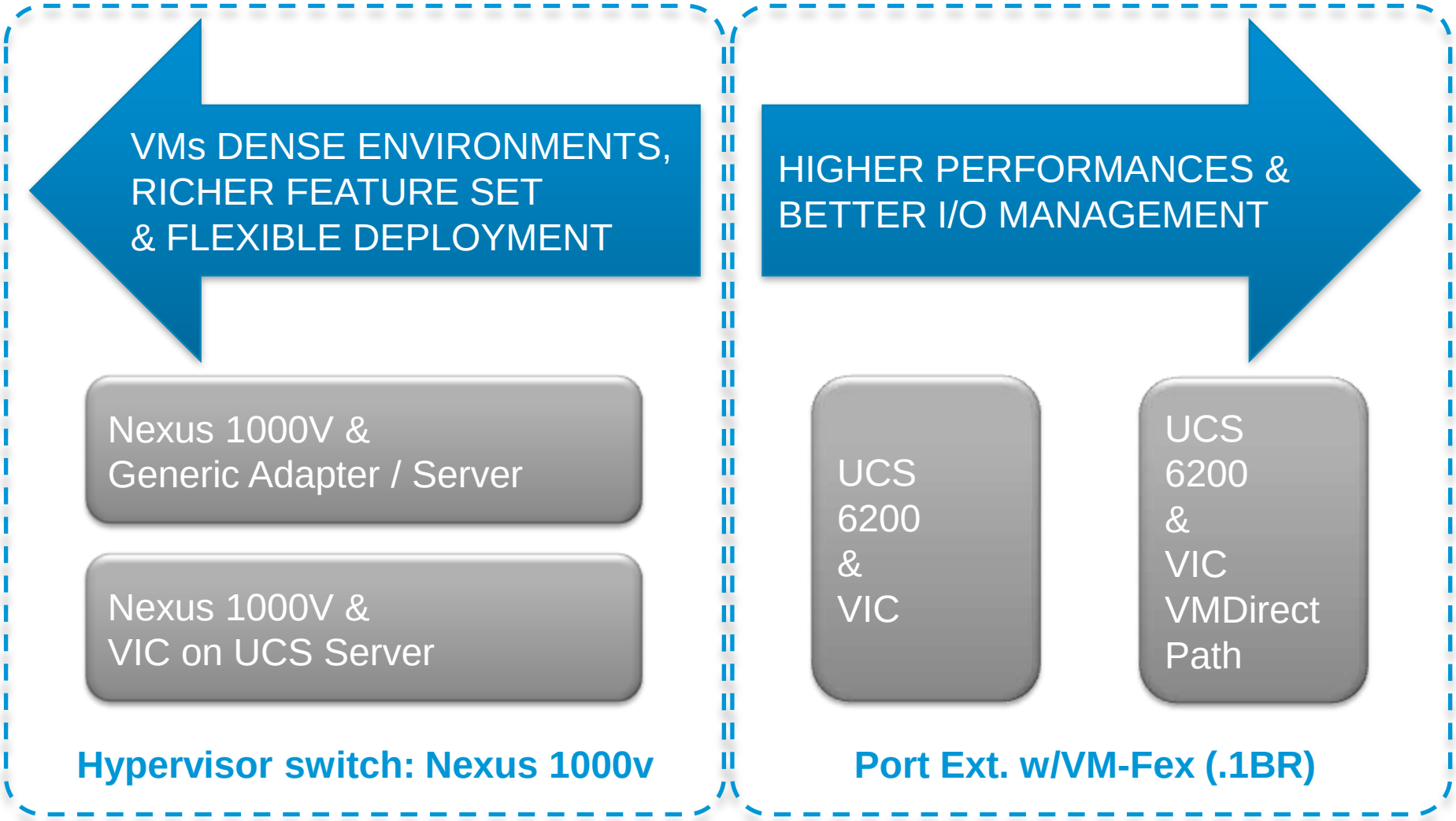
Setup overview



Some key points:

- Always use VEM bits from http://your-UCSM/vmfex/vmfex.html
- One ESX cannot be part of a N1000v and VMFEX in the same time
- If you move one ESX from one to the other, uninstall / reinstall VEM bits will make things easier (VEM keep track of VMFEX / N1kv modes)

Why different models?

