

SBCCon networking refresh 2018



Agenda

- Part 1 – Networking basics
 - Part 2 – CloudStack networking (bootcamp recap)
 - Part 3 – System VM networking troubleshooting
-
- *We have limited time and a lot of material to go through.*
 - *Please give feedback as we go along about what is useful and what we can skip through quicker!*
 - *Please ask questions! If a topic is too big to digest in one sitting we'll arrange followup sessions.....*



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Part 1 – networking basics



OSI (Open Systems Interconnect) model

- International model / standard for telecommunications since the 1970's.
- From a CloudStack code point of view we are mainly interested in layer 2 (VLANs and mac addresses) and layer 3 (IP).
- From a design and project delivery point of view layer 1/2/3 are all important.

Layer	Application/Example	Central Device/Protocols	DOD4 Model
Application (7) Serves as the window for users and application processes to access the network services.	End User layer Program that opens what was sent or creates what is to be sent Resource sharing • Remote file access • Remote printer access • Directory services • Network management	User Applications SMTP	Process
Presentation (6) Formats the data to be presented to the Application layer. It can be viewed as the "Translator" for the network.	Syntax layer encrypt & decrypt (if needed) Character code translation • Data conversion • Data compression • Data encryption • Character Set Translation	JPEG/ASCII EBDIC/TIFF/GIF PICT	
Session (5) Allows session establishment between processes running on different stations.	Synch & send to ports (logical ports) Session establishment, maintenance and termination • Session support - perform security, name recognition, logging, etc.	Logical Ports RPC/SQ/LNFS NetBIOS names	GATEWAY
Transport (4) Ensures that messages are delivered error-free, in sequence, and with no losses or duplications.	TCP Host to Host, Flow Control Message segmentation • Message acknowledgement • Message traffic control • Session multiplexing	F P A C K E T T R A C K E R I N G TCP/SPX/UDP	
Network (3) Controls the operations of the subnet, deciding which physical path the data takes.	Packets ("letter", contains IP address) Routing • Subnet traffic control • Frame fragmentation • Logical-physical address mapping • Subnet usage accounting	Routers IP/IPX/ICMP	Host to Host Internet Can be used on all layers
Data Link (2) Provides error-free transfer of data frames from one node to another over the Physical layer.	Frames ("envelopes", contains MAC address) [NIC card — Switch — NIC card] (end to end) Establishes & terminates the logical link between nodes • Frame traffic control • Frame sequencing • Frame acknowledgment • Frame delimiting • Frame error checking • Media access control	Switch Bridge WAP PPP/SLIP	
Physical (1) Concerned with the transmission and reception of the unstructured raw bit stream over the physical medium.	Physical structure Cables, hubs, etc. Data Encoding • Physical medium attachment • Transmission technique - Baseband or Broadband • Physical medium transmission Bits & Volts	Hub	Network



Layer 1 – physical connectivity

- **Connectivity speeds depend on physical cabling:**
 - 100Mbps - 1Gbps: CAT5 or CAT6 (copper)
 - 1Gbps - 10Gbps: CAT5 or CAT6 / SFP+ (copper or fiber)
 - 10Gbps - 40Gbps: SFP+ / QSFP (copper or fiber)



UNIXD



Layer 2/3 – devices

- **From a CloudStack point of view we are basing all designs around data center distributed switching:**
 - Top-of-rack switches: typically 1Gbps – 10Gbps, handles traffic to / from hypervisor hosts
 - Data center core switches: typically 10Gbps, handles traffic to/ from ToR switches and uplinks to external networks
 - Firewalls: controls ingress / egress traffic from infrastructure / core switches

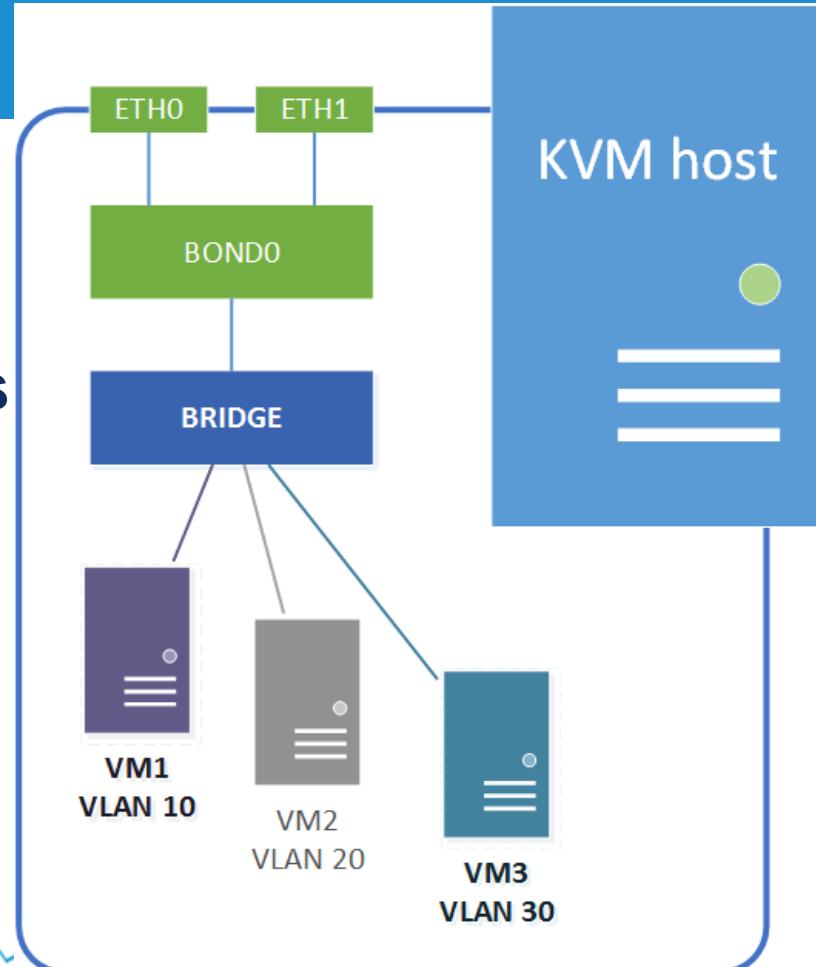


Layer 2/3 - devices

Device	OSI layer	Details	CloudStack
Network switch	2 – data link	Handles traffic based on MAC addresses	ToR and DC core
Router	3 - network	Handles IP routing between networks and interfaces	DC core Virtual Router
Firewalls	3 - network	Handles IP filtering across an interface	DC core Virtual Router
Bridge	(1 – physical) 2 – data link	Connect two separate networks as if they were a single network, often transparently to the traffic flowing through the bridge.	Hypervisor Bridge names used for traffic mapping in CloudStack

Bonds and teams

- **NIC bonds / teams are used for resilience**
- **They will combine the connectivity of two or more NICs to provide:**
 - network failover capability
 - combined bandwidth
- **Note: teams are relatively new (RHEL7 / CentOS7)**



Bonds vs teams

Feature	Bonding	Team
broadcast Tx policy	Yes	Yes
round-robin Tx policy	Yes	Yes
active-backup Tx policy	Yes	Yes
LACP (802.3ad) support	Yes (active only)	Yes
Hash-based Tx policy	Yes	Yes
User can set hash function	No	Yes
Tx load-balancing support (TLB)	Yes	Yes
LACP hash port select	Yes	Yes
load-balancing for LACP support	No	Yes
Ethtool link monitoring	Yes	Yes
ARP link monitoring	Yes	Yes
NS/NA (IPv6) link monitoring	No	Yes
ports up/down delays	Yes	Yes
port priorities and stickiness ("primary" option enhancement)	No	Yes

Feature	Bonding	Team
separate per-port link monitoring setup	No	Yes
multiple link monitoring setup	Limited	Yes
lockless Tx/Rx path	No (rwlock)	Yes (RCU)
VLAN support	Yes	Yes
user-space runtime control	Limited	Full
Logic in user-space	No	Yes
Extensibility	Hard	Easy
Modular design	No	Yes
Performance overhead	Low	Very Low
D-Bus interface	No	Yes
multiple device stacking	Yes	Yes
zero config using LLDP	No	(in planning)
NetworkManager support	Yes	Yes

Bond modes

Bond mode / hypervisor	Details
1 - active-backup (Xen/KVM/VMware)	Active-backup policy: Only one slave in the bond is active. A different slave becomes active if, and only if, the active slave fails. The bond's MAC address is externally visible on only one port (network adapter) to avoid confusing the switch. This mode provides fault tolerance. The primary option affects the behavior of this mode.
2 - balance-xor (KVM)	XOR policy: Transmit based on [(source MAC address XOR'd with destination MAC address) modulo slave count]. This selects the same slave for each destination MAC address. This mode provides load balancing and fault tolerance.
3 – broadcast (KVM)	Broadcast policy: transmits everything on all slave interfaces. This mode provides fault tolerance.
4 - 802.3ad (Xen/KVM/VMware)	IEEE 802.3ad Dynamic link aggregation (LACP / LAG)

Bond modes

Bond mode / hypervisor	Details
5 - balance-tlb (KVM)	Adaptive transmit load balancing: channel bonding that does not require any special switch support. The outgoing traffic is distributed according to the current load (computed relative to the speed) on each slave. Incoming traffic is received by the current slave. If the receiving slave fails, another slave takes over the MAC address of the failed receiving slave.
6 - balance-alb (KVM/Xen/VMware)	Adaptive load balancing: includes balance-tlb plus receive load balancing (rlb) for IPV4 traffic, and does not require any special switch support. The receive load balancing is achieved by ARP negotiation. The bonding driver intercepts the ARP Replies sent by the local system on their way out and overwrites the source hardware address with the unique hardware address of one of the slaves in the bond such that different peers use different hardware addresses for the server.

Bond modes

Bond mode / hypervisor	Details
7 – SLB (Xen)	<p>SLB is an active/active mode, but only supports load balancing of virtual machine traffic across the physical NICs. Provides fail-over support for all other traffic modes. Does not require switch support for Etherchannel or 802.3ad (LACP). Load balances traffic between multiple interfaces at virtual machine granularity by sending traffic through different interfaces based on the source MAC address of the packet. Is derived from the open source ALB mode and reuses the ALB capability to dynamically re-balance load across interfaces.</p>



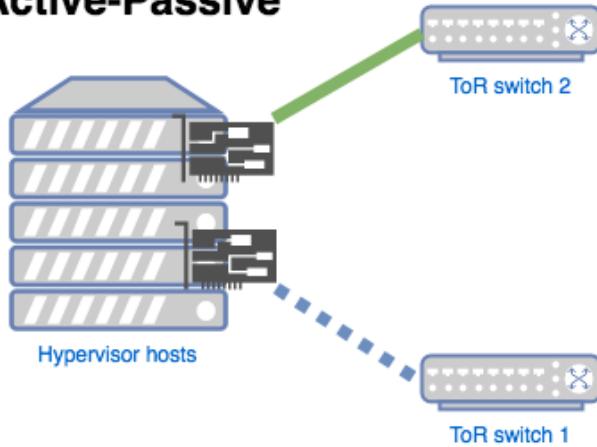
The Cloud Specialists

ShapeBlue.com

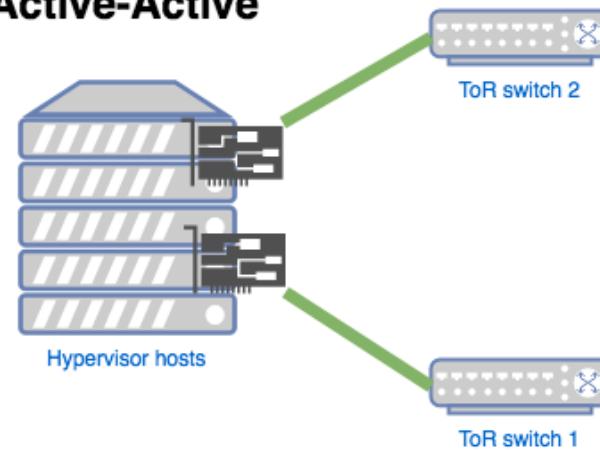
 @ShapeBlue

Bond modes

Active-Passive

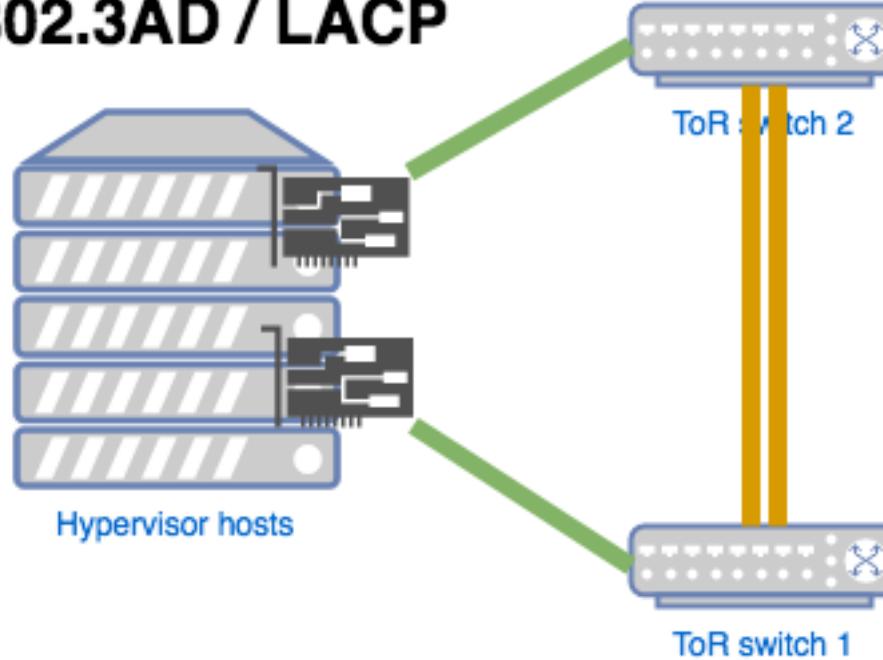


Active-Active



Bond modes

802.3AD / LACP



NICs / bonds / teams in KVM CloudStack agent

- KVM agent looks for underlying device name patterns
- <https://github.com/apache/cloudstack/blob/master/plugins/hypervisors/kvm/src/main/java/com/cloud/hypervisor/kvm/resource/LibvirtComputingResource.java#L1294>

```
1293
1294     String [] _ifNamePatterns = {
1295         "^eth",
1296         "^bond",
1297         "^vlan",
1298         "^vx",
1299         "^em",
1300         "^ens",
1301         "^eno",
1302         "^enp",
1303         "^team",
1304         "^enx",
1305         "^p\\d+p\\d+"
1306     };
1307 }
```

Networking basics

- **MAC addresses:**

```
[root@csman tmp]# ip addr show eth2
4: eth2: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc pfifo_fast state UP qlen 1000
    link/ether 08:00:27:e0:39:59 brd ff:ff:ff:ff:ff:ff
        inet 192.168.56.11/24 brd 192.168.56.255 scope global eth2
            inet6 fe80::a00:27ff:fee0:3959/64 scope link
                valid_lft forever preferred_lft forever
```

- OSI layer 2
- Map physical NIC to an IP address
- Must be unique
- First 3 octets tells which vendor (OUI)
- Tracked in ARP tables
- MAC addresses never “leave” the local subnet.



