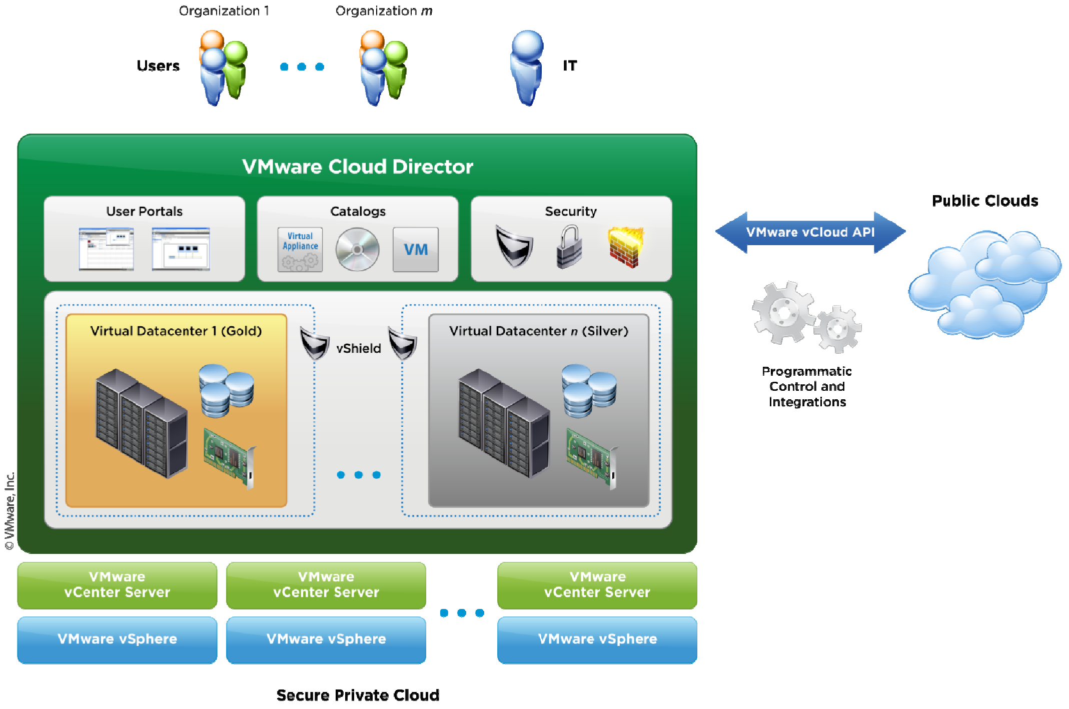


Vcloud Director Networking

Packet Flow
‘Revealed’

(well , many details here not documented yet;-)

Explaining how VCD networking is built and what is the resulted packet flow.

- External routed (currently always NATed) networks.
- External direct networks.
- Network Pools: VCDNI-backed, Vlan-backed and Port-Group-backed.

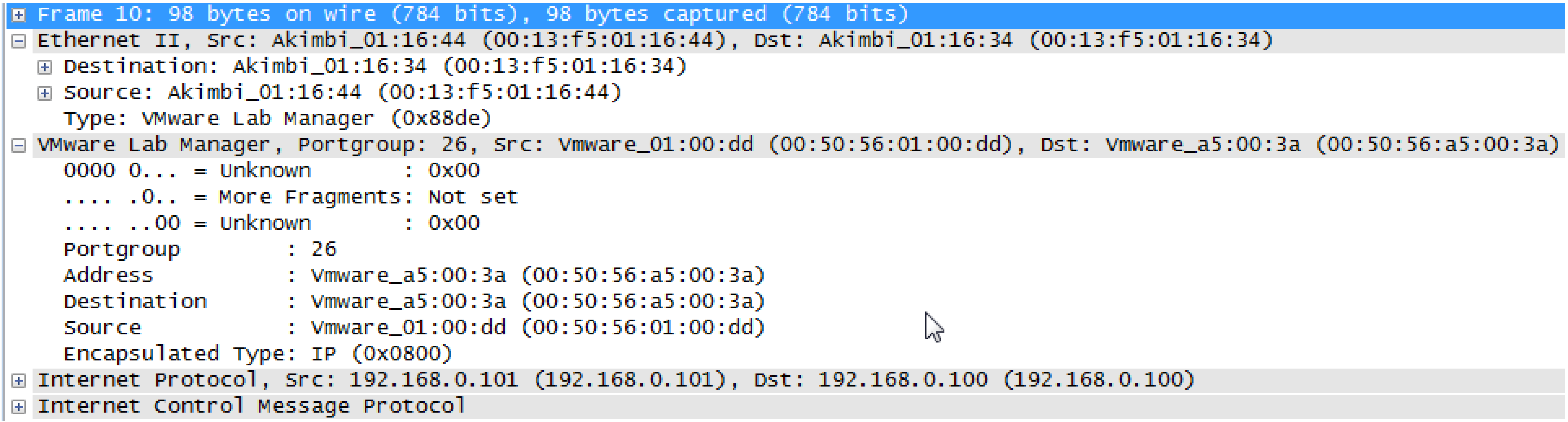
Organization Networks are built from all the above.

You just need to know vcenter and vDS ‘legacy’ networking to understand this …

You also need to know about two vmware services used by VCD:

Vshield-Port-Group-Isolation (PGI) technology

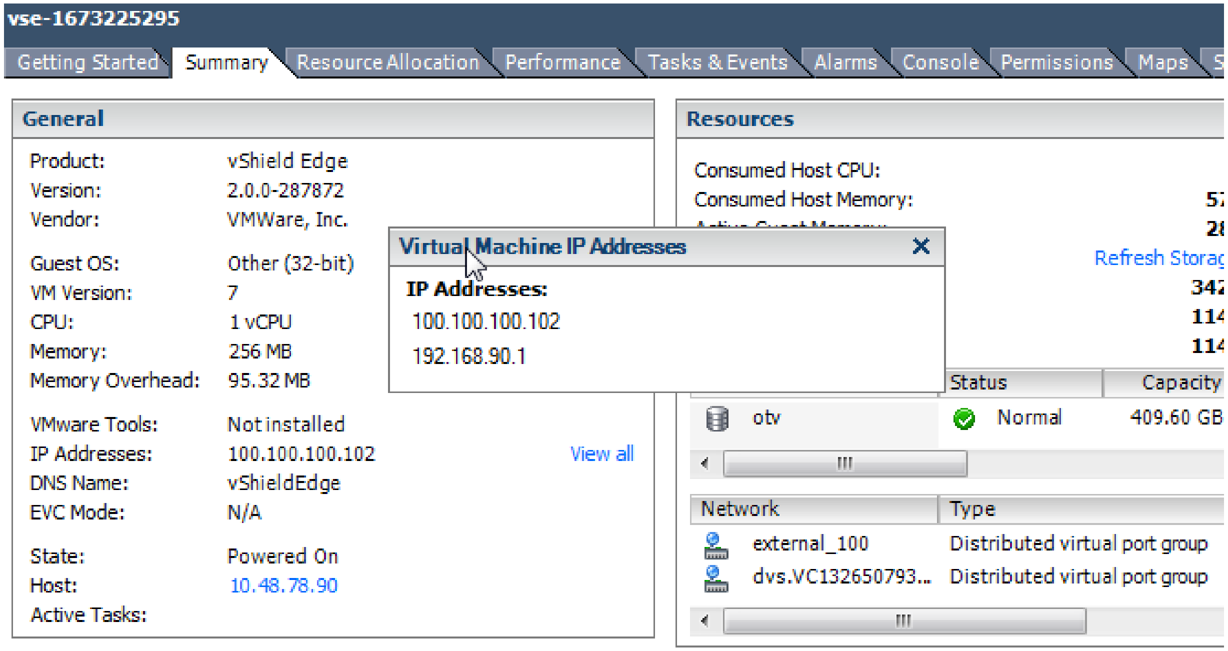
- It was a special vAPP (by akimbi requisition) running as a VM (linux2.4.31)
- PGI-VM was created on each ESX host for ‘isolated networks’ in lab-manager.
- It was implemented as a bridge device placed in-line between VM vNIC and the ESX host external vmNIC. It had many vNIC interfaces (bridged).
- Today (ESX 4.0u2 and up) in VCD, all ORGs (tenants) use a PGI code running on the vDS on each ESX host for doing the encapsulation needed per tenant (more on this later).
- It is encapsulating a VM ‘regular’ Ethernet frames on a special ‘lab-manager’ ‘88de’ ether-type frames (not 802.1ah MAC-in-MAC but something proprietary created by akimbi/vmware), this needs to be supported on external switches:



```
⊞ Frame 10: 98 bytes on wire (784 bits), 98 bytes captured (784 bits)
⊟ Ethernet II, Src: Akimbi_01:16:44 (00:13:f5:01:16:44), Dst: Akimbi_01:16:34 (00:13:f5:01:16:34)
  ⊞ Destination: Akimbi_01:16:34 (00:13:f5:01:16:34)
  ⊞ Source: Akimbi_01:16:44 (00:13:f5:01:16:44)
    Type: VMware Lab Manager (0x88de)
⊟ VMware Lab Manager, Portgroup: 26, Src: Vmware_01:00:dd (00:50:56:01:00:dd), Dst: Vmware_a5:00:3a (00:50:56:a5:00:3a)
    0000 0... = Unknown       : 0x00
    .... .0.. = More Fragments: Not set
    .... ..00 = Unknown       : 0x00
    Portgroup          : 26
    Address            : Vmware_a5:00:3a (00:50:56:a5:00:3a)
    Destination        : Vmware_a5:00:3a (00:50:56:a5:00:3a)
    Source             : Vmware_01:00:dd (00:50:56:01:00:dd)
    Encapsulated Type: IP (0x0800)
⊞ Internet Protocol, Src: 192.168.0.101 (192.168.0.101), Dst: 192.168.0.100 (192.168.0.100)
⊞ Internet Control Message Protocol
```

Vshield-Edge services-VM (VES) (1)

- A special vAPP running as a VM with 2 vNIC interfaces (linux2.6.18.8).
- This VES is at the heart of the VCD networking concepts (many of those used)
- It is created and run any time VCD needs DHCP ,NAT, FW GW services.
- VCD creates one per every network that needs those services (multiple VES might be created for the same Organization).
- It is a 2 vNIC device providing 'inside' and 'outside' L3 IP interfaces to be connected between different port-groups/networks. It is doing NAT between inside and outside (It is NOT a router, it is a NAT-only device).