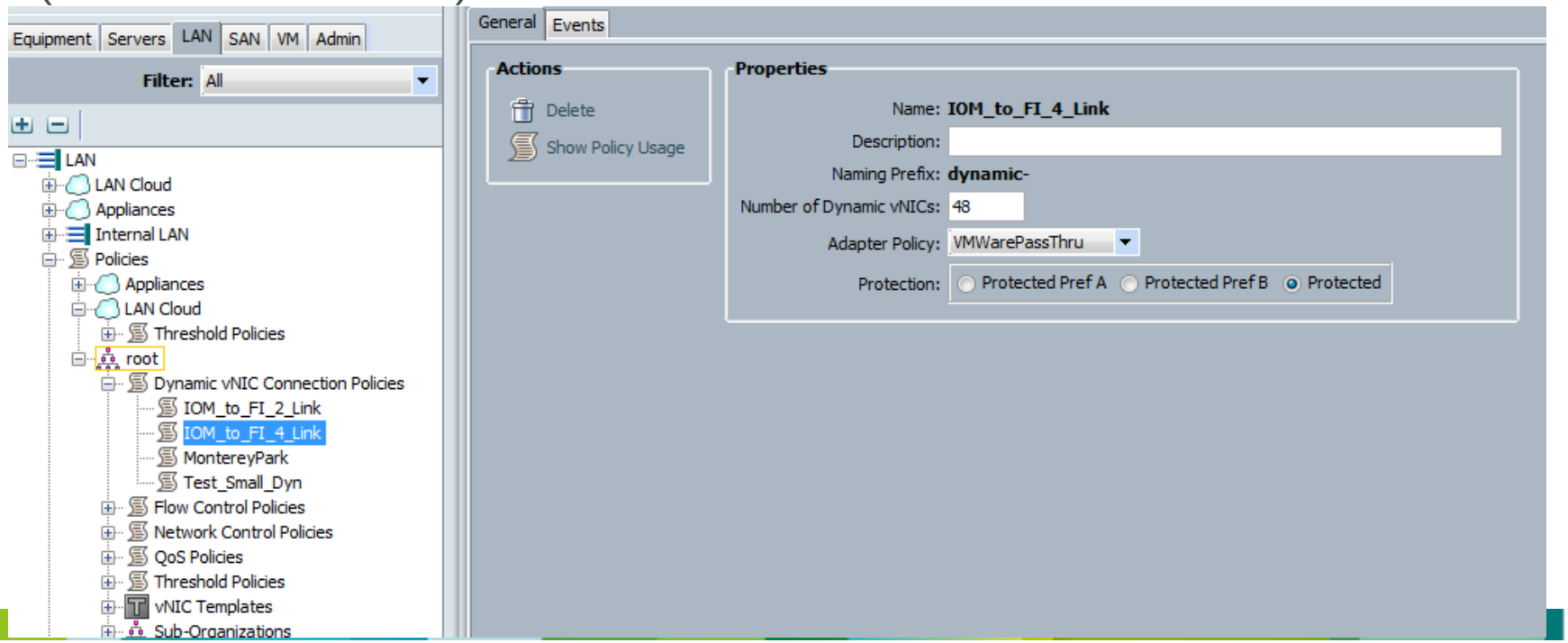


DirectPath I/O with vMotion

Creating Dynamic vNICs

- Policies are to automatically provision dynamics vNICs on Servers
- Dependent on the number of Fabric Interconnect to IO Module connections