The Cloud Specialists

ShapeBlue.com

 @ShapeBlue

Networking basics

- ARP tables keep the MAC to IP cache:

```
[root@tr1-1330-k-cs45-dsonstebo-kvm1 ~]#  
[root@tr1-1330-k-cs45-dsonstebo-kvm1 ~]# arp  
Address          HWtype  HWaddress      Flags Mask   Iface  
10.2.0.16        ether    00:50:56:8e:ad:d8  C       cloudbr0  
tr1-1330-k-cs45-dsonste  ether    06:15:4e:01:07:2a  C       cloudbr0  
10.2.254.254     ether    5c:26:0a:c6:aa:55  C       cloudbr0  
10.2.0.4         ether    00:50:56:8e:70:43  C       cloudbr0  
10.2.2.2         ether    06:cf:3a:01:06:fa  C       cloudbr0  
[root@tr1-1330-k-cs45-dsonstebo-kvm1 ~]
```

- We will cover routing later but.....
 - ARP will allow you assign an IP address to an un-addressed device on the local subnet.
 - This is useful if you just needs to connect to a new IP device and you can't log on to it or get it to pick up a DHCP address

• GARP

- A Gratuitous ARP is a handshake sent from a device to let other network devices know a MAC address has moved



The Cloud Specialists

ShapeBlue.com

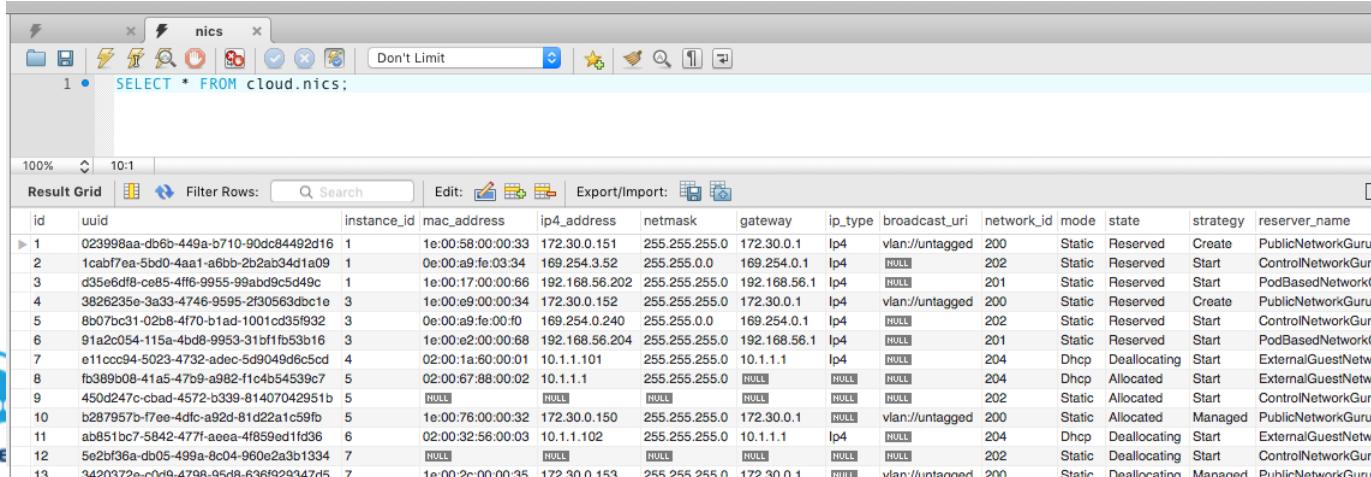
@ShapeBlue

Networking basics

- **MAC addresses and CloudStack:**

- In ACS we generate the MAC addresses for VMs, and as a result the CloudStack code is responsible for uniqueness
- We recently found flaws in the logic around this in environments with a large number of VMs - where we found duplicate MAC addresses used. This leads to issues as you end up with conflict between multiple MAC to IP mappings (fixed by Daan?).

- **Can be found in the "nics" table:**



The screenshot shows a MySQL Workbench interface with the following details:

- Toolbar: Includes standard icons for file operations, search, and database navigation.
- Query Editor: Shows the SQL command: `SELECT * FROM cloud.nics;`
- Result Grid: Displays the data from the 'nics' table. The columns are: id, uuid, instance_id, mac_address, ip4_address, netmask, gateway, ip_type, broadcast_uri, network_id, mode, state, strategy, and reserver_name.
- Data Rows (approximate values):

id	uuid	instance_id	mac_address	ip4_address	netmask	gateway	ip_type	broadcast_uri	network_id	mode	state	strategy	reserver_name
1	023998aa-db6b-449a-b710-90dc84492d16	1	1e:00:58:00:00:33	172.30.0.151	255.255.255.0	172.30.0.1	lp4	vlan://untagged	200	Static	Reserved	Create	PublicNetworkGuru
2	1cabf7ea-5bd0-4aa1-a6bb-2b2ab34d1a09	1	0e:00:a9:fe:03:34	169.254.3.52	255.255.0.0	169.254.0.1	lp4	NULL	202	Static	Reserved	Start	ControlNetworkGuru
3	d35e6df8-ce85-4ff6-9955-99abd9c5d49c	1	1e:00:17:00:00:66	192.168.56.202	255.255.255.0	192.168.56.1	lp4	NULL	201	Static	Reserved	Start	PodBasedNetworkGuru
4	3826235e-3a33-4746-9595-2f30563dbc1e	3	1e:00:e9:00:00:34	172.30.0.152	255.255.255.0	172.30.0.1	lp4	vlan://untagged	200	Static	Reserved	Create	PublicNetworkGuru
5	b028bc31-47b0-b1ad-1001c035932	3	0e:00:a9:fe:00:00	169.254.0.240	255.255.0.0	169.254.0.1	lp4	NULL	202	Static	Reserved	Start	ControlNetworkGuru
6	91a2c054-115a-4bd8-9953-31b1fb53b16	3	1e:00:e2:00:00:68	192.168.56.204	255.255.255.0	192.168.56.1	lp4	NULL	201	Static	Reserved	Start	PodBasedNetworkGuru
7	e11ccc94-5023-4732-adec-5d9049d6c5cd	4	02:00:01:a6:00:01	10.1.1.101	255.255.255.0	10.1.1.1	lp4	NULL	204	Dhcp	Deallocating	Start	ExternalGuestNetwork
8	fb389b08-41a9-47b9-a982-fc4b54539c7	5	02:00:67:88:00:02	10.1.1.1	255.255.255.0	NULL	NULL	NULL	204	Dhcp	Allocated	Start	ExternalGuestNetwork
9	450d247c-cbad-4572-b339-81407042951b	5	NULL	NULL	NULL	NULL	NULL	NULL	202	Static	Allocated	Start	ControlNetworkGuru
10	b287957b-f7ee-4fdc-a92d-81d22a1c59fb	5	1e:00:76:00:00:32	172.30.0.150	255.255.255.0	172.30.0.1	lp4	vlan://untagged	200	Static	Allocated	Managed	PublicNetworkGuru
11	ab851b7c-5842-477f-aee0-4f859ed1fd36	6	02:00:32:56:00:03	10.1.1.102	255.255.255.0	10.1.1.1	lp4	NULL	204	Dhcp	Deallocating	Start	ExternalGuestNetwork
12	5e2bf36a-db05-49fa-8c04-960e2a3b1334	7	NULL	NULL	NULL	NULL	NULL	NULL	202	Static	Deallocating	Start	ControlNetworkGuru
13	3420372e-c0d9-4798-95d8-636f929347d5	7	1e:00:2c:00:00:35	172.30.0.153	255.255.255.0	172.30.0.1	lp4	vlan://untagged	200	Static	Deallocating	Managed	PublicNetworkGuru

- IP is at OSI layer 3.
- IPv4:
 - 32-bit address space
 - Decimal addressing: 10.0.2.15/24
 - Full support in CloudStack
- IPv6:
 - 128-bit address space
 - Hex addressing: fe80::a00:27ff:fe49:64bd/64
 - Only supported in basic zones

IPv4 private addressing – RFC1918

Subnet class	Largest CIDR block	IP addresses in range	Details
Class A	10.0.0.0/8 (255.0.0.0)	16,777,216 10.0.0.0 – 10.255.255.255	8 network bits 24 host bits
Class B	172.16.0.0/12 (255.240.0.0)	1,048,576 172.16.0.0 – 172.31.255.255	12 network bits 20 host bits
Class C	192.168.0.0/16 (255.255.0.0)	65,536 192.168.0.0 – 192.168.255.255	16 network bits 16 host bits

Everything outside these ranges are considered public.



The Cloud Specialists

ShapeBlue.com

 @ShapeBlue

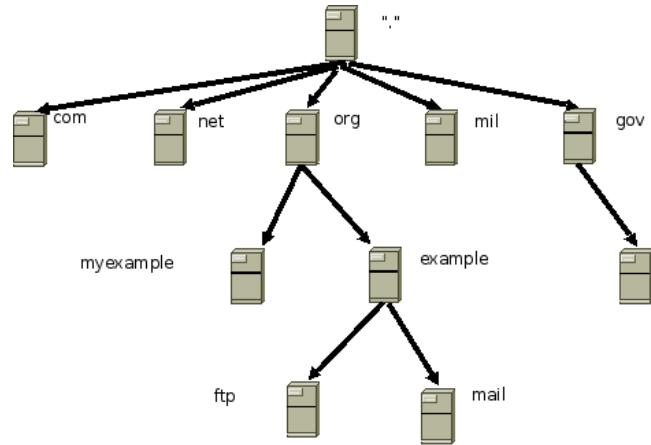
IPv4 subnetting

Decimal	Binary	
192.168.56.11	11000000.10101000.00111000.00001011	4 octets (8 bit number)
255.255.255.0	11111111.11111111.11111111.00000000	
AND Remainder	11000000.10101000.00111000.00000000 -----.-----.-----.00001011	Network is 192.168.56.0 Host is .11

- The first and last address in a range can not be assigned to a host:
 - The first address in a subnet range is the network identifier: **192.168.56.0**
 - The last address in a subnet range is the broadcast address: **192.168.56.255**
- You always have a default gateway on a network, this is typically the first or last address:
192.168.56.1 or .254
- A **.0** IP address is technically valid if subnet mask allows, but is not commonly used (and not used in CloudStack):
 - **192.168.56.0/23** means range **192.168.56.1 to 192.168.57.254**
 - Hence **192.168.57.0** is technically a valid address

DNS

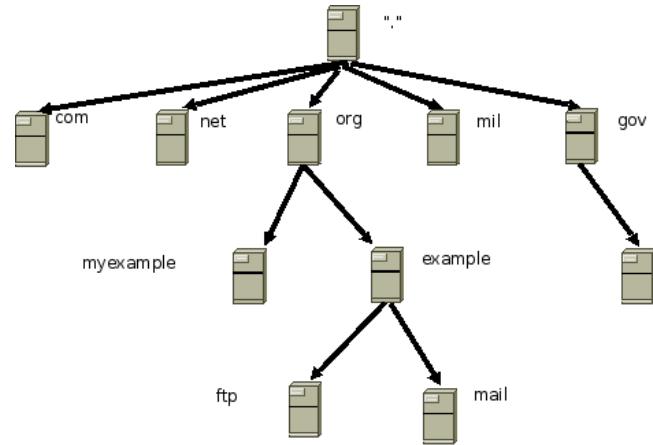
- **Domain Name System:**
 - OSI Layer7 / Application Layer
 - Handles IP to hostname resolution
 - Based on an hierarchical domain naming structure
 - The TTL – Time To Live – for any DNS record determines how long it can be cached at the client end:
 - If you update a DNS record with TTL=1hour, then it may take up to 1hour *per DNS hierarchical server level* before this new DNS record is used by a client.
 - If you use a DNS server twice removed from the DNS SOA (Start Of Authority) server then it can take 3 hours before the cache chain is fully refreshed.