General

Product:	vShield Edge
Version:	2.0.0-287872
Vendor:	VMWare, Inc.
Guest OS:	Other (32-bit)
VM Version:	7
CPU:	1 vCPU
Memory:	256 MB
Memory Overhead:	95.32 MB
VMware Tools:	Not installed
IP Addresses:	100.100.100.102
DNS Name:	vShieldEdge
EVC Mode:	N/A
State:	Powered On
Host:	10.48.78.90
Active Tasks:	

Virtual Machine IP Addresses

IP Addresses:
100.100.100.102
192.168.90.1

Resources

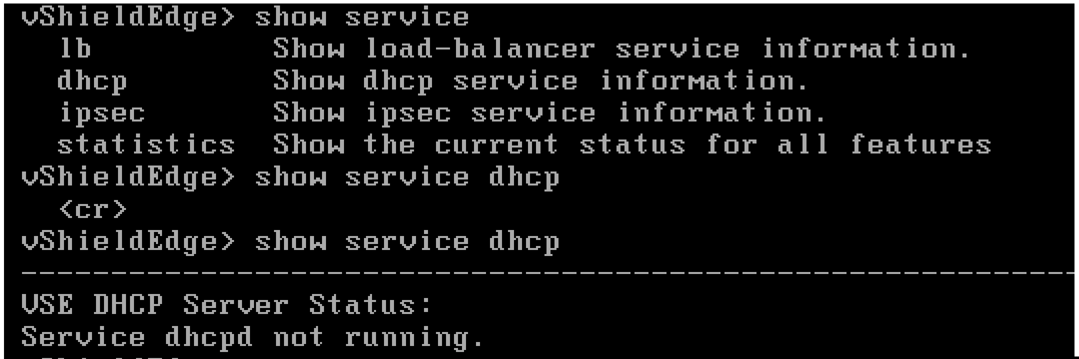
Consumed Host CPU:
Consumed Host Memory: 57
28
342
114
114

	Status	Capacity
otv	Normal	409.60 GB

Network	Type
external_100	Distributed virtual port group
dvs.VC132650793...	Distributed virtual port group

Vshield-Edge services-VM (VES) (2)

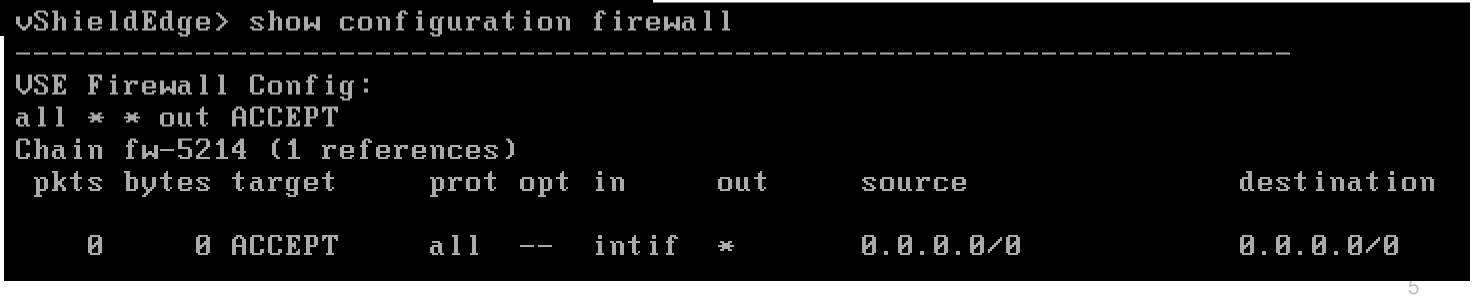
- It's 'outside' interface IP address is defined by the static IP (pool) you define for the external network it is attached to, it also gets a default GW for 'outside'
- It's 'inside' interface IP address is defined by the static/DHCP IP (pool) you define for the internal network it is attached to, it also acts as a default GW for 'inside' network.
- It might be deployed by VCD on a different ESX host then the one that hosts the actual VMs that needs it's services (then DHCP traverse ESX hosts etc..)
- If it is lost (network/server issues) a backup VM will be initiated by vmware-HA capabilities (this is in the minutes, it is not a stateful failover device).



```
vShieldEdge> show service
  lb          Show load-balancer service information.
  dhcp        Show dhcp service information.
  ipsec       Show ipsec service information.
  statistics  Show the current status for all features
vShieldEdge> show service dhcp
  <cr>
vShieldEdge> show service dhcp
-----------------------------------------------------------
VSE DHCP Server Status:
Service dhcpd not running.
vShieldEdge> _
```

It's services are basic (basic FW, VPN, NAT and DHCP).

Note: default FW rule is permit all (this can not be changed)



```
vShieldEdge> show configuration firewall
-----------------------------------------------------------------------
VSE Firewall Config:
all * * out ACCEPT
Chain fw-5214 (1 references)
 pkts bytes target     prot opt in     out     source               destination

    0     0 ACCEPT     all  --  intif  *       0.0.0.0/0            0.0.0.0/0
```

VCD networking concepts:

- VCD uses the PGI technology and the VES devices to create many kinds of networks/port-groups on vDS.
- It provides those networks as resources to ‘Organizations‘ (different tenants on the ‘cloud’).
- VMs inside ‘Organizations’ are attached to those networks by ORG admin , they use their vNIC to attach to these pre-define port-groups/networks.
- ORG Network can be used for VM-to-VM connectivity– VCD ‘internal’.
- ORG Network can be used for VM-to-outside connectivity – VCD ‘external’.
- Both ‘external’ and ‘internal’ networks are, of course, regular port-groups with a vlan, sending frames through external switch between multiple ESX hosts.
- Only exception is VCDNI which is using a single L2 vlan for many internal networks with a special encapsulation per network per ORG.

Network Pools

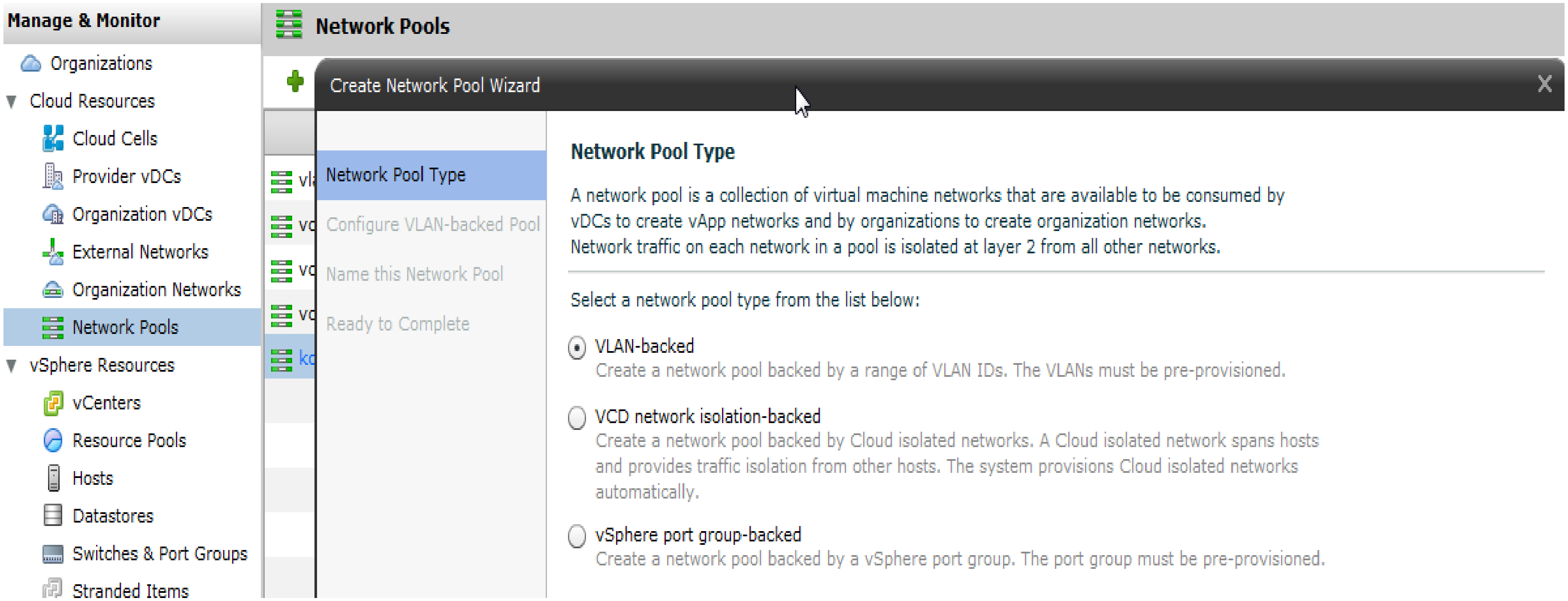
It is NOT a network , it is a Pool of Networks made available ‘on-order’.

But …at the end - it creates networks/vlans that connects between ESX hosts.

They can be used for creating ‘organization internal networks’ ‘on-order’.

They can be used as VSE ‘inside’ networks NATed to external networks on the outside.

It is attached to vDS - it is attached to several ESX host vNICs used as ‘uplinks’.



Network Pool Type

A network pool is a collection of virtual machine networks that are available to be consumed by vDCs to create vApp networks and by organizations to create organization networks. Network traffic on each network in a pool is isolated at layer 2 from all other networks.

Select a network pool type from the list below:

● VLAN-backed
Create a network pool backed by a range of VLAN IDs. The VLANs must be pre-provisioned.

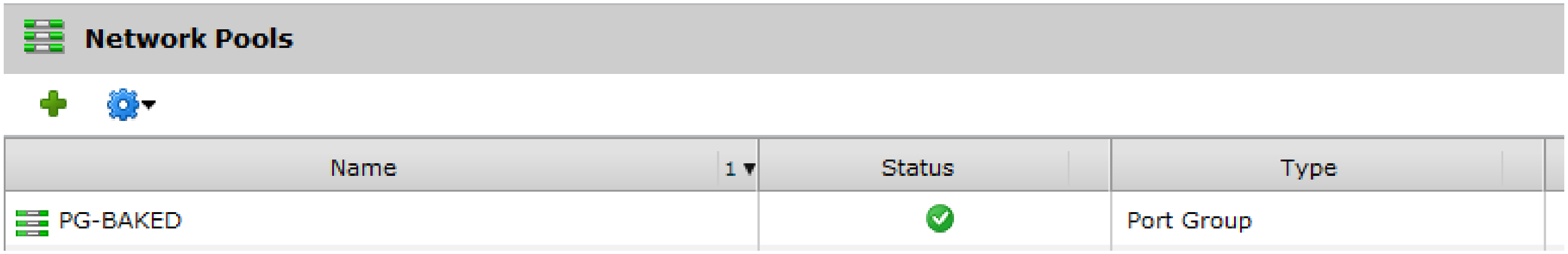
○ VCD network isolation-backed
Create a network pool backed by Cloud isolated networks. A Cloud isolated network spans hosts and provides traffic isolation from other hosts. The system provisions Cloud isolated networks automatically.

○ vSphere port group-backed
Create a network pool backed by a vSphere port group. The port group must be pre-provisioned.