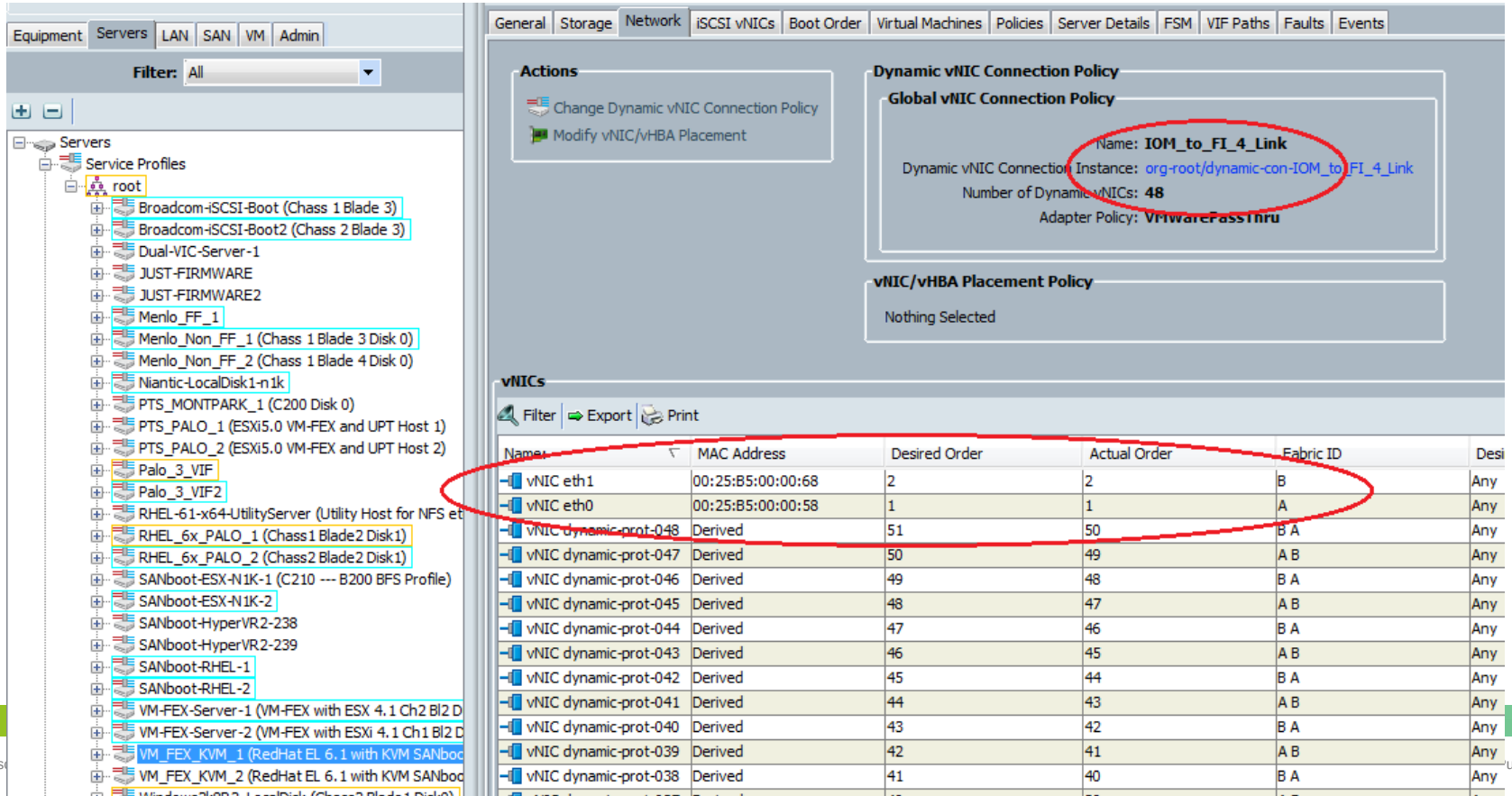
  (# IOM to FI links * 63) – 2 on Gen2



DirectPath I/O with vMotion

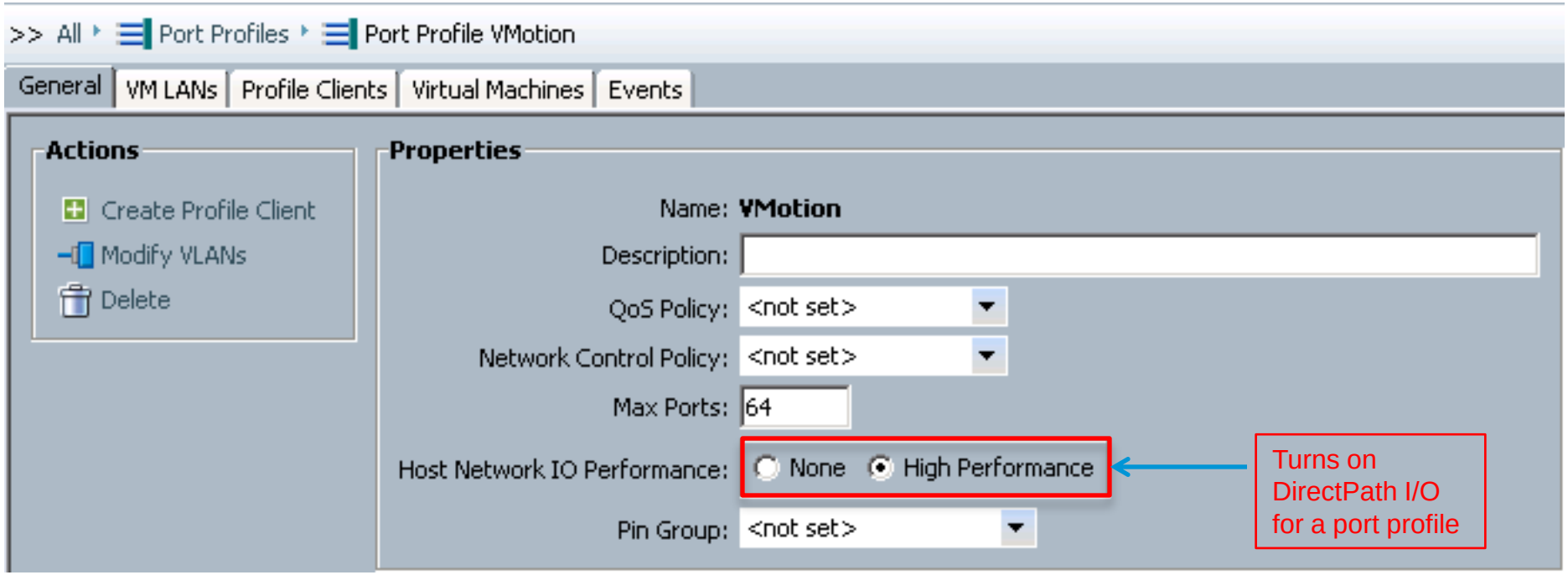
Building Service Profile

- 2 Statics – 1 to each UCS Fabric
- Change dynamic vNIC connection policy to setup dynamics
- Keep cool: Dynamic vNIC are hidden in vCenter (different PCI device ID)