The Domain Name System

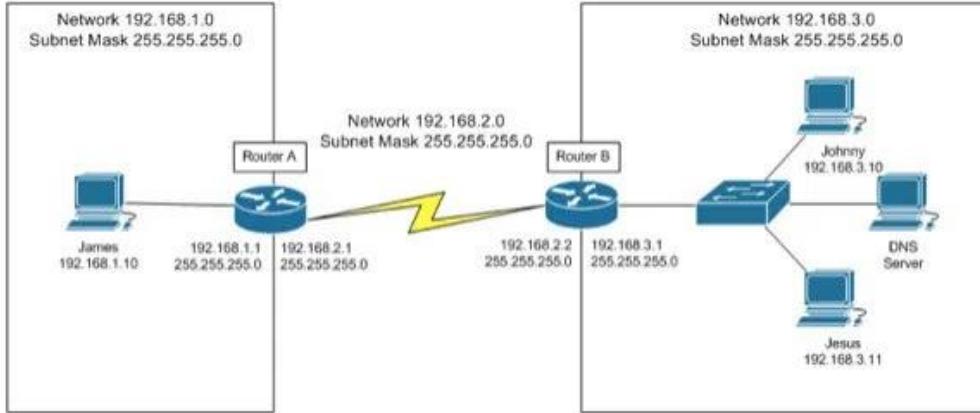
DNS

- **Domain Name System:**
 - Service records can be useful for failover purposes:
 - "clouddbsrv.sblab.local" could point to the master MySQL server IP address, but could be changed to the slave in failover scenario.
 - The FQDN "clouddbsrv.sblab.local" would be used in CloudStack configuration.
 - In this scenario a short TTL of e.g. 60 seconds should be used.
 - Unrelated to CloudStack – but service records are heavily used in Windows Active Directory environments.
 - Depending on your OS the IP to DNS cached records are kept in the local DNS cache for the duration of each records TTL.



The Domain Name System

IP routing



- IP routing takes care of traffic destined for networks outside the local subnet.
- Routing is managed through a number of different routing propagation algorithms
- In its simplest form you find the current routing decisions in the local routing table
- The routing table tells which interface handles the default gateway, and which interfaces handle any specific subnets

IP routing

```
permitted by applicable law.  
root@r-3-VM:~# route -n  
Kernel IP routing table  
Destination     Gateway         Genmask        Flags Metric Ref    Use Iface  
0.0.0.0         10.1.63.254   0.0.0.0        UG    0      0        0 eth1  
10.1.32.0       0.0.0.0       255.255.224.0  U     0      0        0 eth1  
10.100.1.0      0.0.0.0       255.255.255.0  U     0      0        0 eth2  
169.254.0.0     0.0.0.0       255.255.0.0    U     0      0        0 eth0  
root@r-3-VM:~#
```

- Rohit will cover more advanced routing scenarios in his talk later this week.
- This will cover how iptables and routing work together for more granular routing scenarios



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

IP routing

Decimal	Binary	
Local 192.168.56.11/24	11000000.10101000.00111000.00001011 11111111.11111111.11111111.00000000	4 octets (8 bit number)
AND Remainder	11000000.10101000.00111000.00000000 -----.-----.-----.00001011	Network is 192.168.56.0 Host is .11
Remote 192.168.22.231/24	11000000.10101000.00010110.11100111 11111111.11111111.11111111.00000000	
AND Remainder	11000000.10101000.00010110.00000000 -----.-----.-----.11100111	Network is 192.168.22.0 Host is .231 >>> SEND TO DEFAULT GW

- The subnet mask will determine if a destination address is
 - Local: i.e. it is transmitted on the local subnet, or
 - Remote: it is transmitted to the default gateway, which handles onward routing

IP routing

- **Routing can only happen if the source and destination IP subnets don't overlap!**
- **CloudStack takes care of most routing scenarios on system VMs**
- **A few areas we do pay attention to:**
 - **Multiple zones:**
 - If we have multiple zones we ensure the guest isolation CIDR network is different in each zone to allow future routing over VPN between zones.
 - E.g. Zone1 is 10.1.1.0/24, Zone2 is 10.1.2.0/24
 - **VPC:**
 - If we run multiple VPCs in the same application stack we ensure the super-CIDR for each is unique in case we want to do inter-VPC peering.



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

- **How do I calculate network throughput?**
- **Network speed is measured in Mbps or Gbps**
 - 1 Gbps = 1 gigabit per second
 - 8 bits = 1 byte
 - $1\text{Gbps} = 1/8 \text{ GB/s} = 0.125\text{GigaByte per second} - \text{or } 125\text{MB}$
- **Transfer time for a 50GB file:**
 - 50GBytes = 400Gbits
 - Will take 400 seconds to transfer
 - This is theoretical.....

Network Address Translation - NAT

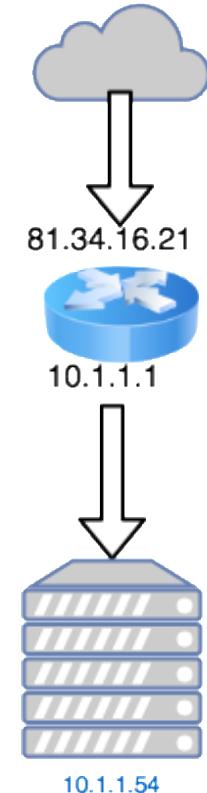
- **Translates an IP address to another**
- **OSI Layer3**
- **Source NAT:**
 - translates IP address during egress network operations
 - This is how your home broadband connection or the WiFi connection you currently use works.
 - Your local private IP address is never exposed outside your local subnet
 - This is the standard implemented on CloudStack virtual routers



Network Address Translation - NAT

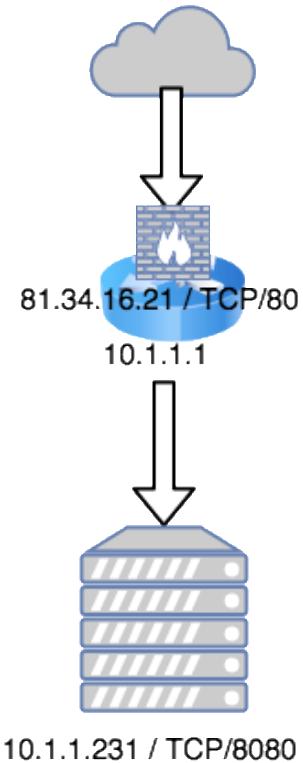
- **Destination NAT:**

- translates IP address during ingress network operations
- Used to present internal servers to the outside world
- In CloudStack we do this by configuring port forwarding.



Port Forwarding

- **Port forwarding:**
 - translates IP address AND TCP port during ingress network operations
 - Used to present internal servers to the outside world
 - Also requires us to open up the firewall to allow inbound traffic
 - E.g. present a web service running on a VM on IP 10.1.1.231 / TCP port 8080 to the outside world as IP 81.34.16.21 / TCP port 80 (HTTP)



Protocols recap

Layer	Application/Example	Central Device/Protocols	DOD4 Model
Application (7) Serves as the window for users and application processes to access the network services.	End User layer Program that opens what was sent or creates what is to be sent Resource sharing • Remote file access • Remote printer access • Directory services • Network management	User Applications SMTP	Process
Presentation (6) Formats the data to be presented to the Application layer. It can be viewed as the "Translator" for the network.	Syntax layer encrypt & decrypt (if needed) Character code translation • Data conversion • Data compression • Data encryption • Character Set Translation	JPEG/ASCII EBDIC/TIFF/GIF PICT	
Session (5) Allows session establishment between processes running on different stations.	Synch & send to ports (logical ports) Session establishment, maintenance and termination • Session support - perform security, name recognition, logging, etc.	Logical Ports RPC/SQL/NFS NetBIOS names	
Transport (4) Ensures that messages are delivered error-free, in sequence, and with no losses or duplications.	TCP Host to Host, Flow Control Message segmentation • Message acknowledgement • Message traffic control • Session multiplexing	F A C K E T T R I L F I L T E R I N G TCP/SPX/UDP	Host to Host
Network (3) Controls the operations of the subnet, deciding which physical path the data takes.	Packets ("letter", contains IP address) Routing • Subnet traffic control • Frame fragmentation • Logical-physical address mapping • Subnet usage accounting	Routers IP/IPX/ICMP	Internet
Data Link (2) Provides error-free transfer of data frames from one node to another over the Physical layer.	Frames ("envelopes", contains MAC address) [NIC card —> Switch —> NIC card] (end to end) Establishes & terminates the logical link between nodes • Frame traffic control • Frame sequencing • Frame acknowledgment • Frame delimiting • Frame error checking • Media access control	Switch Bridge WAP PPP/SLIP	Can be used on all layers
Physical (1) Concerned with the transmission and reception of the unstructured raw bit stream over the physical medium.	Physical structure Cables, hubs, etc. Data Encoding • Physical medium attachment • Transmission technique - Baseband or Broadband • Physical medium transmission Bits & Volts	Hub	Network

- **Layer 2 (data link)**
 - VLANs
- **Layer 3 (network)**
 - IP
 - ICMP
- **Layer 4 (transport)**
 - TCP
 - UDP
- **Layer 7 (application)**
 - DNS

- **Layer 3 (network)**
 - IP: IPv4, IPv6, handles IP addressing
 - ICMP: ping, traceroute, route advertisement
- **Layer 4 (transport)**
 - TCP: reliable, ordered, error checked delivery of packets, retransmits
 - UDP: no handshaking, no duplication protection, no guarantee of delivery

Unicast / multicast / broadcast

Transport	Details
Unicast	<ul style="list-style-type: none">• One-to-one IP transmission over TCP• E.g. HTTP from PC to webserver, i.e. both source and destination IP addresses clear and defined• Scope: internal or external to IP subnet
Multicast	<ul style="list-style-type: none">• One-to-several transmission• Receiver must be a member of a group to receive data – handshake uses IGMP• TCP only uses unicast, hence multicast must use UDP• Multicast address range: 224.0.0.0 – 239.255.255.255• Use cases: stream of video or financial data
Broadcast	<ul style="list-style-type: none">• One-to-all transmission• Scope: same subnet only• Broadcast address: last IP in range, e.g. 192.168.56.255• Use cases: DHCP, ARP

- **CIDR:**
 - Introduced by IETF in 1993 to slow growth of routing tables and slow the exhaustion of IPv4 addresses
 - Allows for further subdivision of IP classes and is based on variable subnet masking
 - In CloudStack we rely on this for VPC networking
- **Example:**
 - A /24 bit network has 254 IP addresses
 - This can be subdivided into:
 - 2 x /25 bit networks with 126 IP addresses each
 - 4 x /26 bit networks with 62 IP addresses each
 - 8 x /27 bit networks with 30 IP addresses each
 - Etc.....

CIDR – VPC example

- **I need:**
 - A 3-tier web / application / DB stack.
 - I will run maximum 64 hosts in my web tier, and less in the other tiers
 - I need full control of traffic between tiers
- **How do I build my CIDR network?**
 - 64 hosts can be not be handled by a /26 bit network (62 hosts), but is OK in a /25 bit network (126 hosts)
 - We choose to use the same size subnet for all three tiers:
 $3 \times 126 = 378$ IP addresses, this can only be met by a /23 with 512 addresses.



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

CIDR – VPC example

CIDR	Details	Host count and range
Super-CIDR	192.168.100.0 / 23	512 hosts 192.168.100.1 - 192.168.101.254
Web tier	192.168.100.0 / 25	126 hosts 192.168.100.1-127
Application tier	192.168.100.128 / 25	126 hosts 192.168.100.129-254
DB tier	192.168.101.0 / 25	126 hosts 192.168.101.1-127
Spare range	192.168.101.128 / 25	126 hosts 192.168.101.129-254



The Cloud Specialists

ShapeBlue.com

 @ShapeBlue

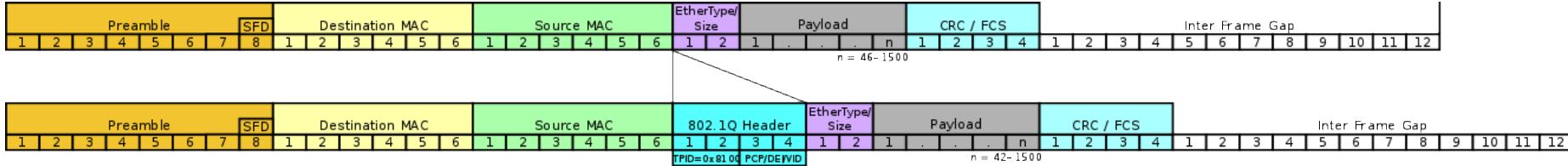
VLANs / 802.1Q

- VLANs are a Layer 2 mechanism**
- Ethernet frame is modified with additional header information

Layer	Application/Example	Central Device/Protocols	DOD4 Model
Application (7) Serves as the window for users and application processes to access the network services.	End User layer Program that opens what was sent or creates what is to be sent Resource sharing • Remote file access • Remote printer access • Directory services • Network management	User Applications SMTP	Process
Presentation (6) Formats the data to be presented to the Application layer. It can be viewed as the "Translator" for the network.	Syntax layer encrypt & decrypt (if needed) Character code translation • Data conversion • Data compression • Data encryption • Character Set Translation	JPEG/ASCII EBDIC/TIFF/GIF PICT	GATEWAY
Session (5) Allows session establishment between processes running on different stations.	Synch & send to ports (logical ports) Session establishment, maintenance and termination • Session support - perform security, name recognition, logging, etc.	Logical Ports RPC/SQ/LNFS NetBIOS names	Host to Host
Transport (4) Ensures that messages are delivered error-free, in sequence, and with no losses or duplications.	TCP Host to Host, Flow Control Message segmentation • Message acknowledgement • Message traffic control • Session multiplexing	F I L T E R P A C K E T TCP/SPX/UDP	Internet
Network (3) Controls the operations of the subnet, deciding which physical path the data takes.	Packets ("letter", contains IP address) Routing • Subnet traffic control • Frame fragmentation • Logical-physical address mapping • Subnet usage accounting	Routers IP/IPX/ICMP	Can be used on all layers
Data Link (2) Provides error-free transfer of data frames from one node to another over the Physical layer.	Frames ("envelopes", contains MAC address) [NIC card — Switch — NIC card] (end to end) Establishes & terminates the logical link between nodes • Frame traffic control • Frame sequencing • Frame acknowledgment • Frame delimiting • Frame error checking • Media access control	Switch Bridge WAP PPP/SLIP	Network
Physical (1) Concerned with the transmission and reception of the unstructured raw bit stream over the physical medium.	Physical structure Cables, hubs, etc. Data Encoding • Physical medium attachment • Transmission technique - Baseband or Broadband • Physical medium transmission Bits & Volts	Hub Land Based Layers	



VLANs / 802.1Q



- This longer frame size requires the network hardware to be VLAN aware
- The actual VLAN ID header is 12 bits: $2^{12} = 4096$ maximum VLANs

- **Pro's**

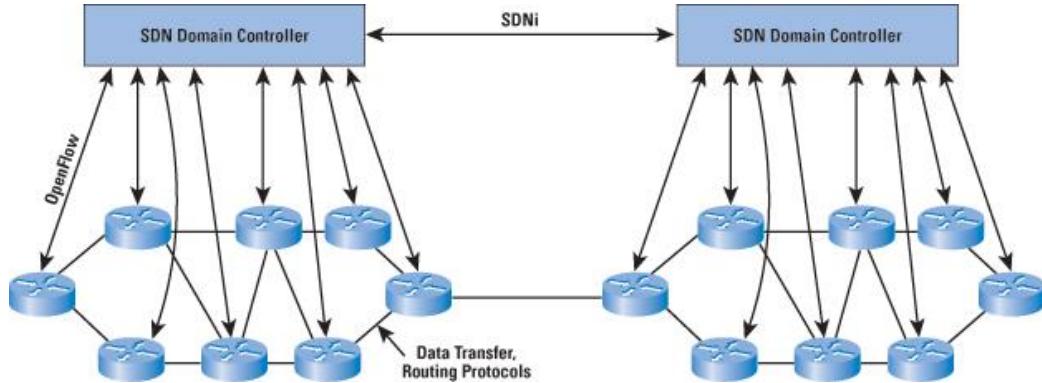
- Trusted, secure and mature technology
- Easy to implement
- Accepted as a networking security boundary, hence used in most CloudStack environments to segregate customer isolated networks

- **Con's**

- Theoretical limit of 4096 VLANs in any single environment
- From a CloudStack point of view this is within a single zone
- However – most ToR switches can only handle a subset – e.g. 1024 VLANs at any time
- This means 1024 isolated networks max in that rack!