Network Pool type: Port-Group-backed

First you pre-define this port-group on vDS in vcenter, using a regular vlan.

Then you name it in VCD to be used as a pre-defined vlan for ORG VMs.



Network Pools

Name	Status	Type
PG-BAKED	✓	Port Group

It is just like a ‘normal’ definition of a port-group on vDS, you can also define it on Cisco Nexus-1000v and connect VMs to it.

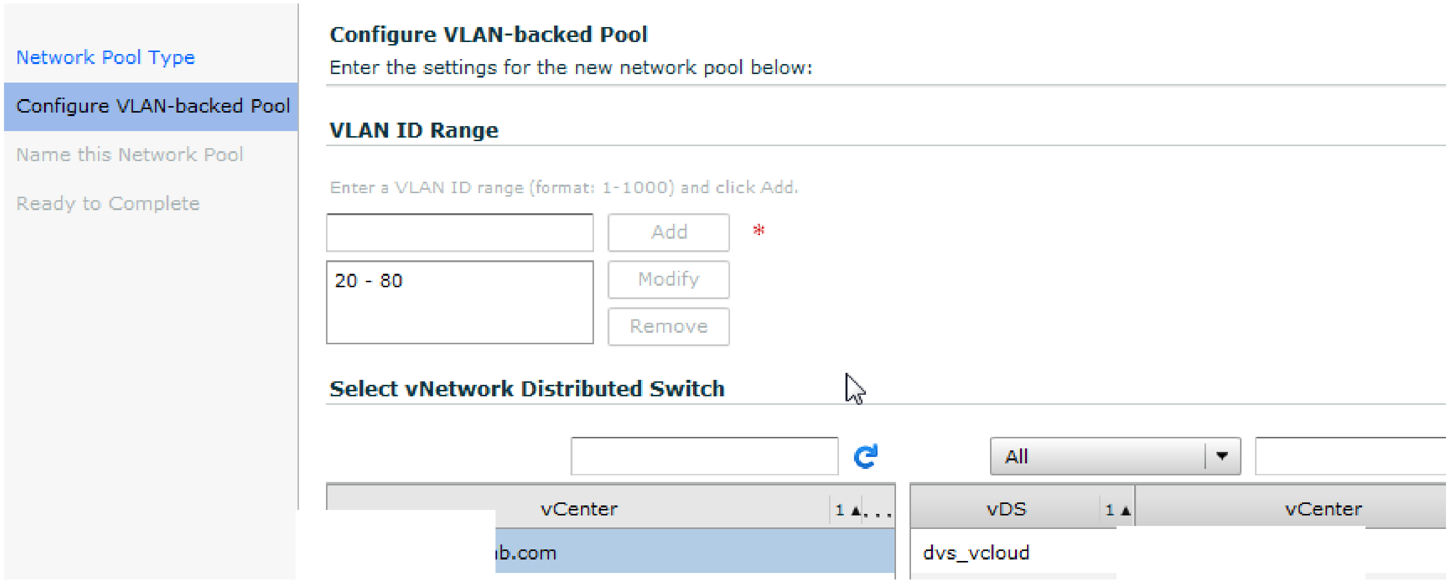
VLAN is sent between different ESX hosts on the external physical switch.

It is exactly the same as ‘external network direct’ connection (see later on).

Network Pool type: Vlan-backed

It is defined only on VCD, it creates regular vlans/port-groups ‘dynamically’ on vDS in vcenter, the key is ‘created when needed’.

You define vlan-range and you name this ‘pool’ to be used for ORG VMs.



Configure VLAN-backed Pool

Enter the settings for the new network pool below:

VLAN ID Range

Enter a VLAN ID range (format: 1-1000) and click Add.

20 - 80

Select vNetwork Distributed Switch

vCenter	vDS	vCenter
b.com	dvs_vcloud	

It is just like ‘normal’ definition of a port-group on vDS, you can define it on Cisco Nexus-1000v and connect VMs to it.

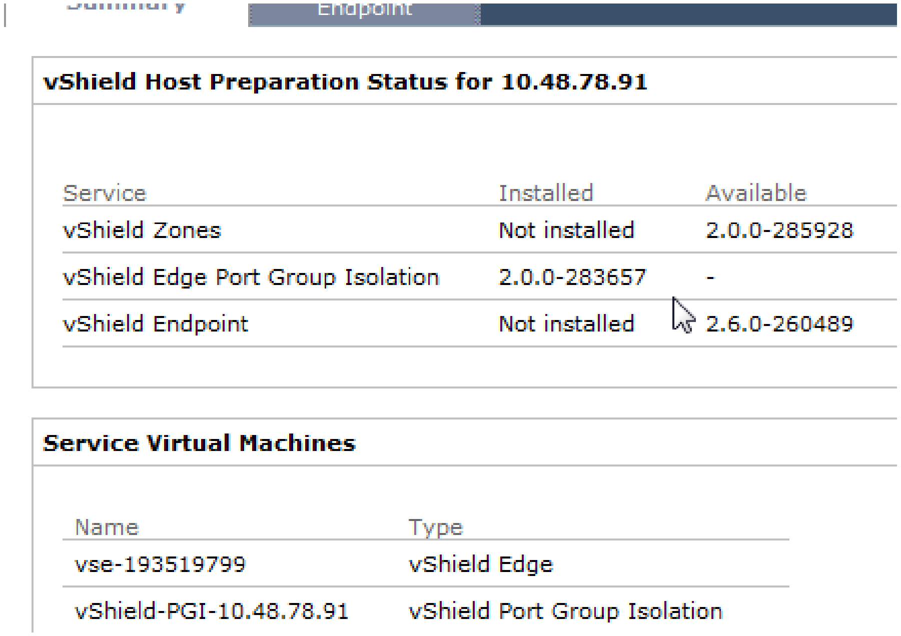
it is <u>exactly the same</u> as ‘external network direct’ connection, once created and connected to external switch (after the dynamic allocation by VCD).

Network Pool type: VCDNI-backed (1)

It is defined only on VCD, it creates new kind of networks, ‘isolated’ by Vmware’s proprietary frame encapsulation (akimbi, ‘lab manager’) made by PGI bridge service on vDS on the ESX host.

Basically it uses a regular vlan on the ‘outside’, but adding another header per-port-group to ‘isolate’ it from other port-groups, so all port-groups will share a common L2 vlan but still be ‘isolated’ from each other by the PGI on vDS.

In order for the isolation technology to be able to work, you need first of all to totally isolate this shared vlan from any other end-host and external routers.



vShield Host Preparation Status for 10.48.78.91

Service	Installed	Available
vShield Zones	Not installed	2.0.0-285928
vShield Edge Port Group Isolation	2.0.0-283657	-
vShield Endpoint	Not installed	2.6.0-260489

Service Virtual Machines

Name	Type
vse-193519799	vShield Edge
vShield-PGI-10.48.78.91	vShield Port Group Isolation

All frames pass through PGI service on vDS in the ESX host in order for it to do the encapsulation of the frames, before they are sent out through the ESX host physical vmNIC.

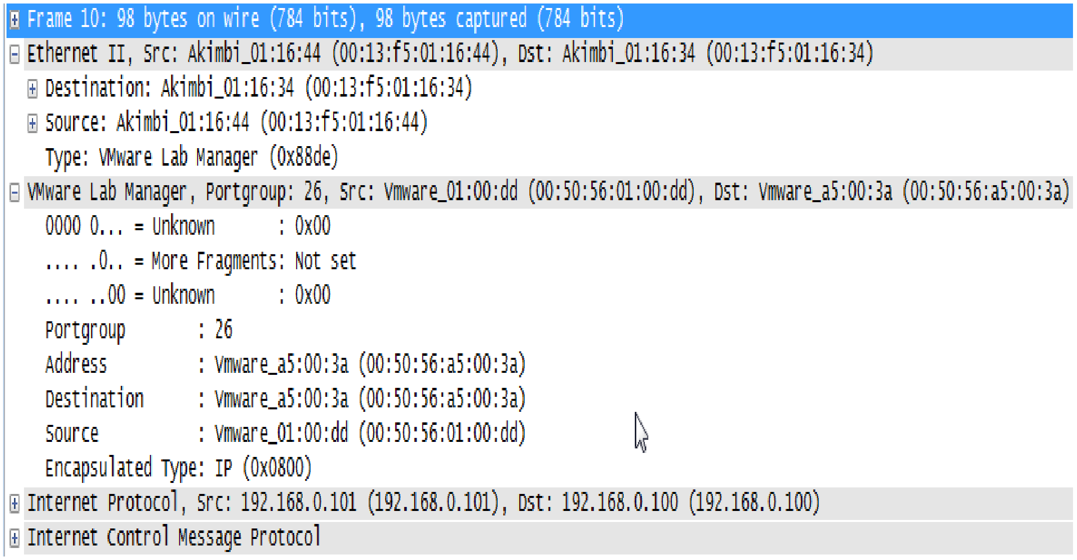
VCD create this service for you automatically when you choose VCDNI network pool…

Since 4.0.u2 it is part of the vDS code.

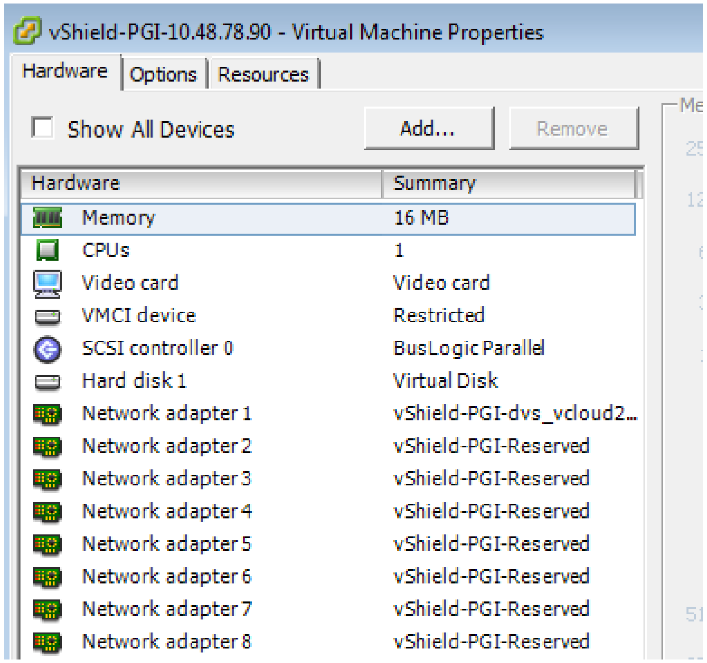
Network Pool type: VCDNI-backed (2)

In VCD you define the number of VCDNI networks and the ‘shared vlan’ used for all ‘isolated’ networks created from this pool.