DirectPath I/O with vMotion

Configure port-profile

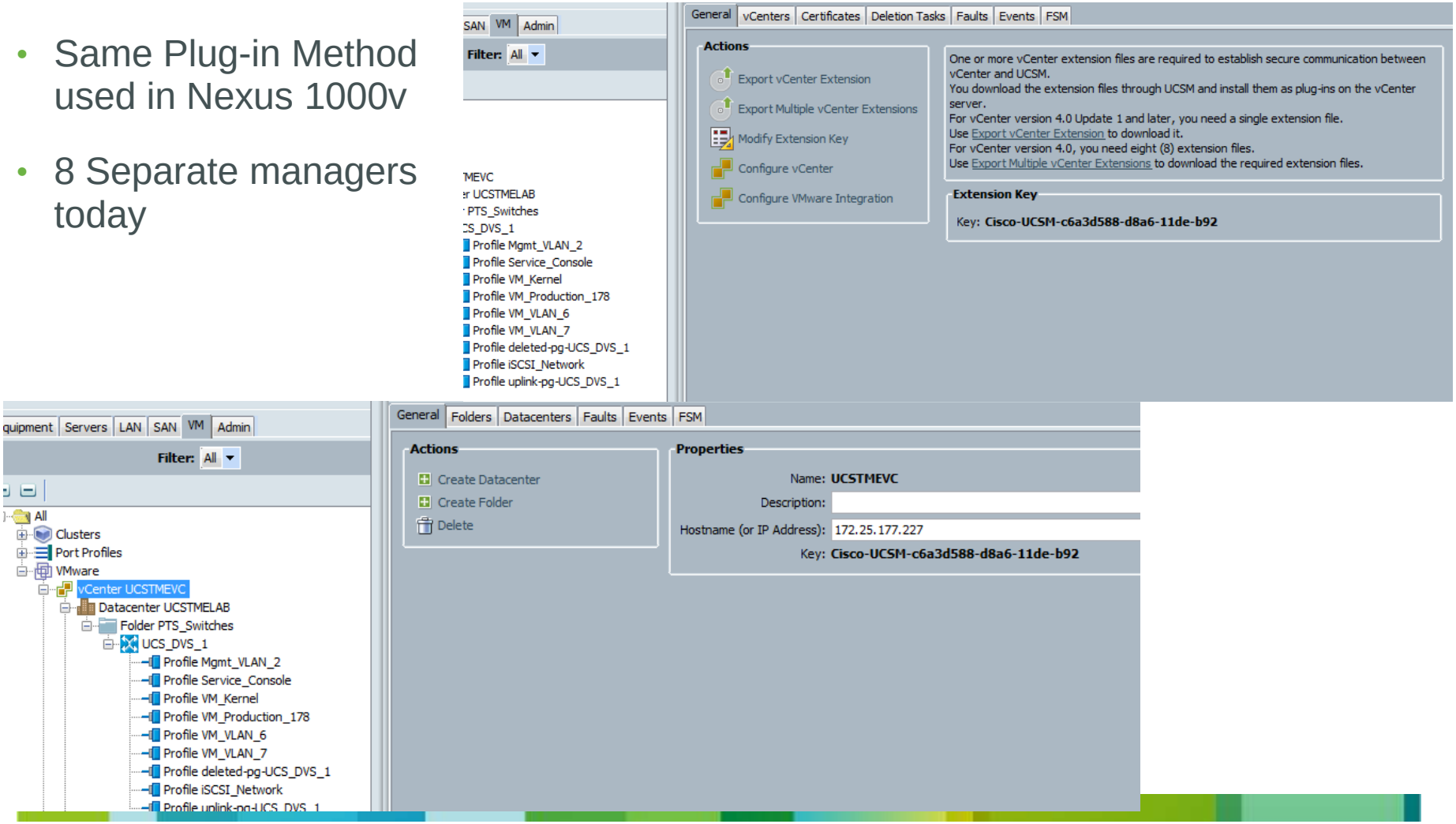


You can / should use VM FEX for all your interfaces (VMs, vmk…)