Software Defined Networking - SDN

- **Cloud era networking concept.**
- **Different implementations work at layer 2/3/4**
- **Tries to centralize networking in one network component by disassociating the forwarding of network packets (data plane) from the routing process (control plane).**



- **Works with various protocols and tools – based around OpenFlow**
- **Requires centralised control in most cases – e.g. OpenDaylight**
- **Buy-in from a number of network vendors: Cisco, Arista, HP, Juniper, etc.**
- **In CloudStack we have various SDN integrations used for guest network isolation:**
 - Nicira / VMware NSX
 - VXLAN
 - Nuage
 - Various other protocols: GRE, L3VPN tunnelling

- **These isolation methods are not widely used:**
 - Using an enterprise grade SDN solution like NSX or Nuage is expensive
 - They also add more complexity that is required in most CloudStack environments need (keep in mind VLANs still provide up to 4096 isolated networks)
- **The lighter touch technologies like GRE and L3VPN rely on hypervisors maintaining “tunnels”:**
 - This means every inter-hypervisor connection needs a separate tunnel on a per isolated network basis.
 - This is very CPU intensive.

Networking basics

End of part 1 – any questions?



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Part 2 - CloudStack networking

Real world examples

Pictures speak a thousand words...



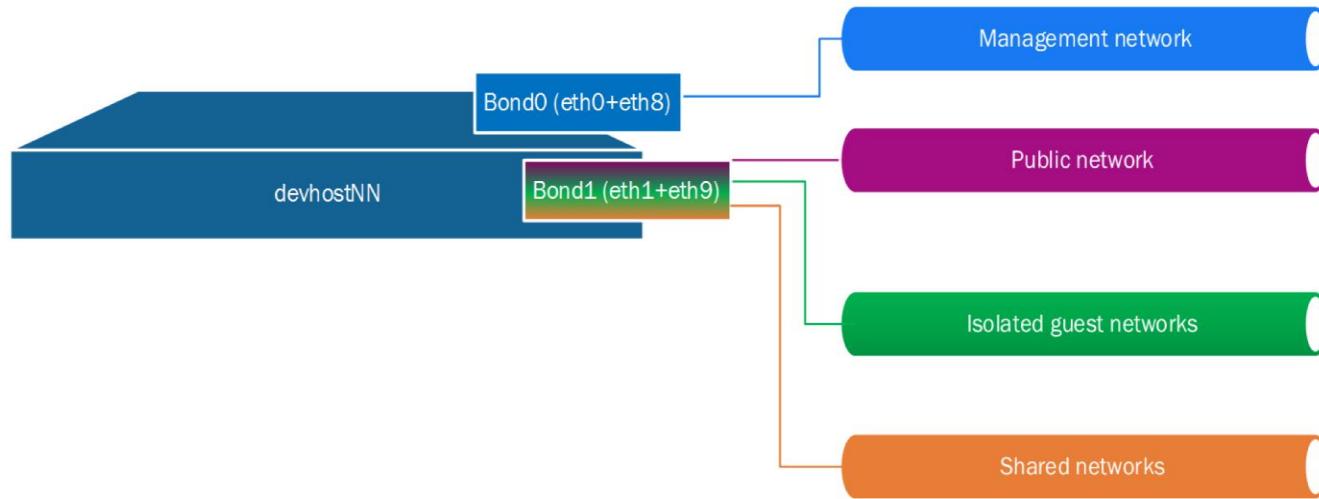
Blue The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Example network architecture – host networking



VMware

Navigator

Back

vCenter-sblab.sblab.local

SBLab2

P1-C1

10.2.0.12

10.2.0.13

10.2.0.14

10.2.0.22

/vmfs/volumes/57ceefac-e73efbd0-2c5...

/vmfs/volumes/57ceefac-e73efbd0-2c5...

/vmfs/volumes/57ceefac-e73efbd0-2c5...

/vmfs/volumes/57ceefac-e73efbd0-2c5...

/vmfs/volumes/57ceefac-e73efbd0-2c5...

/vmfs/volumes/57ceefac-e73efbd0-2c5...

/vmfs/volumes/57ceefac-e73efbd0-2c5...

/vmfs/volumes/57ceefac-e73efbd0-2c5...

i-11-94-VM

i-14-14036-VM

i-14-81-VM

i-14-8720-VM

i-140-11857-VM

i-1437-5636-VM

i-1453-12517-VM

i-1453-12710-VM

i-1453-13610-VM

i-1453-8715-VM

i-15-10012-VM

i-15-10062-VM

i-15-11462-VM

i-15-13972-VM

i-15-1864-VM

i-15-65-VM

i-15-79-VM

i-1982-7614-VM

i-1984-7622-VM

i-1984-7623-VM

i-1984-7624-VM

i-1984-7624-VM

10.2.0.14

Actions

Summary Monitor Configure Permissions VMs Datastores Networks Update Manager

Virtual switches

Storage Adapters Storage Devices Datastores Host Cache Configuration Protocol Endpoints I/O Filters

Virtual switches

VMKernel adapters Physical adapters TCP/IP configuration Advanced

Virtual Machines VM Startup/Shutdown Agent VM Settings Swap file location Default VM Compatibility

System Licensing Time Configuration Authentication Services Certificate Power Management Advanced System Settings System Resource Reservation Security Profile System Swap Host Profile

Switch

vSwitch0 vSwitch1

Discovered Issues

Management Network

VLAN ID: 6
VMKernel Ports (1)
vmk0 : 10.2.0.14

VM Network

VLAN ID: --
Virtual Machines (3)

cloud.guest.10.0.1-vSwit...

VLAN ID: 10
Virtual Machines (0)

cloud.guest.100.0.1-vSwit...

VLAN ID: 100
Virtual Machines (1)

cloud.guest.12.0.1-vSwit...

VLAN ID: 12
Virtual Machines (2)

cloud.guest.213.0.1-vSwit...

VLAN ID: 213
Virtual Machines (0)

Physical Adapters

vmnic3 10000 Full

The screenshot shows the VMware vSphere Web Client interface. The left sidebar displays a tree view of the vCenter server and its hosts. The main pane is titled 'Configure' and shows the 'Virtual switches' section. It lists two vSwitches: 'vSwitch0' and 'vSwitch1'. Below this, there are sections for 'Management Network', 'VM Network', and several other virtual networks, each with their respective VLAN IDs and associated virtual machines. A specific entry for 'Physical Adapters' is highlighted with an orange border, showing 'vmnic3' with a speed of '10000 Full'. The bottom right corner features the 'deBlue' logo.

XenServer

XenCenter

File View Pool Server VM Storage Templates Tools Window Help

Back Forward Add New Server New Pool New Storage New VM Shut Down Reboot Suspend System Alerts: 1

Views: Server View Logged in as: Local root account

xenserver.bootcamp.local

Search General Memory Storage Networking NICs Console Performance Users Logs

Server Networks

Networks

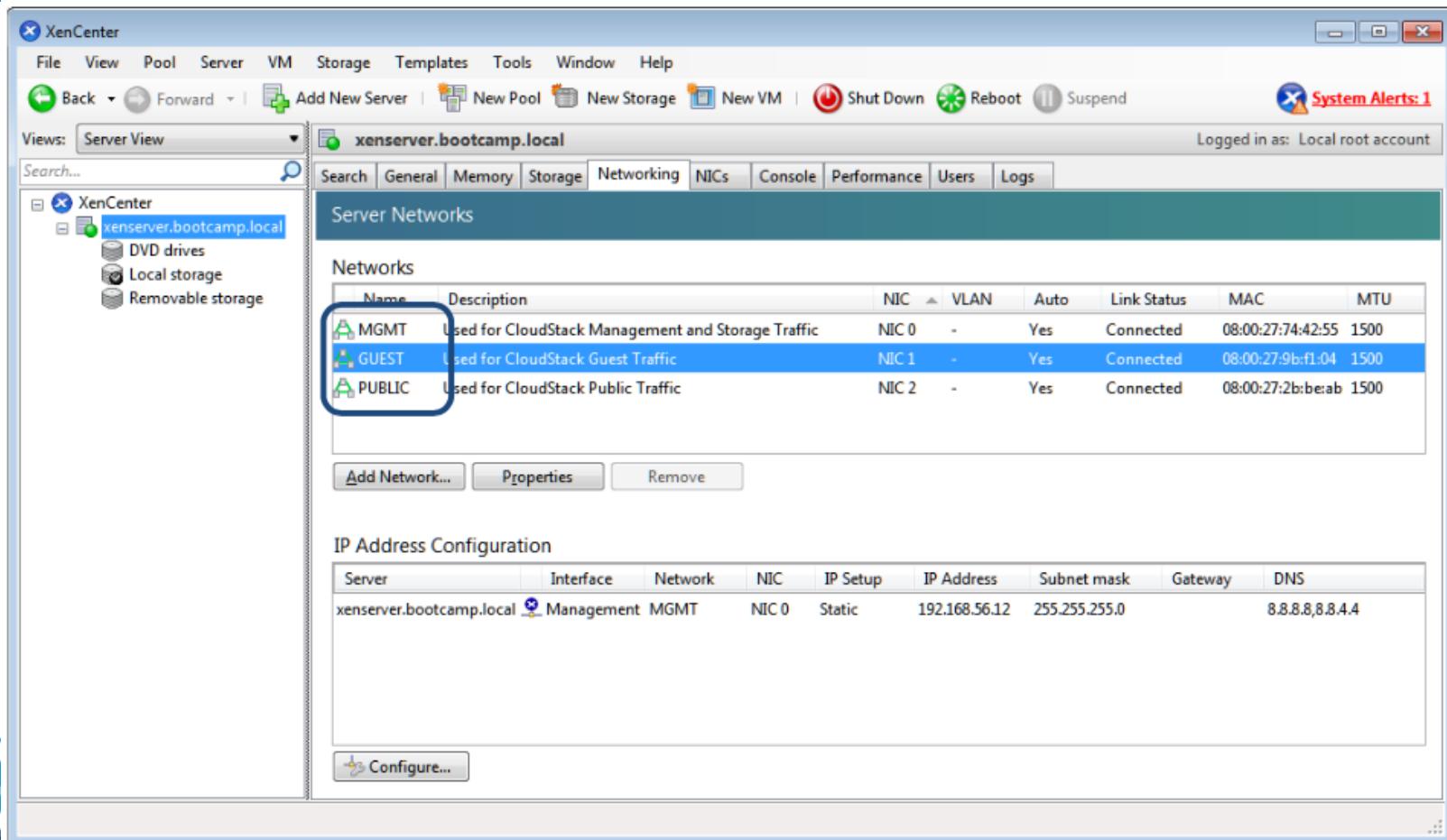
Name	Description	NIC	VLAN	Auto	Link Status	MAC	MTU
MGMT	Used for CloudStack Management and Storage Traffic	NIC 0	-	Yes	Connected	08:00:27:74:42:55	1500
GUEST	Used for CloudStack Guest Traffic	NIC 1	-	Yes	Connected	08:00:27:9b:f1:04	1500
PUBLIC	Used for CloudStack Public Traffic	NIC 2	-	Yes	Connected	08:00:27:2b:be:ab	1500

Add Network... Properties Remove

IP Address Configuration

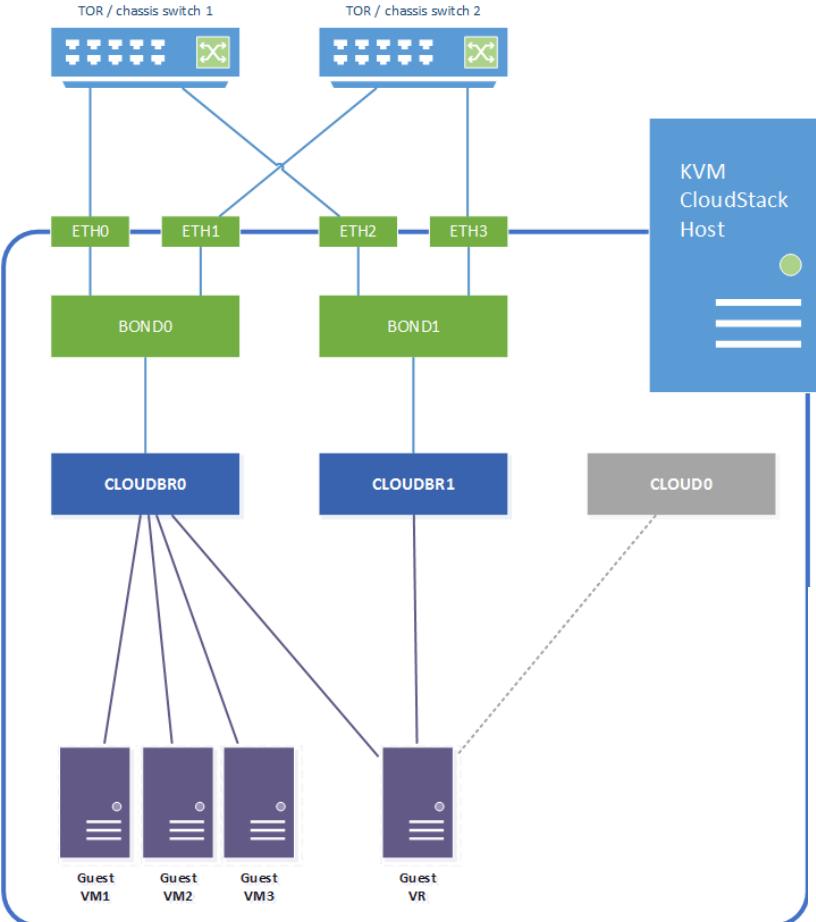
Server	Interface	Network	NIC	IP Setup	IP Address	Subnet mask	Gateway	DNS
xenserver.bootcamp.local	Management	MGMT	NIC 0	Static	192.168.56.12	255.255.255.0	8.8.8.8,8.8.4.4	

Configure...