Note: Use VShield-manager and ESX, pre u2, to create a ‘PGI-VM’ for managing the encapsulation of frames from ORG VMs and test it’s security if you wish to understand more …



```
⊞ Frame 10: 98 bytes on wire (784 bits), 98 bytes captured (784 bits)
⊟ Ethernet II, Src: Akimbi_01:16:44 (00:13:f5:01:16:44), Dst: Akimbi_01:16:34 (00:13:f5:01:16:34)
  ⊞ Destination: Akimbi_01:16:34 (00:13:f5:01:16:34)
  ⊞ Source: Akimbi_01:16:44 (00:13:f5:01:16:44)
    Type: VMware Lab Manager (0x88de)
⊟ VMware Lab Manager, Portgroup: 26, Src: Vmware_01:00:dd (00:50:56:01:00:dd), Dst: Vmware_a5:00:3a (00:50:56:a5:00:3a)
    0000 0... = Unknown       : 0x00
    .... .0.. = More Fragments: Not set
    .... ..00 = Unknown       : 0x00
    Portgroup       : 26
    Address         : Vmware_a5:00:3a (00:50:56:a5:00:3a)
    Destination     : Vmware_a5:00:3a (00:50:56:a5:00:3a)
    Source          : Vmware_01:00:dd (00:50:56:01:00:dd)
    Encapsulated Type: IP (0x0800)
⊞ Internet Protocol, Src: 192.168.0.101 (192.168.0.101), Dst: 192.168.0.100 (192.168.0.100)
⊞ Internet Control Message Protocol
```



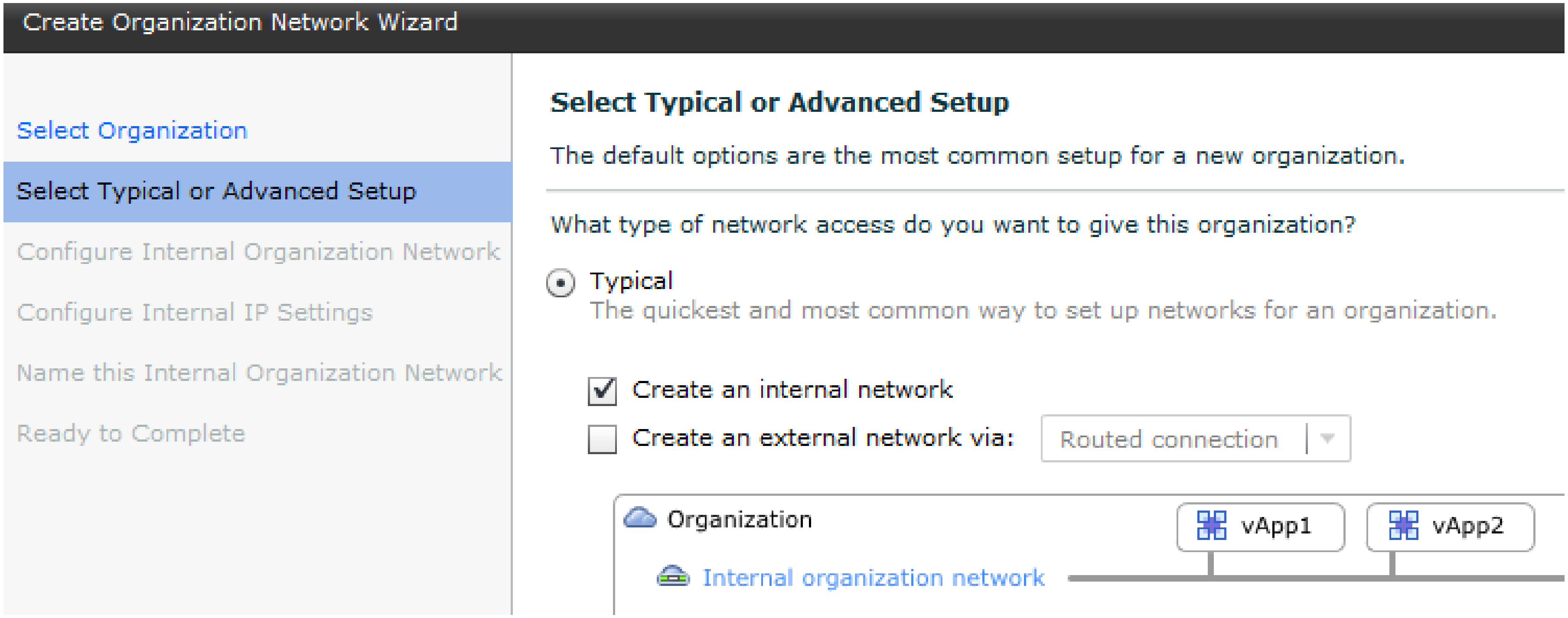
vShield-PGI-10.48.78.90 - Virtual Machine Properties

Hardware	Summary
Memory	16 MB
CPUs	1
Video card	Video card
VMCI device	Restricted
SCSI controller 0	BusLogic Parallel
Hard disk 1	Virtual Disk
Network adapter 1	vShield-PGI-dvs_vcloud2...
Network adapter 2	vShield-PGI-Reserved
Network adapter 3	vShield-PGI-Reserved
Network adapter 4	vShield-PGI-Reserved
Network adapter 5	vShield-PGI-Reserved
Network adapter 6	vShield-PGI-Reserved
Network adapter 7	vShield-PGI-Reserved
Network adapter 8	vShield-PGI-Reserved

VCDNI is NOT supported by Nexus1000V, it is not doing this encapsulation, It is still vlan based.

Organization ‘internal’ networks

- A network defined for an organization in VCD, it is created from the 3 types of ‘network pools’ and it is ‘isolated’ for specific set of organization VMs that connects to it if needed by the organization administrator.
- In any case it is deployed on multiple ESX hosts using vDS or Nexus1000V, and frames are sent between ESX hosts for any type of ‘isolated network’.


Create Organization Network Wizard
Select Organization
Select Typical or Advanced Setup
Configure Internal Organization Network
Configure Internal IP Settings
Name this Internal Organization Network
Ready to Complete

Select Typical or Advanced Setup

The default options are the most common setup for a new organization.

What type of network access do you want to give this organization?

● Typical
The quickest and most common way to set up networks for an organization.

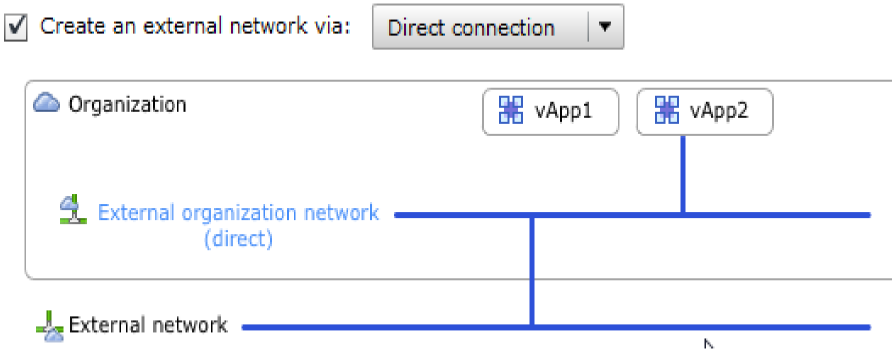
☑ Create an internal network

□ Create an external network via: Routed connection

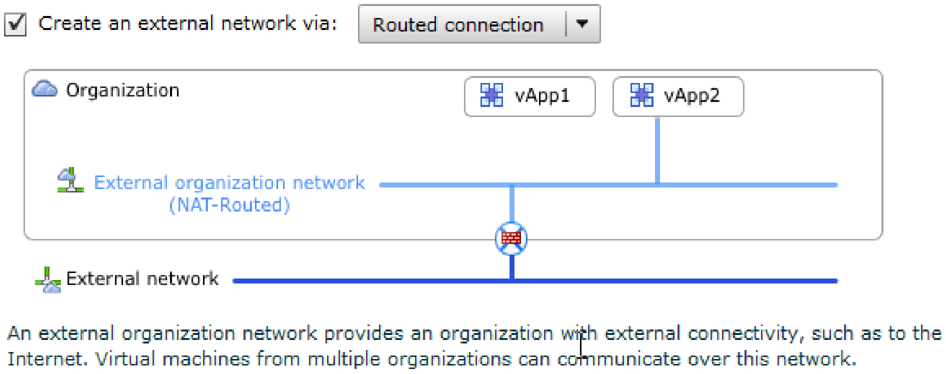
Organization
vApp1
vApp2
Internal organization network

Organization ‘external’ networks

- A network defined for an organization in VCD, it is created from ‘external networks’ (regular port-groups) and might use ‘internal’ Network Pools.
- In any case it is deployed on multiple ESX hosts using vDS or Nexus1000V, and frames are sent between ESX hosts for any type of ‘external network’.
- It is used by VCD to define a communication path from VMs to external users/hosts, like the internet or a private IP network ‘outside’ the ‘cloud’.
- It can be defined as ‘direct connection’ or ‘routed connection’.



☑ Create an external network via: Direct connection



☑ Create an external network via: Routed connection

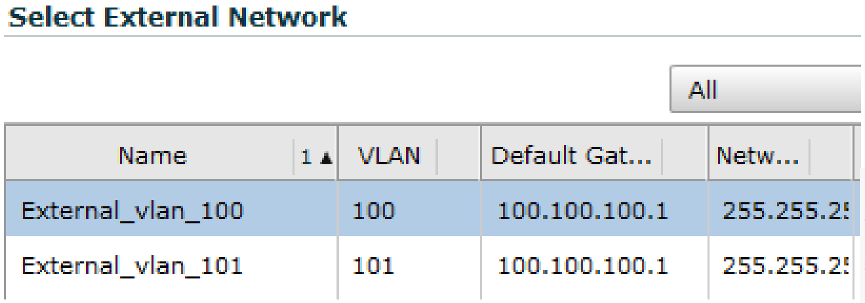
An external organization network provides an organization with external connectivity, such as to the Internet. Virtual machines from multiple organizations can communicate over this network.

- VSE is at the heart of external networks – it can be a GW to ‘routed’(NATed) networks.
- Same Network Pool (either VCDNI,PG or VLAN backed) can be used to create organization internal and organization external networks !

Organization ‘external’ network type: direct

- You attach it to regular port-group/vlan (external network), used by ORG VMs.