DirectPath I/O with vMotion

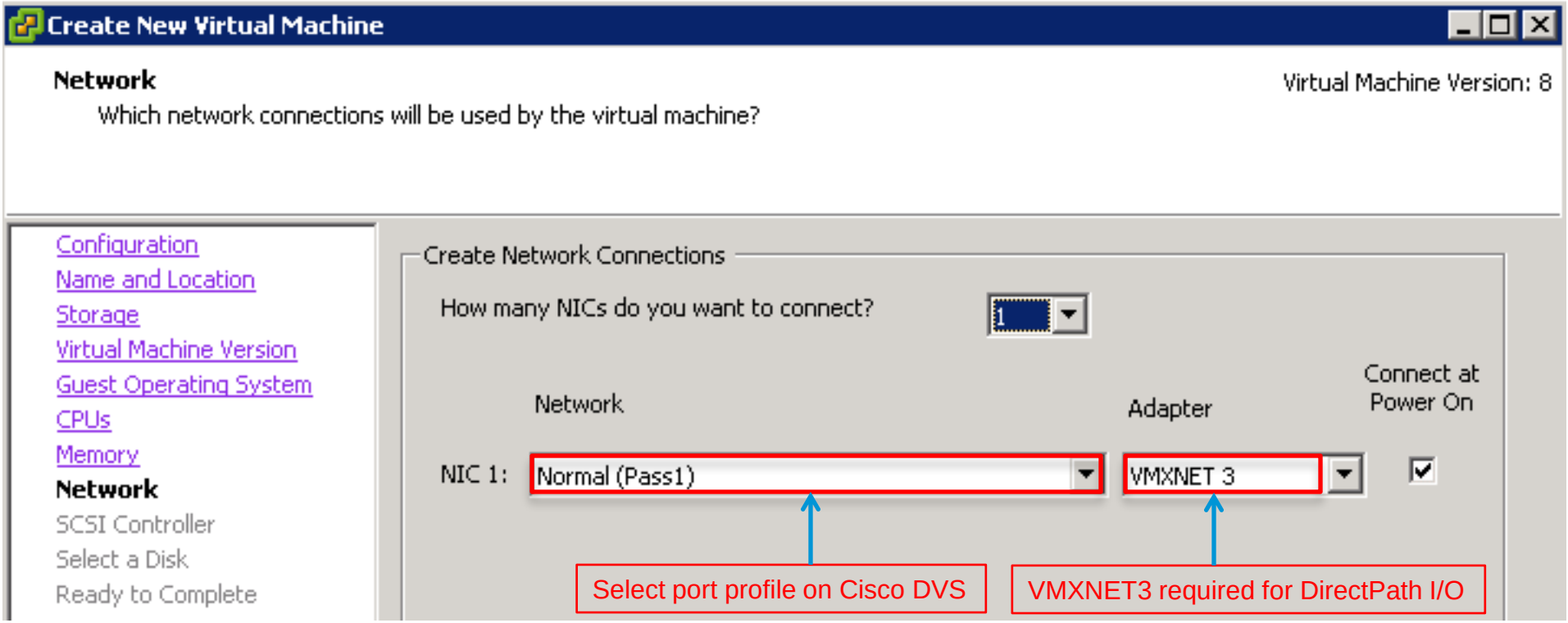
Communication with Manager

- Same Plug-in Method used in Nexus 1000v
- 8 Separate managers today



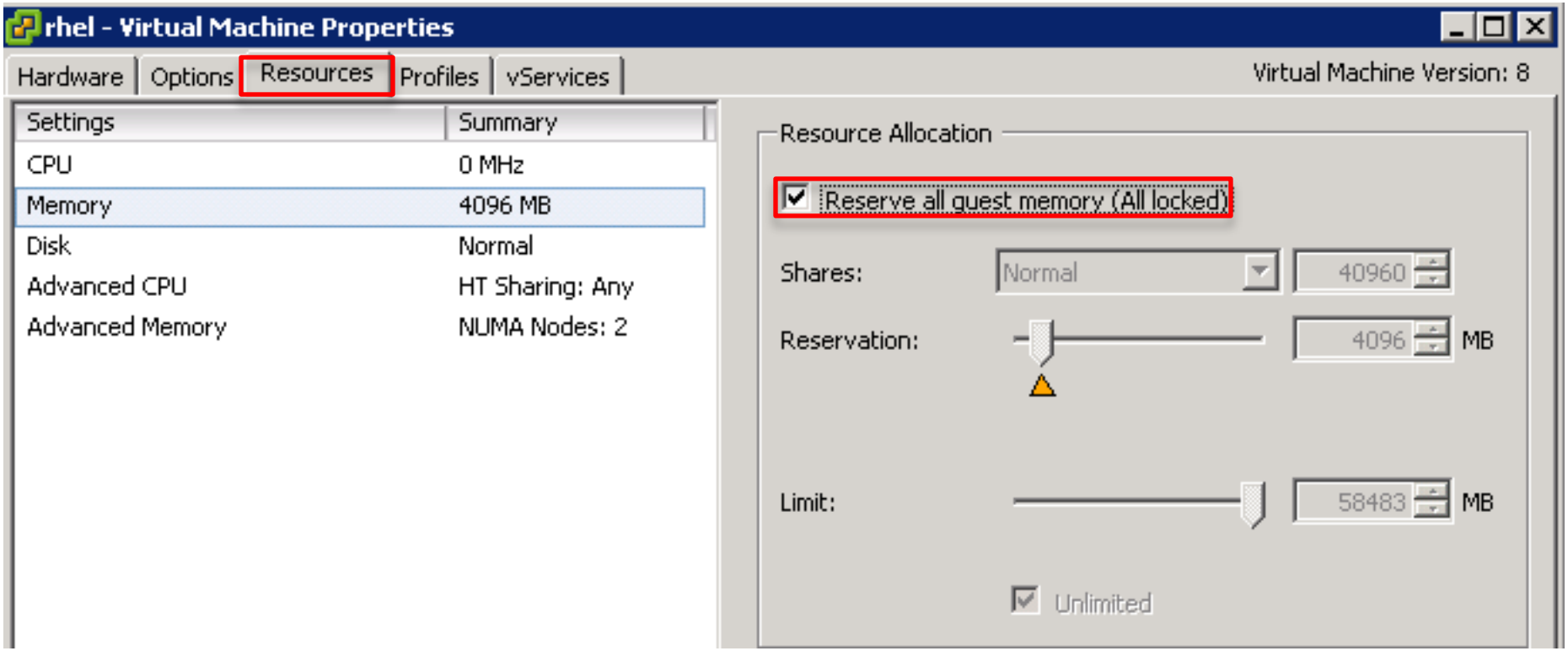