Share Blue

KVM

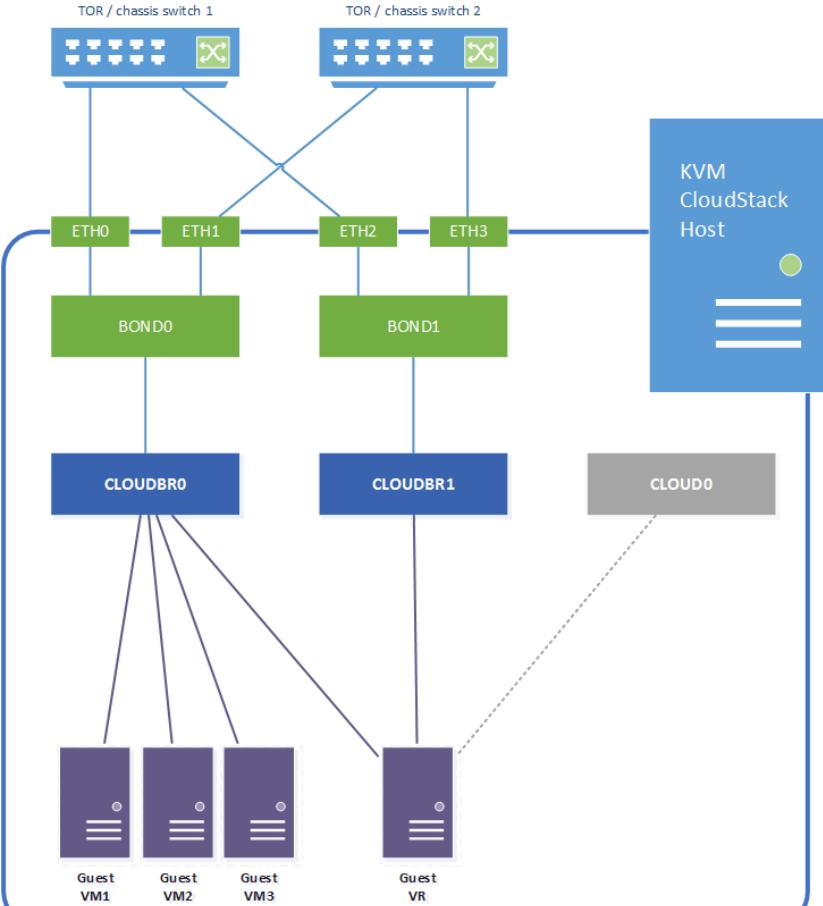


```
1 # vi /etc/sysconfig/network-scripts/ifcfg-eth0
2 DEVICE=eth0
3 TYPE=Ethernet
4 USERCTL=no
5 MASTER=bond0
6 SLAVE=yes
7 BOOTPROTO=none
8 HWADDR=00:0C:12:xx:xx:xx
9 NM_CONTROLLED=no
ONBOOT=yes
```

```
1 # vi /etc/sysconfig/network-scripts/ifcfg-bond0
2 DEVICE=bond0
3 ONBOOT=yes
4 BONDING_OPTS='mode=1 miimon=100'
5 BRIDGE=cloudbr0
6 NM_CONTROLLED=no
```

```
1 # vi /etc/sysconfig/network-scripts/ifcfg-cloudbr0
2 DEVICE=cloudbr0
3 ONBOOT=yes
4 TYPE=Bridge
5 IPADDR=192.168.100.20
6 NETMASK=255.255.255.0
7 GATEWAY=192.168.100.1
8 NM_CONTROLLED=no
DELAY=0
```

KVM

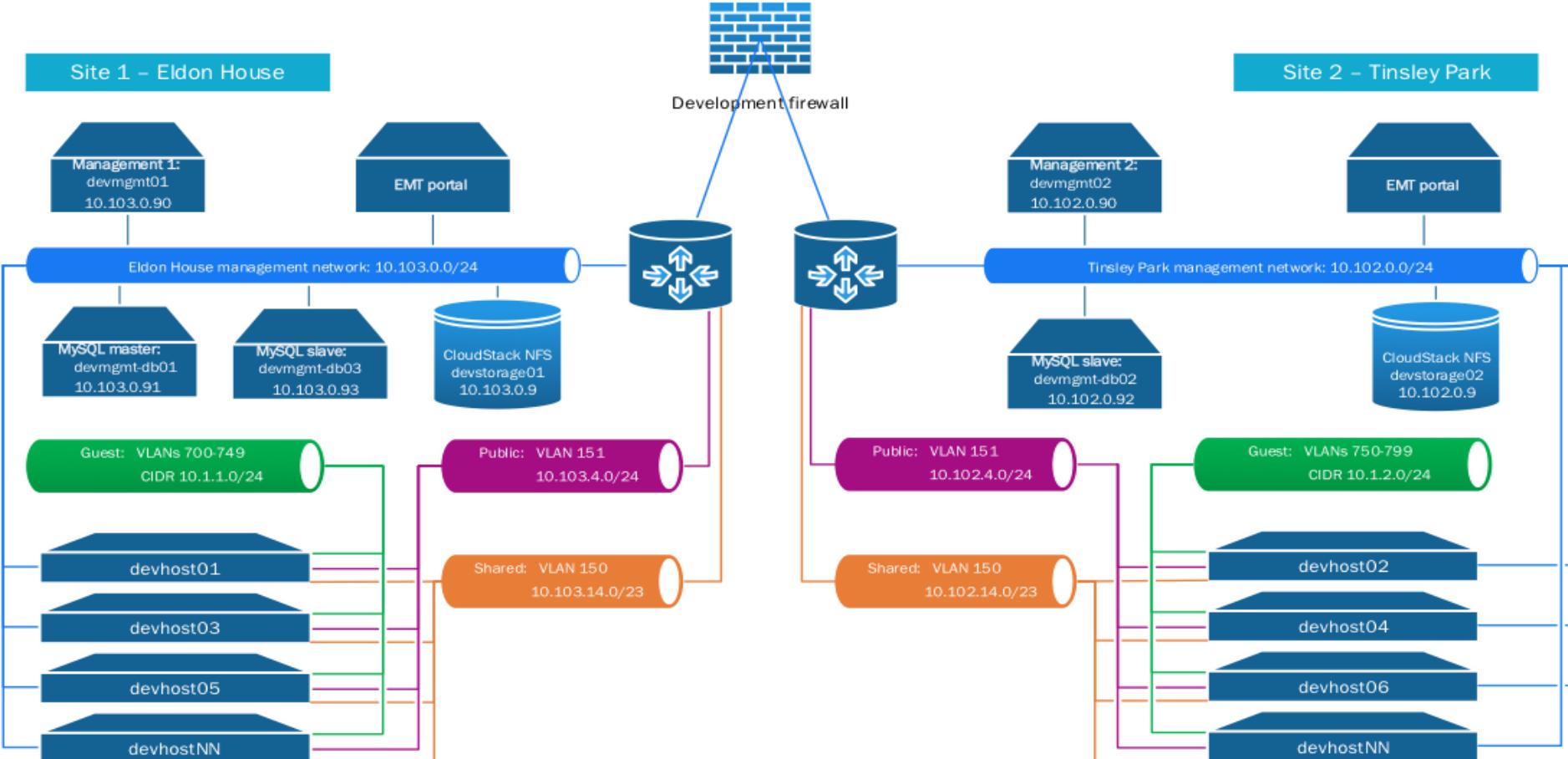


```
[root@tr1-1330-k-cs45-dsonstebo-kvm1 ~]# ifconfig | egrep 'mtulinet|breth1-1001|cloud0|cloudbr0|cloudbr1|eth0|eth1|eth1.1001|lo|virbr0|vnet0'
breth1-1001: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
cloud0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
          inet 169.24.0.1 netmask 255.255.0.0 broadcast 0.0.0.0
cloudbr0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
          inet 10.2.2.138 netmask 255.255.0.0 broadcast 10.2.255.255
cloudbr1: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
eth0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
eth1: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
eth1.1001: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
lo: flags=73<UP,LOOPBACK,RUNNING>  mtu 65536
          inet 127.0.0.1 netmask 255.0.0.0
virbr0: flags=4099<UP,BROADCAST,MULTICAST>  mtu 1500
          inet 192.168.122.1 netmask 255.255.255.0 broadcast 192.168.122.255
vnet0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
[root@tr1-1330-k-cs45-dsonstebo-kvm1 ~]#
```

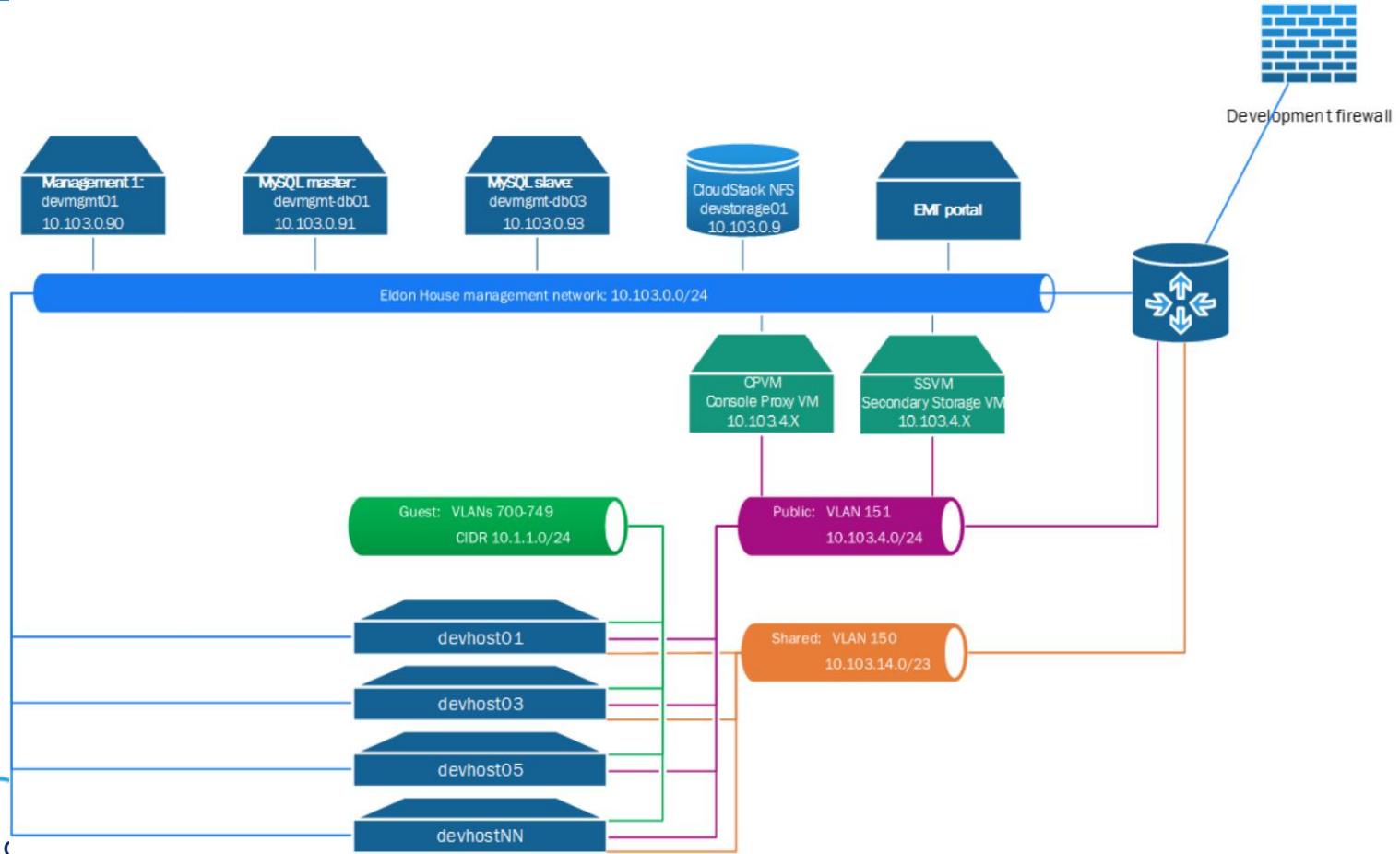
```
[root@tr1-1330-k-cs45-dsonstebo-kvm1 ~]# virsh dumpxml 2 | grep vnet
<target dev='vnet0' />
[root@tr1-1330-k-cs45-dsonstebo-kvm1 ~]# virsh list
Id  Name                           State
-----
2   i-2-3-VM                         running

[root@tr1-1330-k-cs45-dsonstebo-kvm1 ~]# virsh dumpxml 2 | grep vnet
<target dev='vnet0' />
[root@tr1-1330-k-cs45-dsonstebo-kvm1 ~]# brctl show
bridge name     bridge id      STP enabled    interfaces
breth1-1001     8000.06269c00094d    no          eth1.1001
                vnet0
cloud0          8000.000000000000    no
cloudbr0         8000.0668da010782    no          eth0
cloudbr1         8000.06269c00094d    no          eth1
virbr0          8000.525400733548    yes         virbr0-nic
[root@tr1-1330-k-cs45-dsonstebo-kvm1 ~]#
```