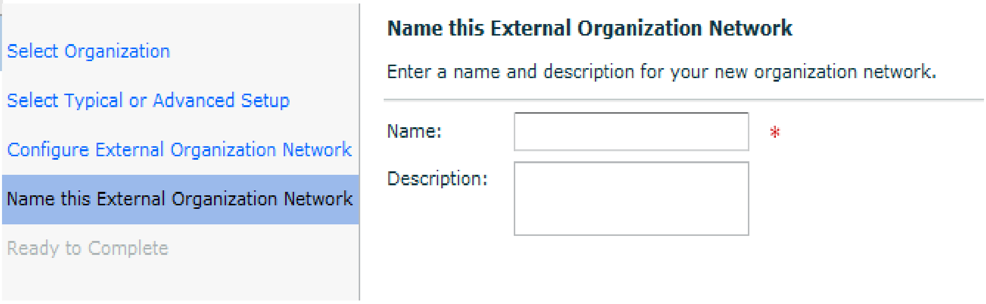
  It is a port-group with a vlan, like ‘internal pg-backed’ or ‘internal vlan-backed’



Select External Network

All

Name 1 ▲	VLAN	Default Gat...	Netw...
External_vlan_100	100	100.100.100.1	255.255.25
External_vlan_101	101	100.100.100.1	255.255.25



Select Organization
Select Typical or Advanced Setup
Configure External Organization Network
Name this External Organization Network
Ready to Complete

Name this External Organization Network

Enter a name and description for your new organization network.

Name: *

Description:

It needs a pre-defined port-group with a ‘regular’ vlan on vDS in Vcenter.

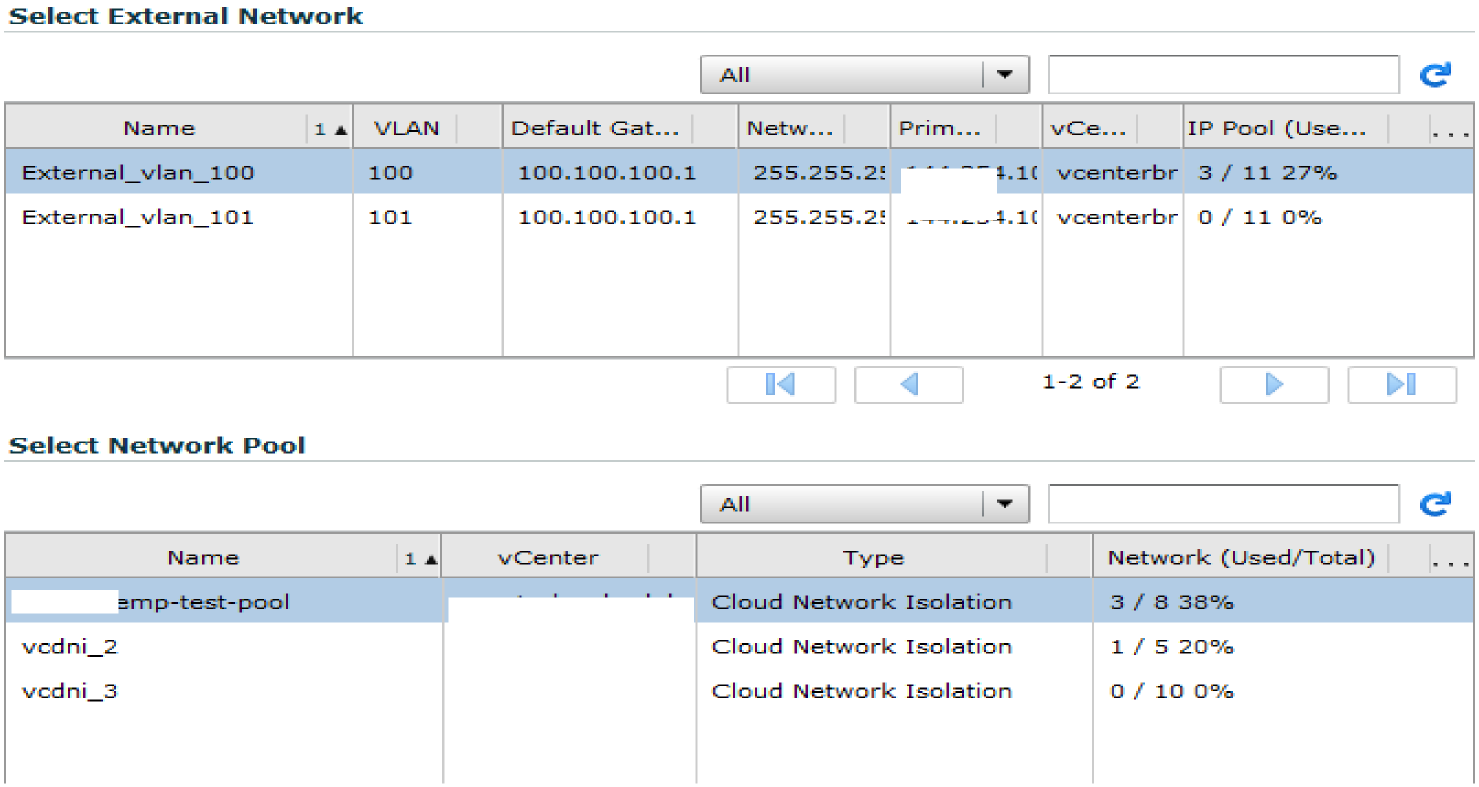
No need for ANY VSE in this case, VMs can attach directly to this port-group and use a vlan to communicate to outside world, using external router.

A VM can have multiple vnic interfaces attached to ‘external’ and ‘internal’ networks, in any case they are sent outside ESX and you might start to use static routes on the VM OS level if one attached to many.

Easiest way is to have an ORG VM connected to a single network.

Organization ‘external’ network type: routed (1)

- It is actually only NATed (name misleading) - You create it using both external pre-define port-group/vlan and network pool for the VMs to connect to.



Select External Network

All ▾

Name 1 ▲	VLAN	Default Gat...	Netw...	Prim...	vCe...	IP Pool (Use...	...
External_vlan_100	100	100.100.100.1	255.255.2!	[redacted]	vcenterbr	3 / 11 27%	
External_vlan_101	101	100.100.100.1	255.255.2!	[redacted]	vcenterbr	0 / 11 0%	

1-2 of 2

Select Network Pool

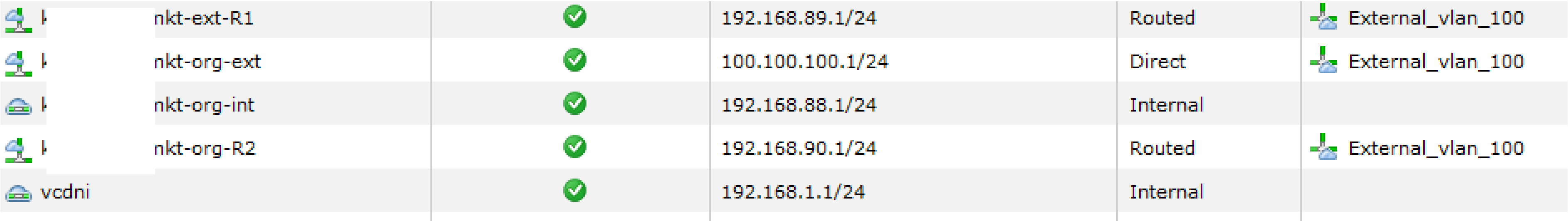
All ▾

Name 1 ▲	vCenter	Type	Network (Used/Total)	...
[redacted]emp-test-pool	[redacted]	Cloud Network Isolation	3 / 8 38%	
vcdni_2		Cloud Network Isolation	1 / 5 20%	
vcdni_3		Cloud Network Isolation	0 / 10 0%	

VSE VM device is created in this case and deployed as a gateway of the VMs on it’s internal vNIC (network pool) and also connected with it’s external vNIC to external port-group/vlan (external network) per the definition.

Organization ‘external’ network type: routed (2)

- A single VSE VM instance is created for every external-routed port-group. Multiple VSE instances might be needed for many networks of the same organization, that needs external communications.



k[redacted]nkt-ext-R1	✓	192.168.89.1/24	Routed	External_vlan_100
k[redacted]nkt-org-ext	✓	100.100.100.1/24	Direct	External_vlan_100
k[redacted]nkt-org-int	✓	192.168.88.1/24	Internal	
k[redacted]nkt-org-R2	✓	192.168.90.1/24	Routed	External_vlan_100
vcdni	✓	192.168.1.1/24	Internal	

This is true even if 2 external networks are using the same external port-group and same vlan, no routing available only NAT is deployed on each VSE.

Traffic between different ‘routed’ external networks of the same organization needs to pass through at least 2 VSE devices with corresponding end-to-end NAT rules and FW rules - if possible per application needs.

It is hard to predict the packet flows in those cases, it is difficult to find which VSE controls which networks and on which ESX host they are deployed at a specific timeframe (VSE naming convention is unclear).

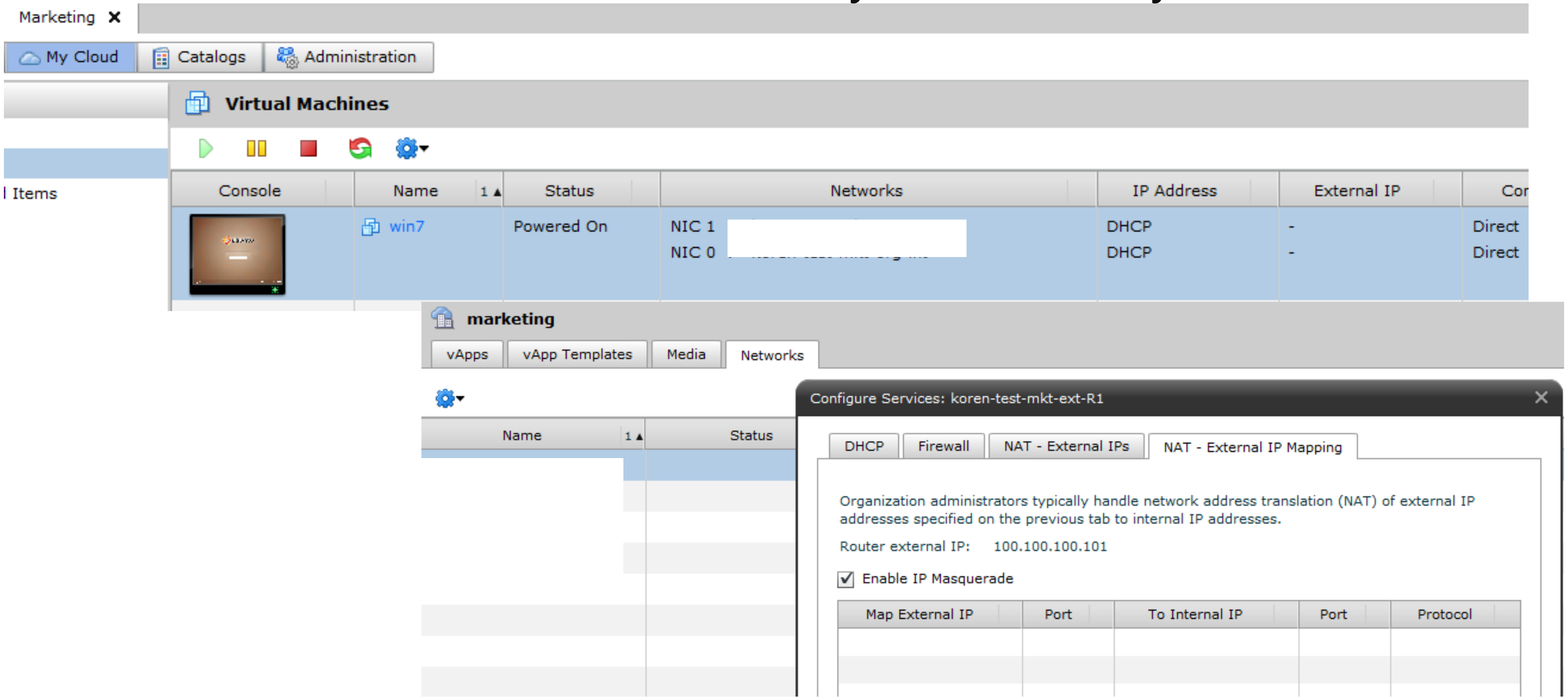
VSE introduces duplicate BW per network per ORG as packets needs to be sent to the VSE ESX host prior to sending finally to external destination.