DirectPath I/O with vMotion

Configuration in VCenter

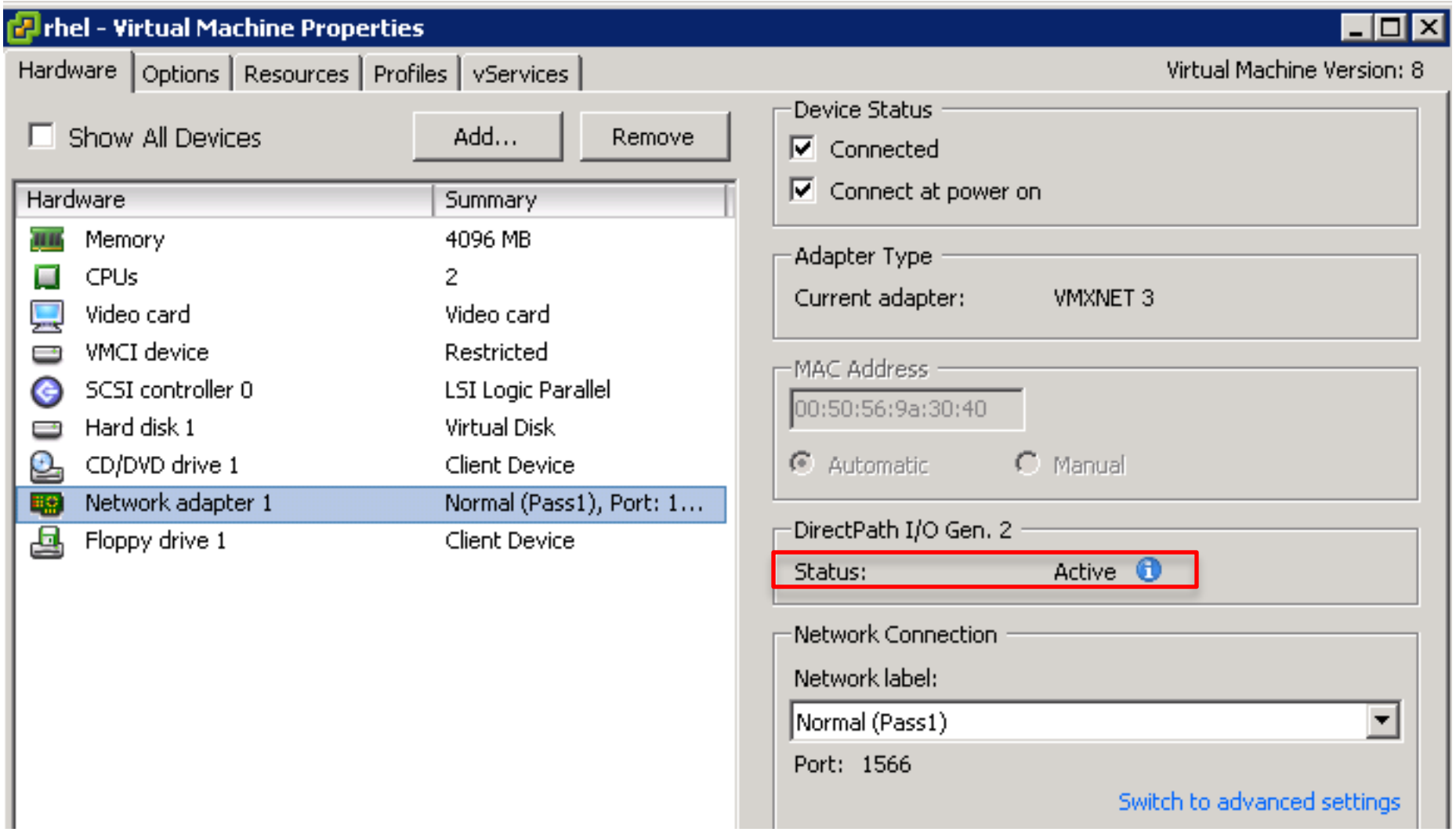


DirectPath I/O with vMotion

Configuration in VCenter



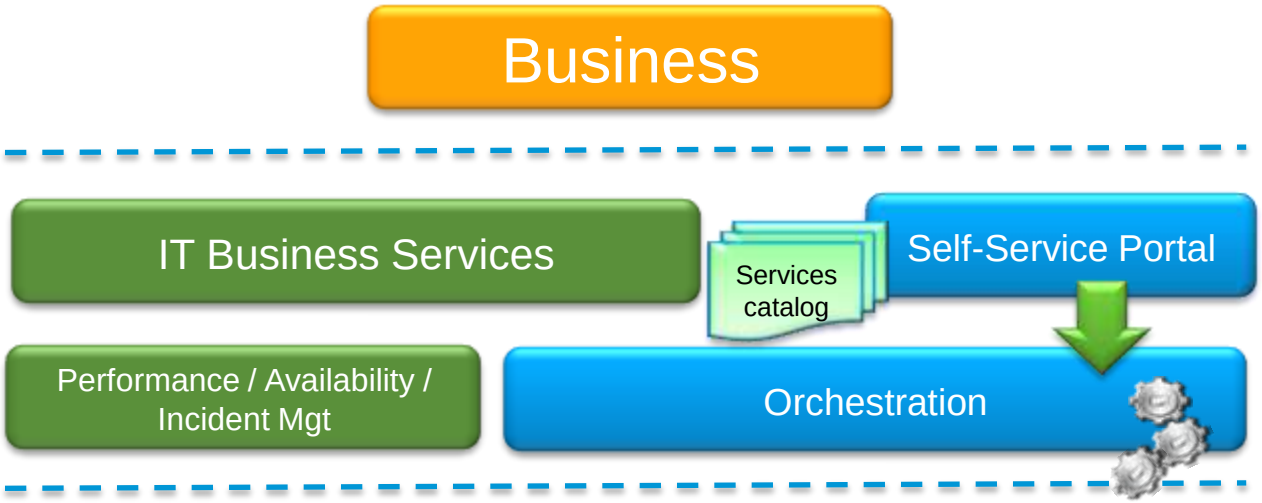
DirectPath I/O Status