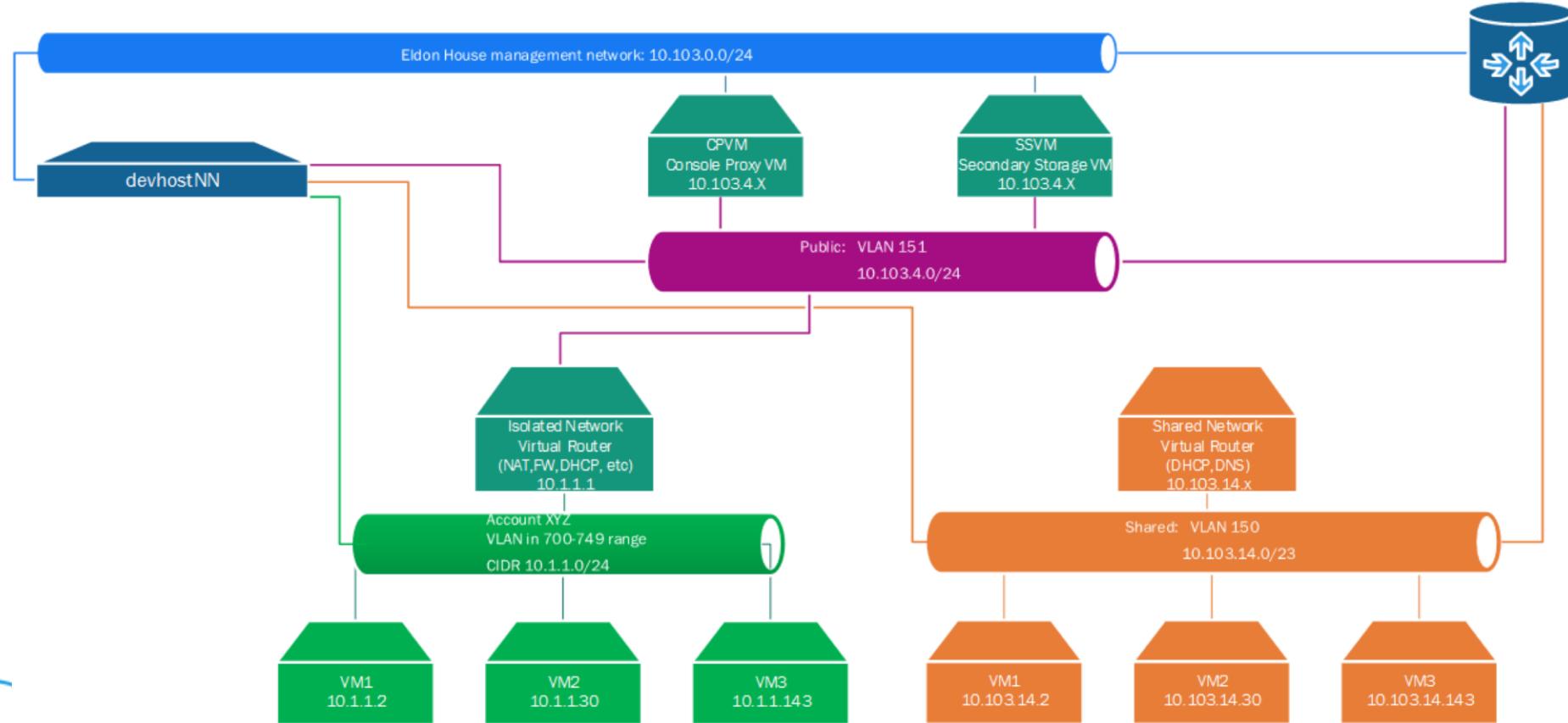
Example network architecture – dual site



Example network architecture – single site



Example network architecture – guest topology



CloudStack networking - bootcamp recap



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

CloudStack Architecture



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

- **Hierarchical structure enables massive scale.**
 - Region
 - A grouping of Availability Zones within a geographic area
 - Dedicated Management Server infrastructure to manage the Region and all of its Zones
 - Availability Zone
 - Typically one Zone per DC
 - Contains at least 1 POD, 1 Cluster and Secondary Storage
 - Network scope for advanced zones using L2 networking / VLANs.

CloudStack architecture

- **Pod**

- Logical entity, typically a rack containing one or more clusters and networking
- Uses concept of something shared i.e. switch stack or storage array
- Network scope (broadcast domain) for L3 networks in basic zones.

- **Cluster**

- Group of identical hosts running a common hypervisor
- Primary storage



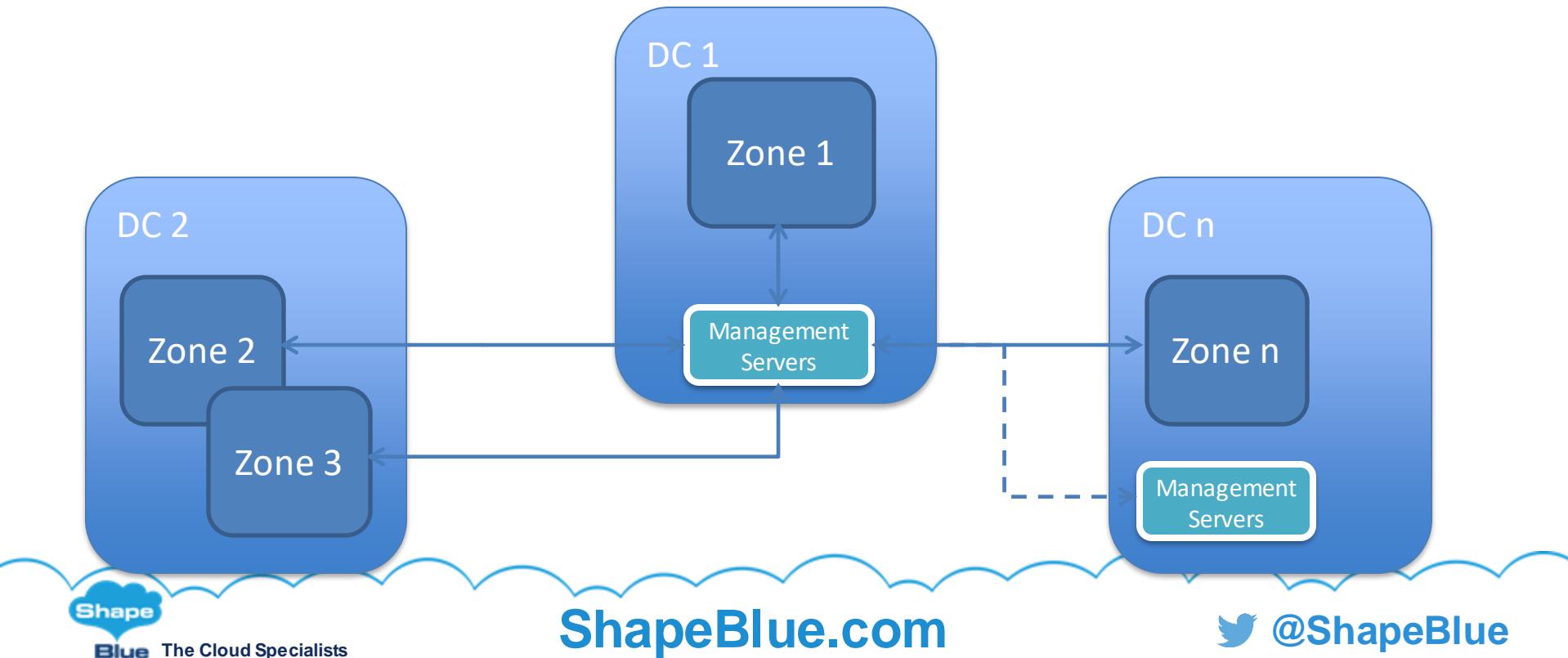
The Cloud Specialists

ShapeBlue.com



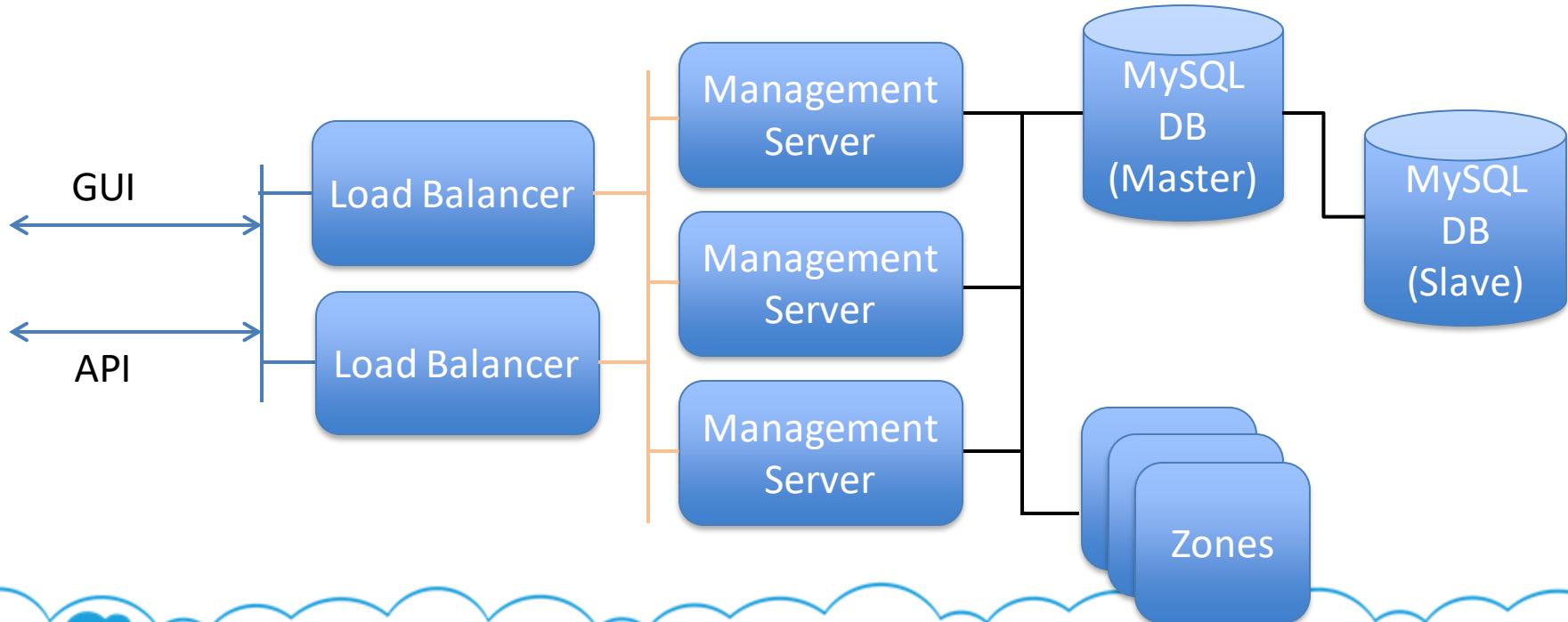
@ShapeBlue

Multiple Availability Zones within a Region



Management Server Deployment Architectures

- **Multi-Node Deployment**



CloudStack Networking



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

- **Logical Networking Models**
 - Basic
 - Advanced
 - Advanced with Security Groups

- **Basic zones**

- Guest isolation is provided through layer-3 security groups (IP address source filtering) by the hypervisor host.
- *Note this is only available on XenServer and KVM.*
- No VLANs.



- **Advanced zones:**

- Guest isolation is provided through layer-2 VLANs (or SDN technologies)
- This network model provides the most flexibility in defining guest networks and providing custom network offerings such as firewall, VPN, Load Balancer & VPC functionality.



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

- **Security Groups in Advanced Zones**

- Enables the deployment of multiple ‘Basic’ style networks which use security groups for isolation of VMs, but with each network isolated VLAN (or SDN).



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Physical Networks

- There are 4 ‘physical’ network types:
 - Management
 - Guest
 - Public
 - Storage
- Public and storage networks might not appear in all CloudStack deployments
- There may be multiple guest networks in an advanced zone.