If connectivity breaks between VMs and their VSE / GW, the VMs are cut off.

VSE device is currently not capable of stateful failover functionality.

VCD Networking as seen by an organization admin

- All you need to do now is attach a specific VM to some networks that you have created, VM can attach to many networks if you want:



Console	Name	Status	Networks	IP Address	External IP	Cor
	win7	Powered On	NIC 1 [redacted] NIC 0	DHCP DHCP	- -	Direct Direct

Name	Status

Configure Services: koren-test-mkt-ext-R1

DHCP | Firewall | NAT - External IPs | NAT - External IP Mapping

Organization administrators typically handle network address translation (NAT) of external IP addresses specified on the previous tab to internal IP addresses.

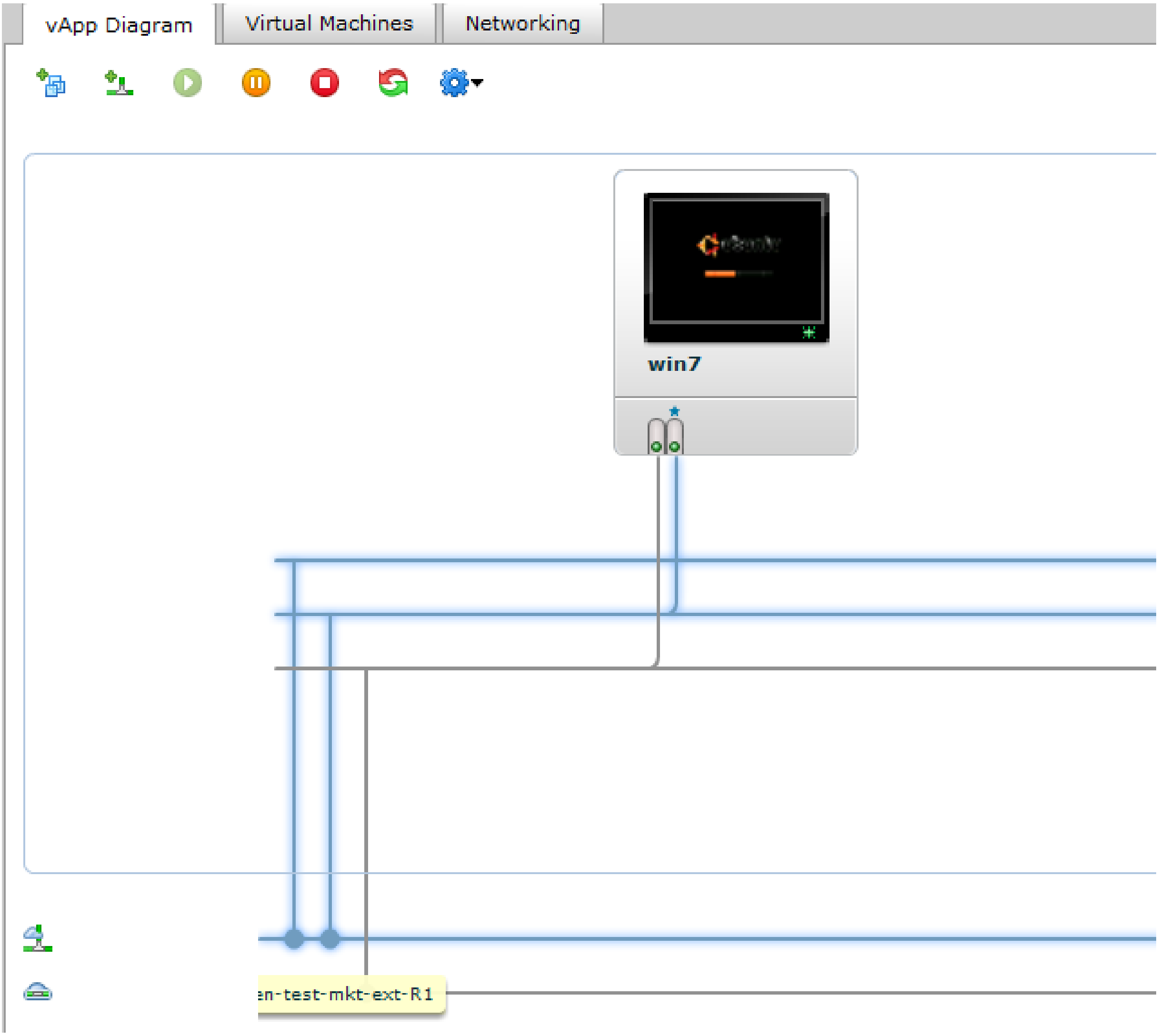
Router external IP: 100.100.100.101

☑ Enable IP Masquerade

Map External IP	Port	To Internal IP	Port	Protocol

- NAT and other services can be managed by organization admin.
- The underlying network (3 types of internal isolated, 2 types of external) will determine the IP scheme, L2 path and IP path a VM frame will take until it reaches it’s final destination outside the cloud, this most probably will be several L2 ‘hops’ and L3 ‘hops’ in different ESX hosts controlled by VCD.

Nice GUI illustration also currently NOT revealing underlying connectivity details – L2 and L3 paths.



A simple use-case for organization network on VCD

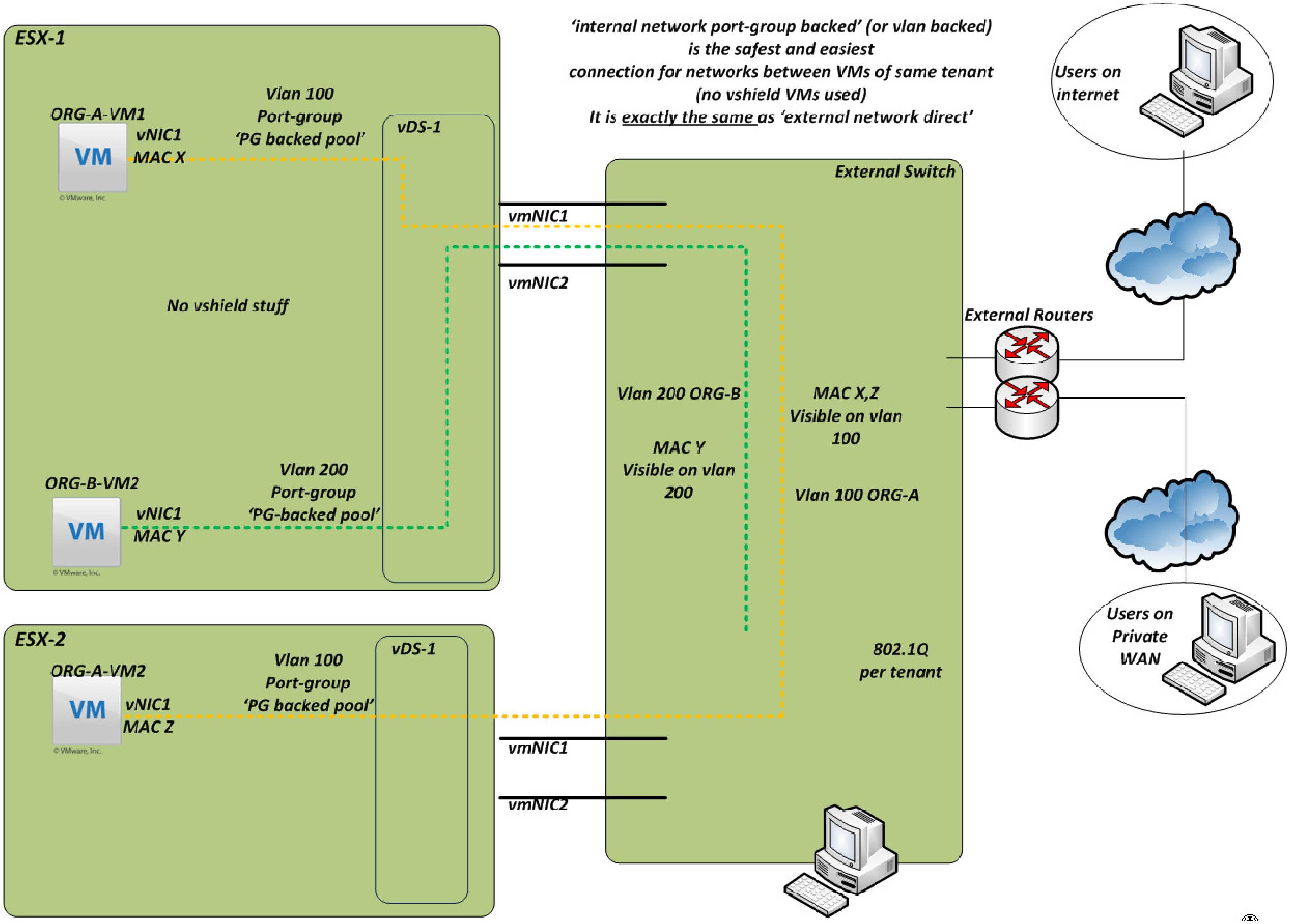
- An organization DC is built with 3 internal networks/vlans (let’s keep it as simple as possible) each network have multiple VMs, 2 ESX hosts used.
- 2 networks/vlans needs access to internet through a FW/router.
- Those 2 networks needs communication between each other through a FW.
- An internal isolated networks are needed for some hosts on 3rd vlan.
- Another organization DC is bulit with 1 networks …

VCD solution:

2 external-routed networks, 3 vshield-edge devices , 5 vlans/port-groups and 2 vshield-PGI device for the VCDNI isolated network pool.

Let’s see how it might be built with VCD and what might be the packet flow…..