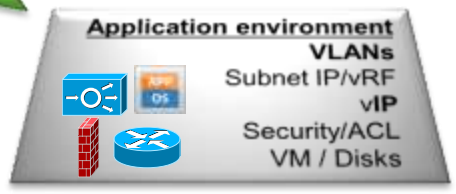


Customers' new subject of interest

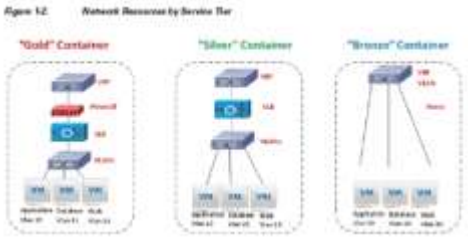
Cloud-ready platform

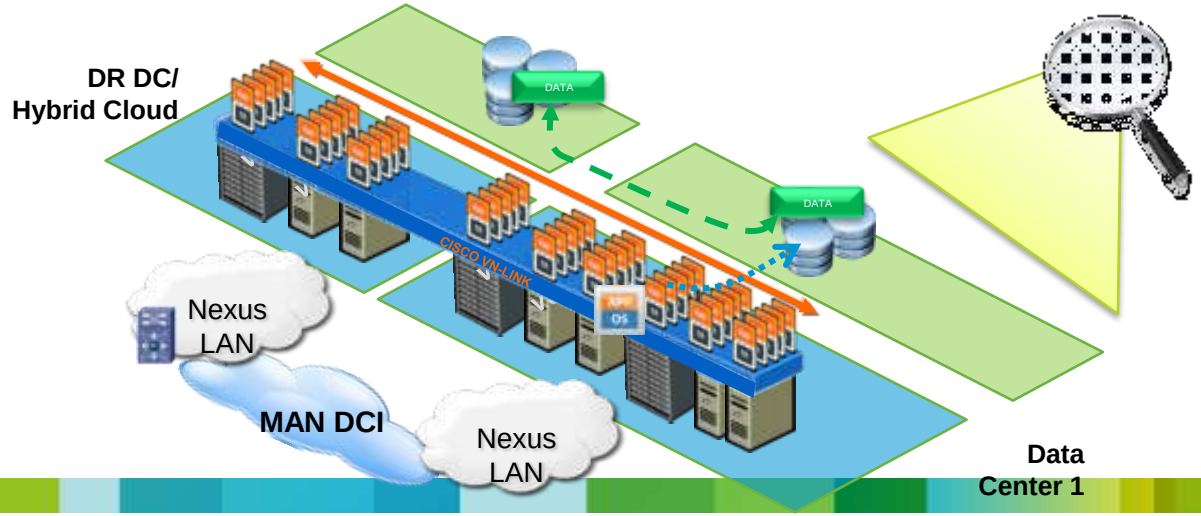
On-demand/Self provisioning, Elastic, Pay-per-use, Secure