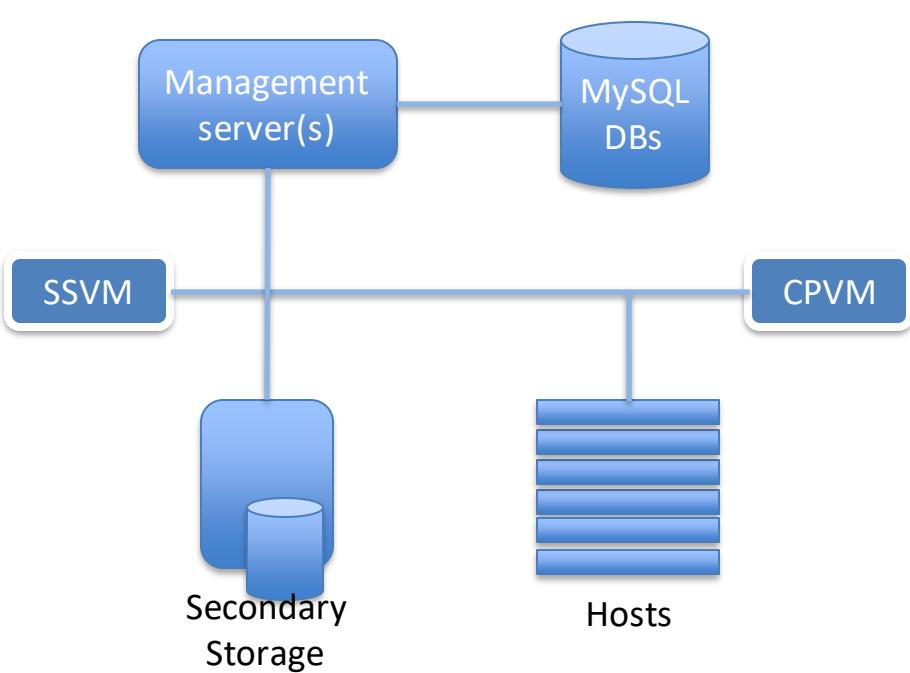


The Cloud Specialists

ShapeBlue.com

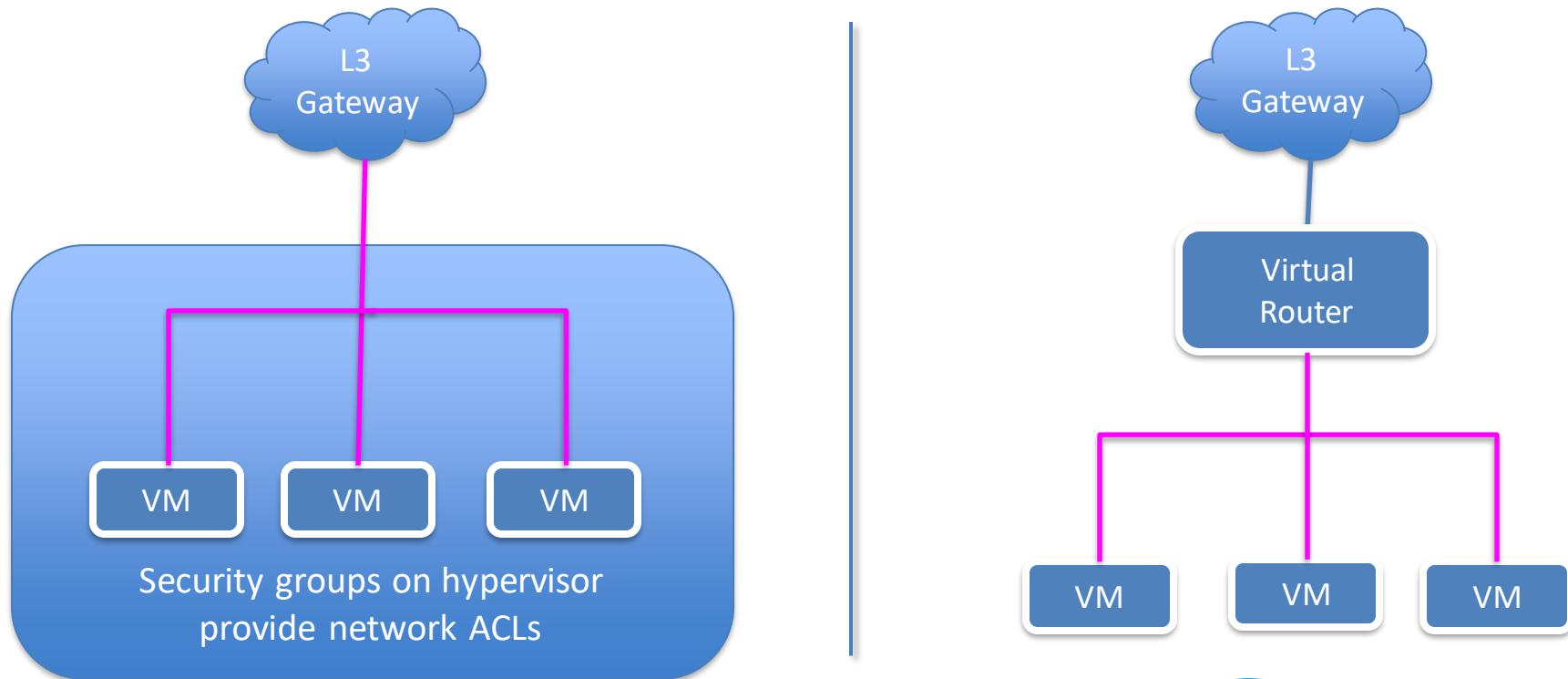
 @ShapeBlue

Management network



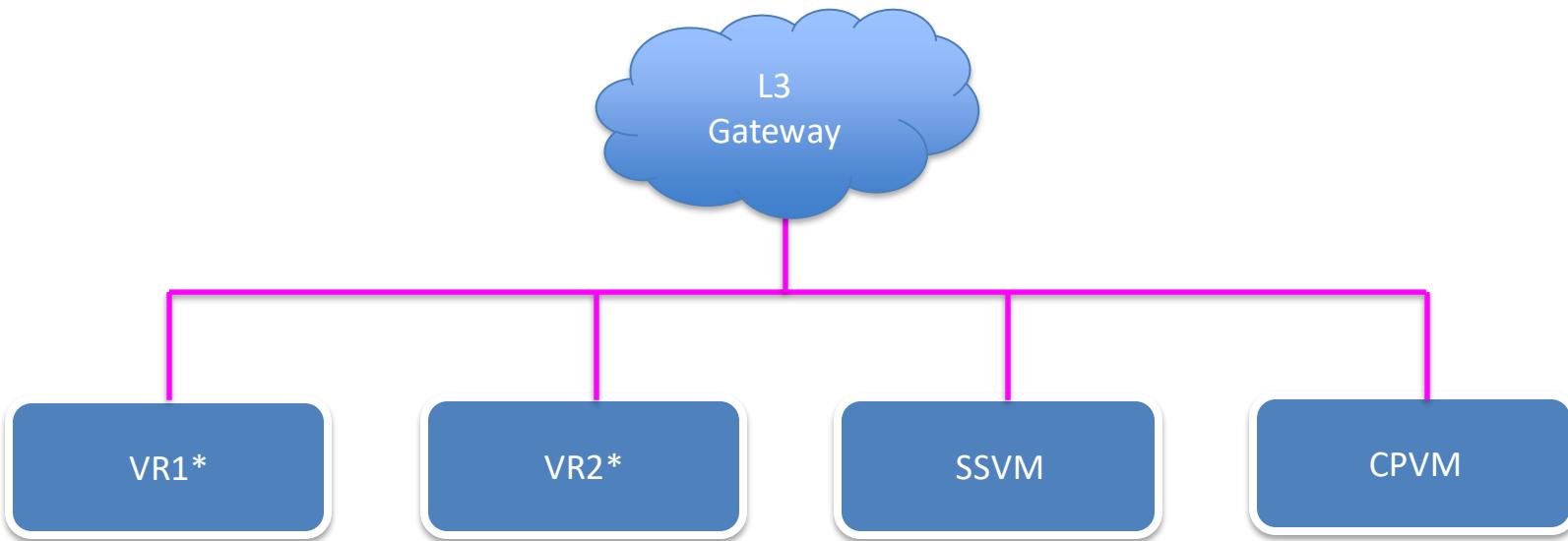
Traffic between CloudStack Management Servers and the various cloud components (Hosts, System VMs, Storage*, vCenter etc)

Guest Networks – Basic & Advanced



- **Public network ranges can be:**
 - Public internet exposed IP ranges.
 - Any other internal company wide private IP network.
- **Basic zone:**
 - End user VMs get assigned IP addresses in the public range (i.e. the guest IP range and the public IP range is the same).
 - When using Netscalers for EIP/ELB the public network is exposed on the northbound interface.
- **Advanced zone:**
 - Public network IP addresses are assigned to the public interface of the customer Virtual Router.

Public Network – System VMs

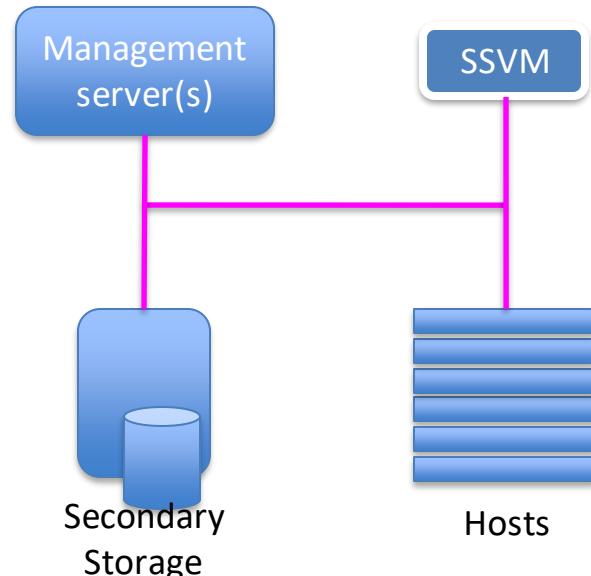


CPVM, SSVM & VRs have a connection to the Public Network

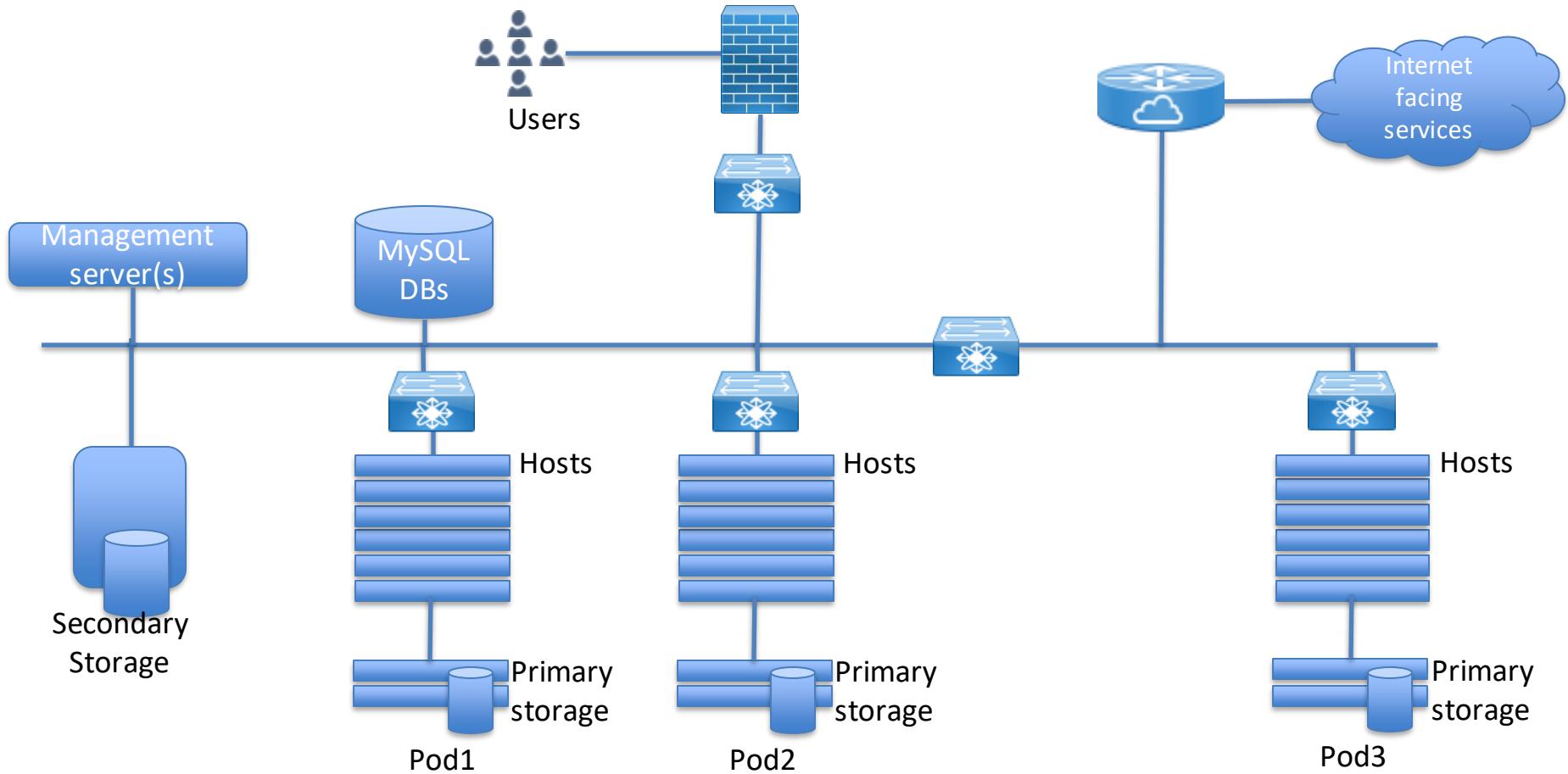
*VRs only have public connection in Advanced Network

Storage network

- Traffic between SSVM, secondary storage and hypervisors
- This is an optional network, traffic will use the management network if not configured.
- If configured, management server must still be able to communicate with ALL secondary storage pools to manage systemvm.iso and secondary storage housekeeping.
- It is NOT used for primary storage traffic.