ORG Internal network: port-group backed / vlan backed



'internal network port-group backed' (or vlan backed)
is the safest and easiest
connection for networks between VMs of same tenant
(no vshield VMs used)
It is exactly the same as 'external network direct'

ESX-1

ORG-A-VM1

VM

vNIC1
MAC X

Vlan 100
Port-group
'PG backed pool'

vDS-1

No vshield stuff

ORG-B-VM2

VM

vNIC1
MAC Y

Vlan 200
Port-group
'PG-backed pool'

vmNIC1

vmNIC2

ESX-2

ORG-A-VM2

VM

vNIC1
MAC Z

Vlan 100
Port-group
'PG backed pool'

vDS-1

vmNIC1

vmNIC2

External Switch

Vlan 200 ORG-B

MAC Y
Visible on vlan
200

MAC X,Z
Visible on vlan
100

Vlan 100 ORG-A

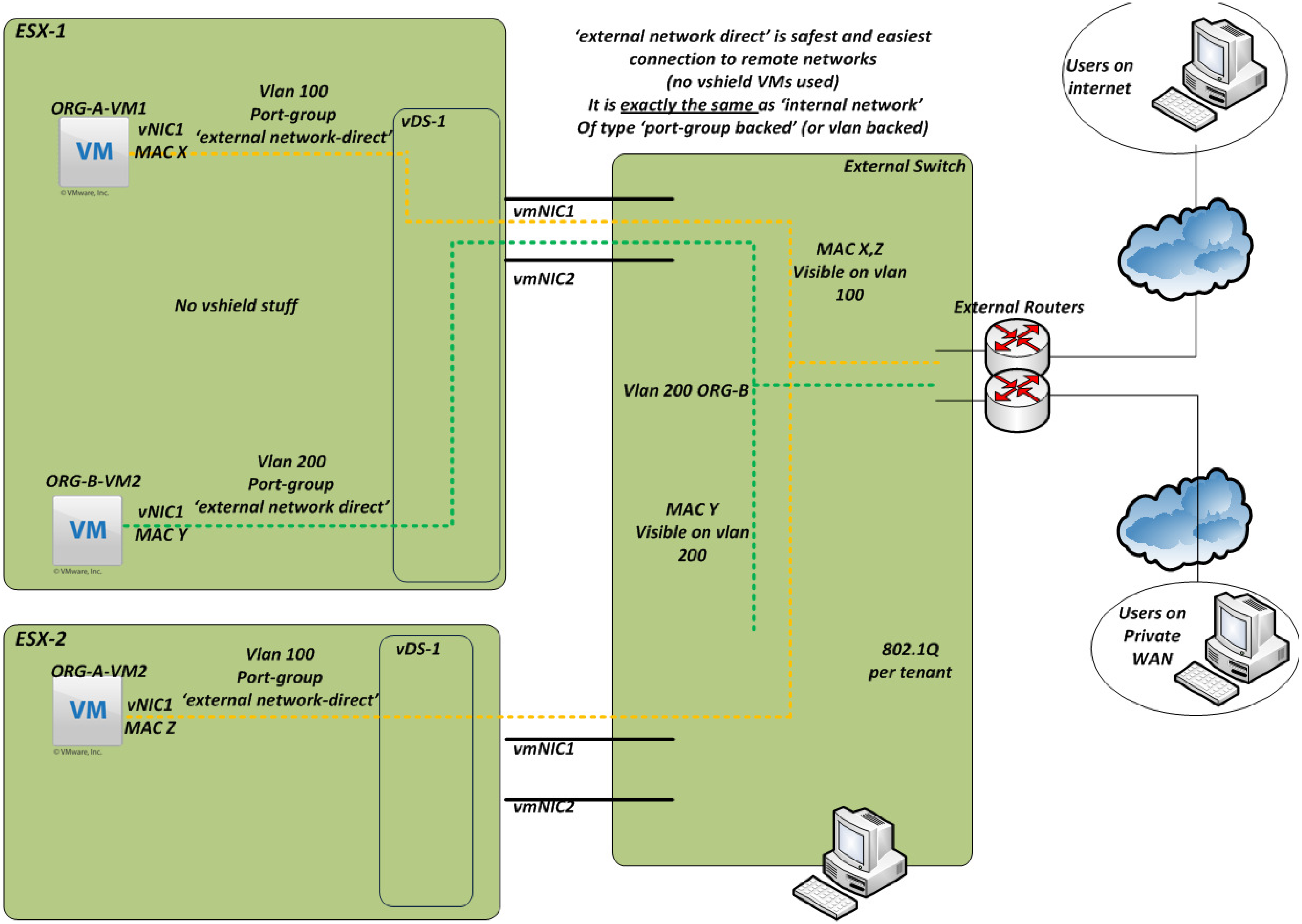
802.1Q
per tenant

External Routers

Users on
internet

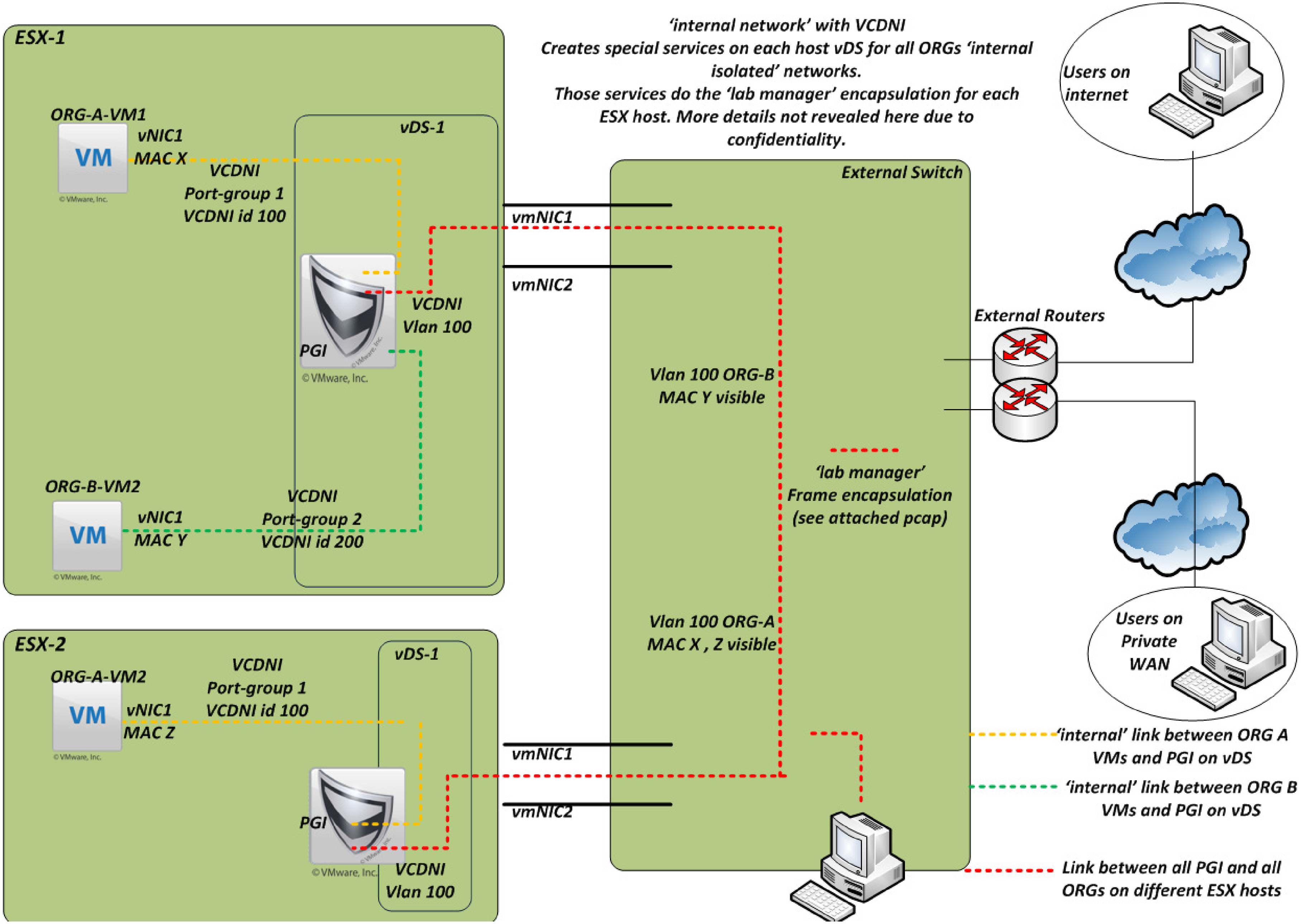
Users on
Private
WAN

ORG External network: direct



'external network direct' is safest and easiest connection to remote networks (no vshield VMs used)
It is exactly the same as 'internal network'
Of type 'port-group backed' (or vlan backed)

‘Internal network’: VCDNI backed



‘internal network’ with VCDNI
Creates special services on each host vDS for all ORGs ‘internal isolated’ networks.
Those services do the ‘lab manager’ encapsulation for each ESX host. More details not revealed here due to confidentiality.

ESX-1

ORG-A-VM1

VM

vNIC1
MAC X

VCDNI
Port-group 1
VCDNI id 100

vDS-1

PGI

VCDNI
Vlan 100

ORG-B-VM2

VM

vNIC1
MAC Y

VCDNI
Port-group 2
VCDNI id 200

ESX-2

ORG-A-VM2

VM

vNIC1
MAC Z

VCDNI
Port-group 1
VCDNI id 100

vDS-1

PGI

VCDNI
Vlan 100

vmNIC1

vmNIC2

vmNIC1

vmNIC2

External Switch

Vlan 100 ORG-B
MAC Y visible

‘lab manager’
Frame encapsulation
(see attached pcap)