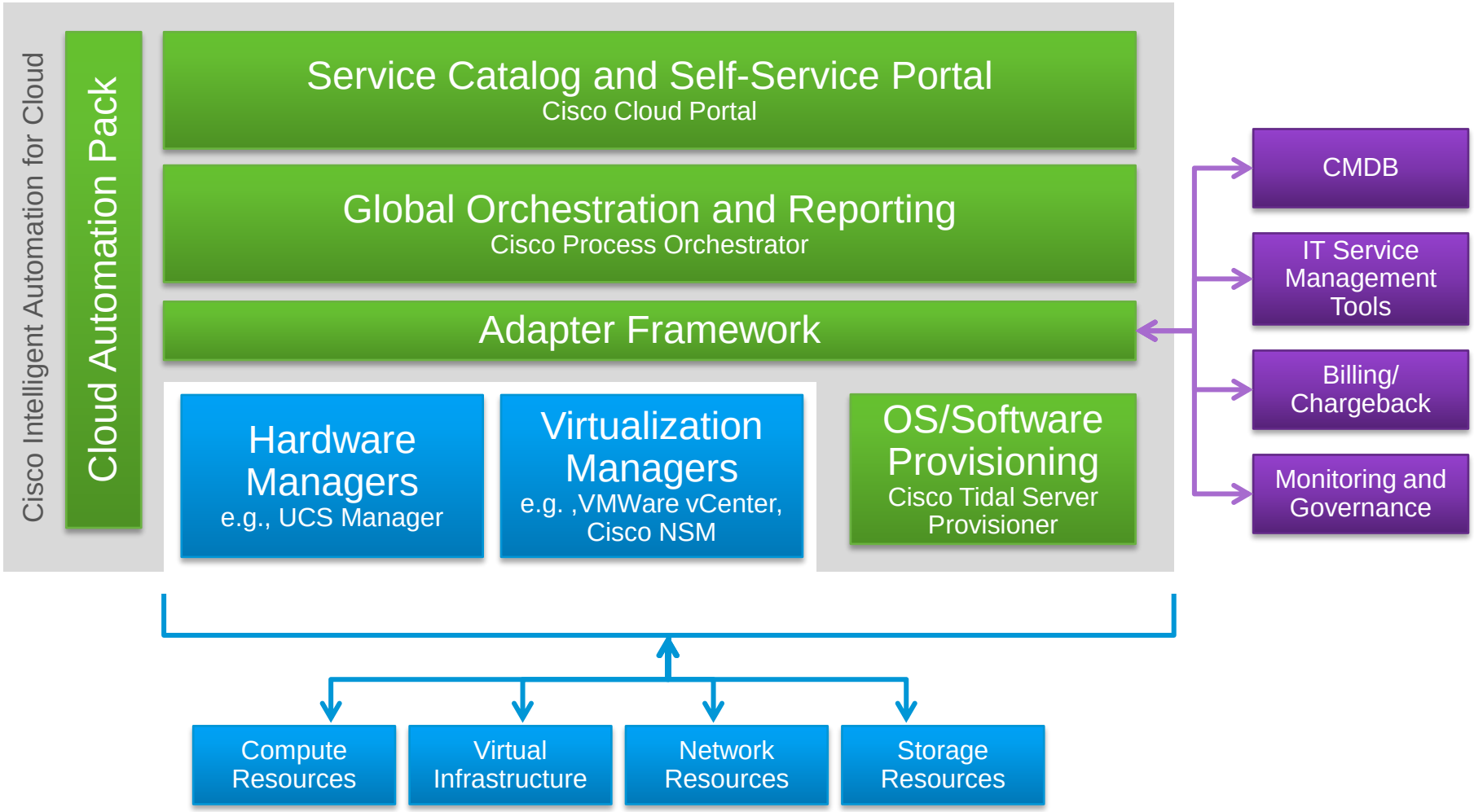


Service Tiering in Environment Container





Cisco Intelligent Automation for Cloud



Useful links

- UCS B-Series documentation roadmap
  - Search „UCS Roadmap“ on cisco.com

- Cisco Developer Network
  - UCS Manager Emulator
  - UCS Dashboard
  - goUCS
  - http://developer.cisco.com/web/unifiedcomputing/home

Otázky a odpovědi

- Twitter www.twitter.com/CiscoCZ
- Talk2Cisco www.talk2cisco.cz/dotazy
- SMS 721 994 600

- Zveme Vás na **Ptali jste se…** v sále **LEO**
  1.den 17:45 – 18:30
  2.den 16:30 – 17:00

Kód přednášky

Prosíme, ohodnoťte tuto přednášku.