Vlan 100 ORG-A
MAC X , Z visible

Users on internet

External Routers

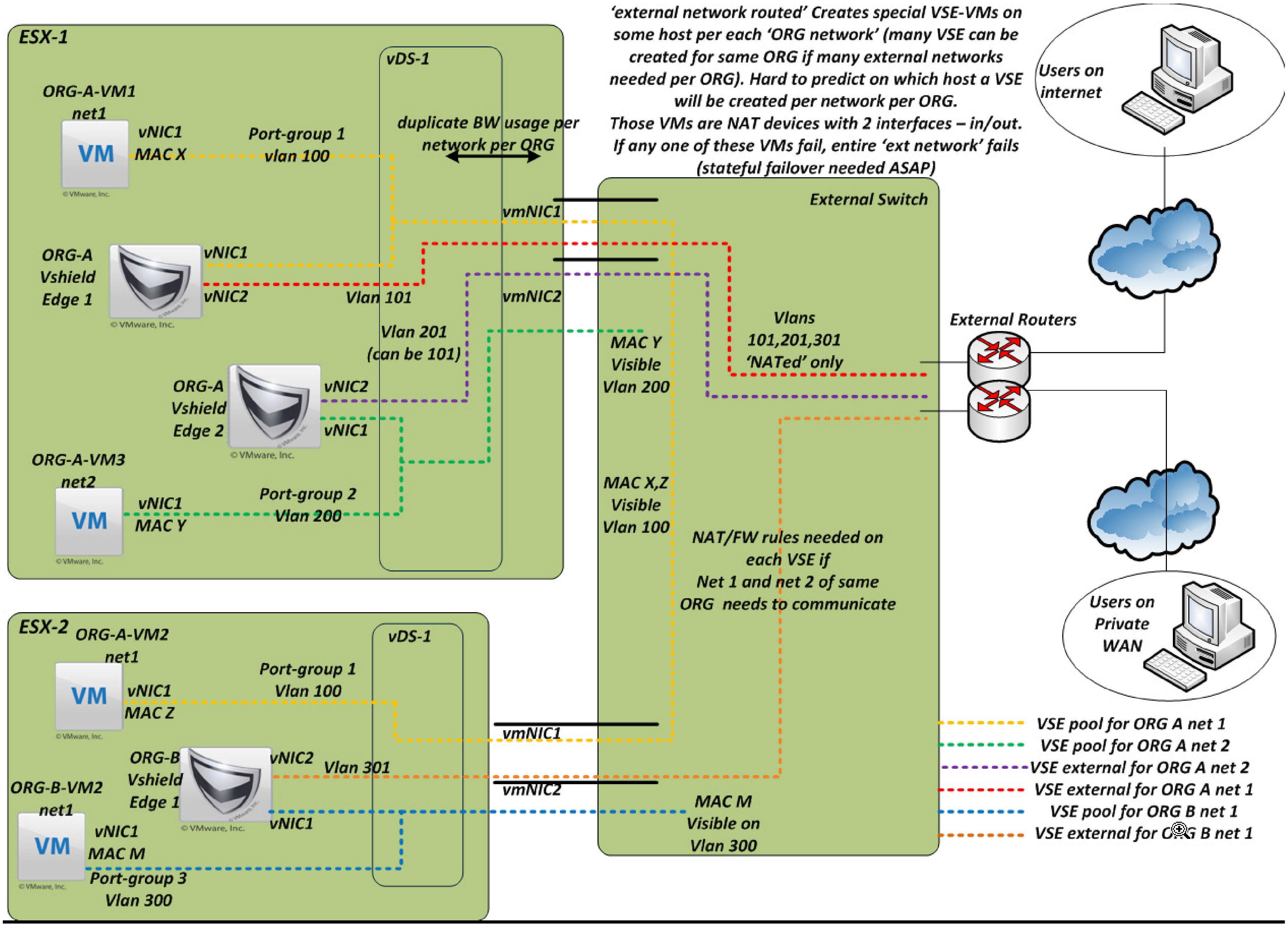
Users on Private WAN

‘internal’ link between ORG A VMs and PGI on vDS

‘internal’ link between ORG B VMs and PGI on vDS

Link between all PGI and all ORGs on different ESX hosts

external network: ‘routed’, ‘PG-backed’ pool inside



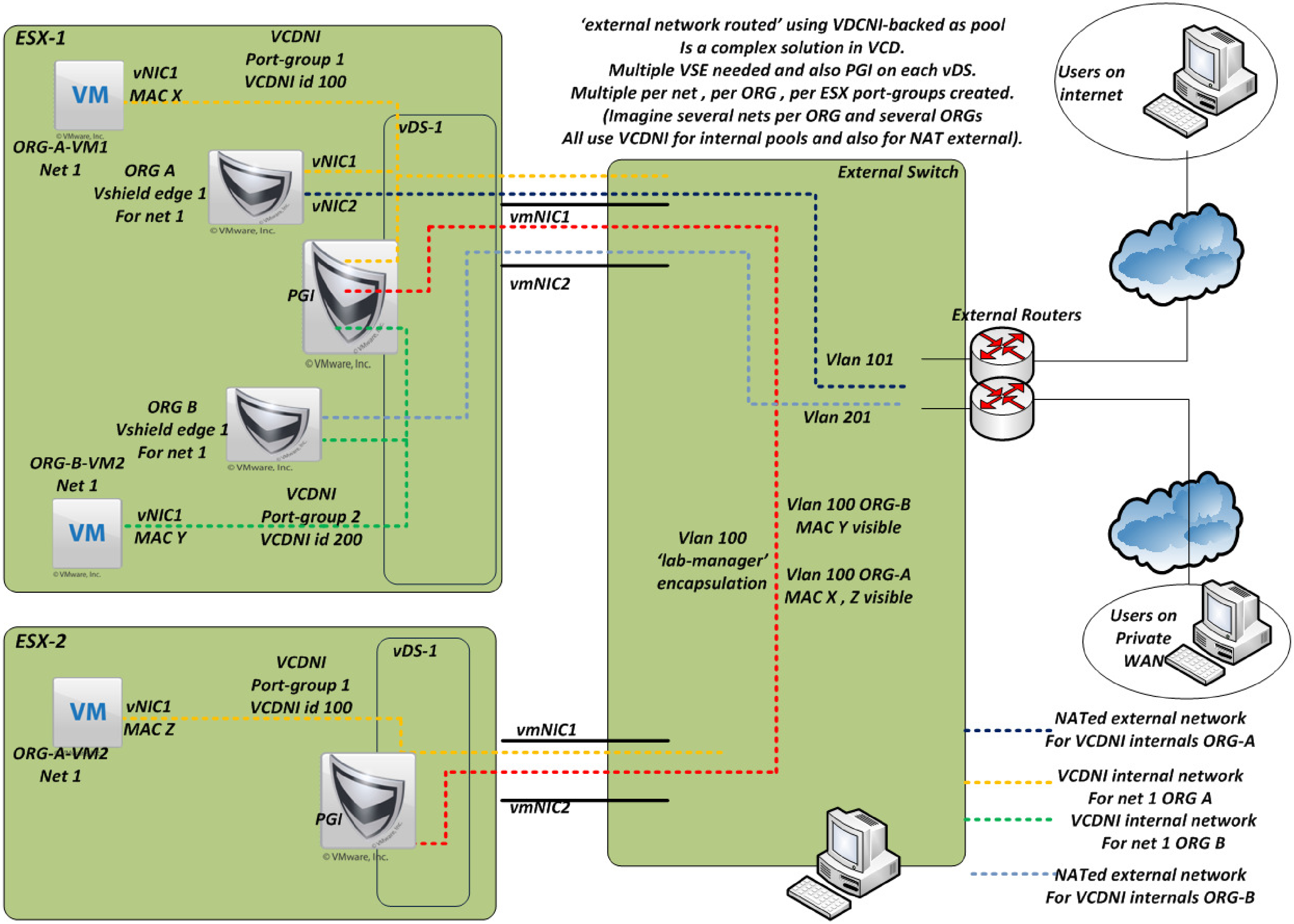
‘external network routed’ Creates special VSE-VMs on some host per each ‘ORG network’ (many VSE can be created for same ORG if many external networks needed per ORG). Hard to predict on which host a VSE will be created per network per ORG.

Those VMs are NAT devices with 2 interfaces – in/out.

If any one of these VMs fail, entire ‘ext network’ fails (stateful failover needed ASAP)

NAT/FW rules needed on each VSE if Net 1 and net 2 of same ORG needs to communicate

external network: ‘routed’, ‘VCDNI backed’ pool inside



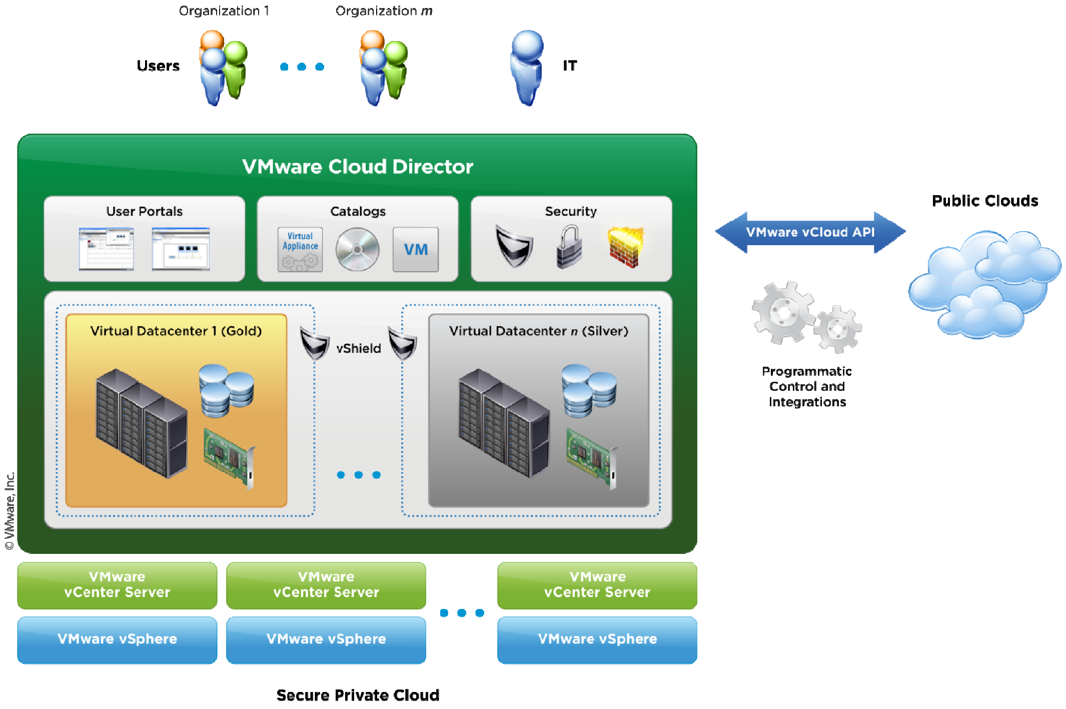
‘external network routed’ using VDCNI-backed as pool
Is a complex solution in VCD.
Multiple VSE needed and also PGI on each vDS.
Multiple per net , per ORG , per ESX port-groups created.
(Imagine several nets per ORG and several ORGs
All use VCDNI for internal pools and also for NAT external).

NATed external network
For VCDNI internals ORG-A

VCDNI internal network
For net 1 ORG A

VCDNI internal network
For net 1 ORG B

NATed external network
For VCDNI internals ORG-B



Vcloud Director Networking

Key Takeaway :

‘Explore end-to-end packet L2 and L3 flows for your ORG’

1. Predict it for the BW usage per ESX for SLA and QOS.
2. Predict it for understanding security impacts.
3. Predict it for understanding effects of in-line NAT and FW.
4. predict it for failure analysis and troubleshooting.