Physical connectivity



Basic Networking



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Basic networking

- **AWS Style L3 isolation – enables massive scale**
- **Simple routed network for each pod**
- **Each pod has a unique CIDR (broadcast domain)**
- **Optional guest isolation via security groups**
- Optional NetScaler integration gives elastic IPs and elastic LB
- **Optional VMware NSX integration**



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Security Groups

- **Isolate traffic between VMs.**
- **Available in both Basic and Advanced zone networking models.**
- **Only supported on XenServer and KVM.**
- **XenServer must use Linux Bridge and not OpenvSwitch**
 - xe-switch-network-backend bridge
 - Edit sysctl to enable *net.bridge.bridge-nf-call-iptables* and *net.bridge.bridge-nf-call-arptables*
 - Must be implemented before adding to CloudStack



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

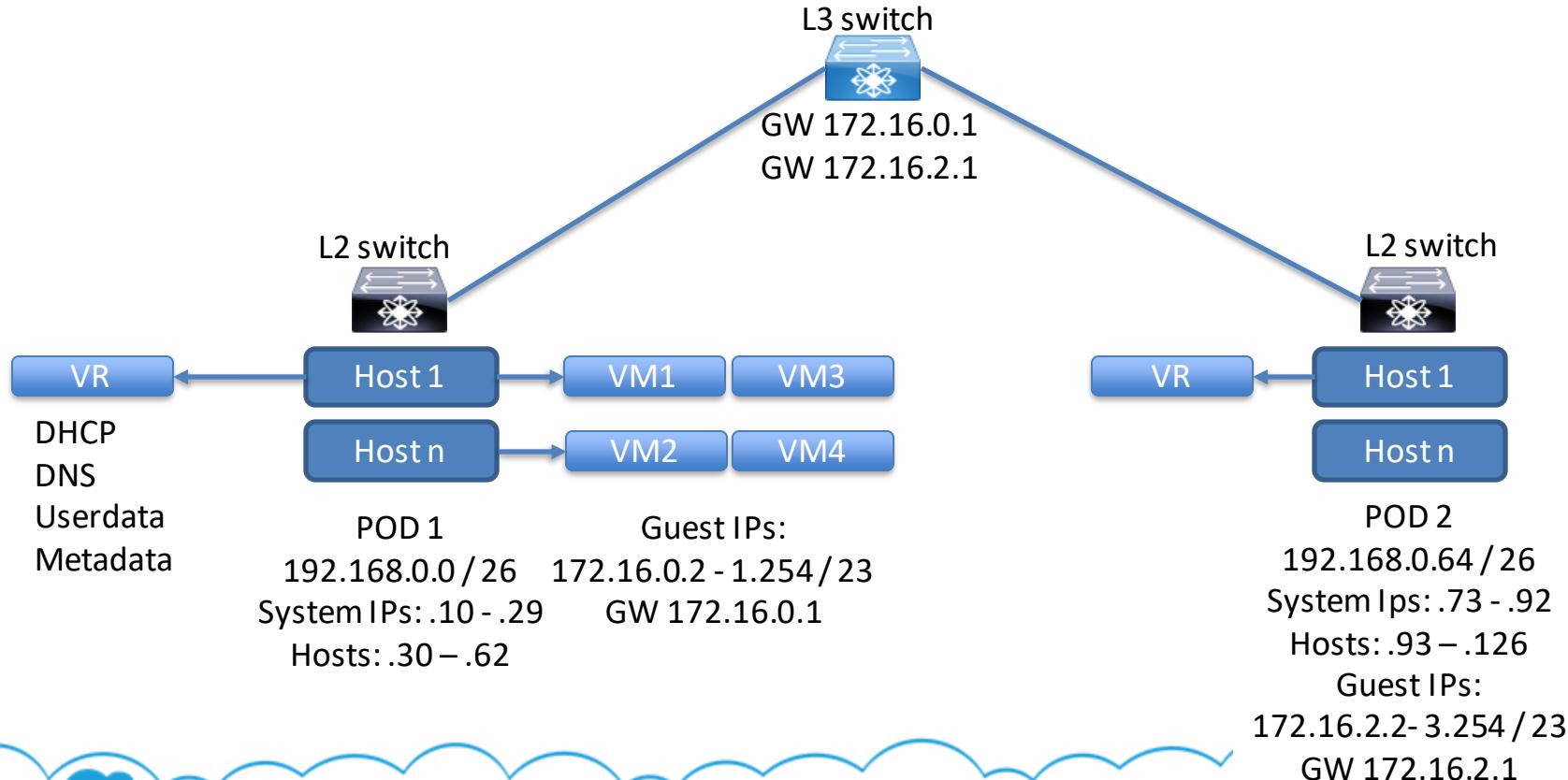
Security Groups

- Must be switched on when the zone is created (Advanced zones). Always enabled in basic zones.
- Uses ingress and egress rules to control traffic flow.
- Default is all outbound traffic allowed, all inbound denied.
- Rules can be mapped to CIDR or another account/security group.

Protocol	Start Port	End Port	CIDR	Add
TCP				<button>Add</button>
TCP	80	80	0.0.0.0/0	X

Protocol	Start Port	End Port	Account, Security group	Add
TCP				<button>Add</button>
TCP	3306	3306	geoff - App...	X

Basic zone – example IP schema



Advanced Networking



The Cloud Specialists

ShapeBlue.com

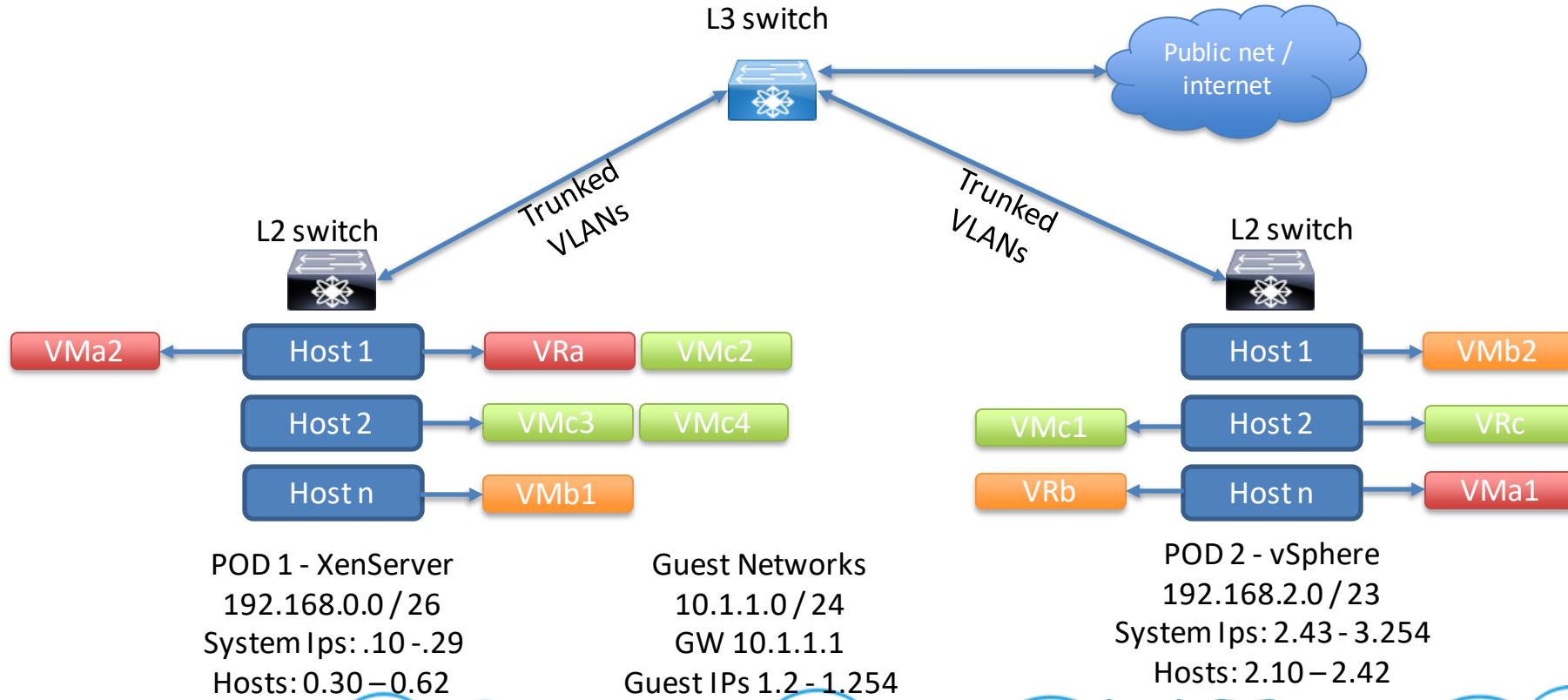


@ShapeBlue

- **Guest networks isolated by VLANs / SDN technologies**
- **Private and shared guest networks**
- **Multiple ‘physical’ guest networks possible**
- **Virtual Router for each network providing any of:**
 - DNS & DHCP
 - Firewall
 - Client VPN
 - Load balancing
 - Source / static NAT
 - Port forwarding



Advanced zone – example IP schema



Network Service Providers

- A hardware or virtual appliance that provide network services to CloudStack e.g.:

Virtual Router
VPC Virtual Router
Internal LBVM
Citrix NetScaler
F5 load balancer
Juniper SRX firewall
VMware NSX (Nicira)

Midokura Midonet
BigSwitch Vns
Cisco VNMC
Baremetal DHCP
Baremetal PXE
Palo Alto
OVS

System VM networking



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Accessing System VMs

- **XenServer / KVM (via host over the link local network)**
 - ssh onto the host which the VM is running on
 - ssh -i /root/.ssh/id_rsa.cloud -p 3922 root@169.254.n.n
-
- **VMware (via CloudStack management server)**
 - ssh onto the management Server
 - ssh -i /usr/share/cloudstack-common/scripts/vm/systemvm/id_rsa.cloud -p 3922 root@private-ip



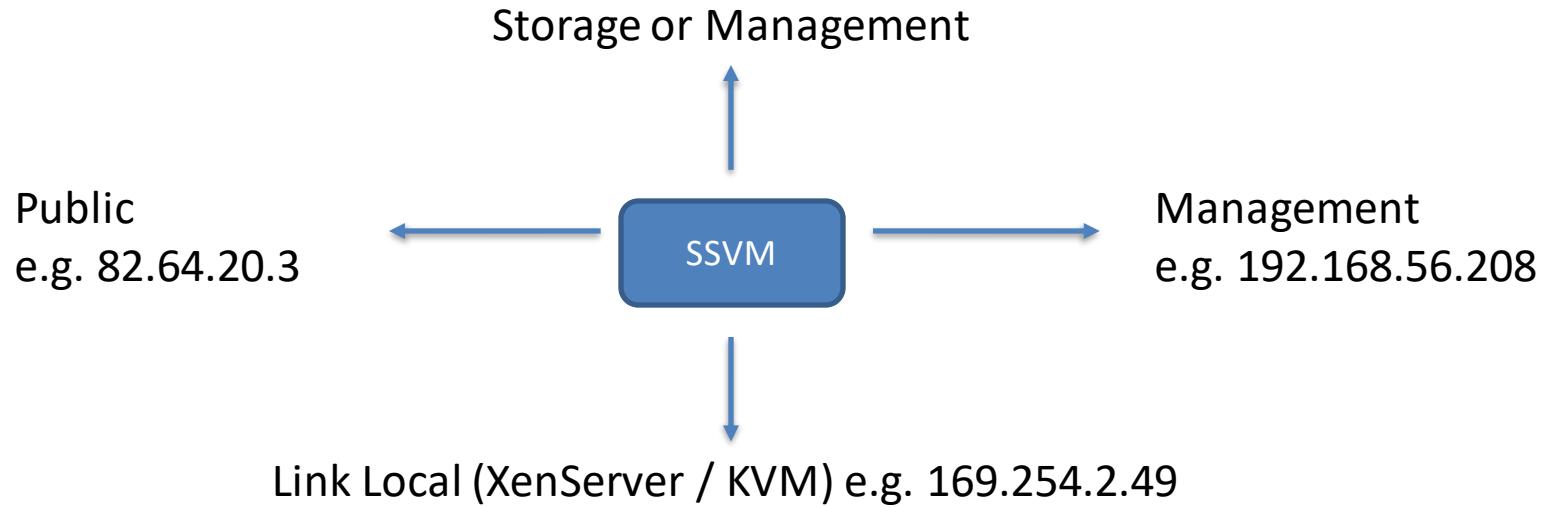
The Cloud Specialists

ShapeBlue.com

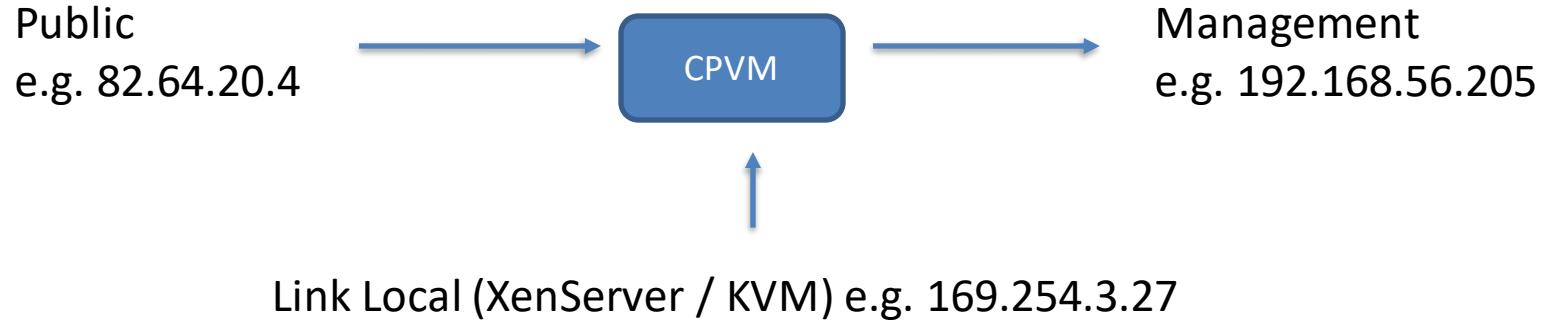


@ShapeBlue

SSVM networking



Console Proxy VM networking



Note direction of communications



Blue The Cloud Specialists

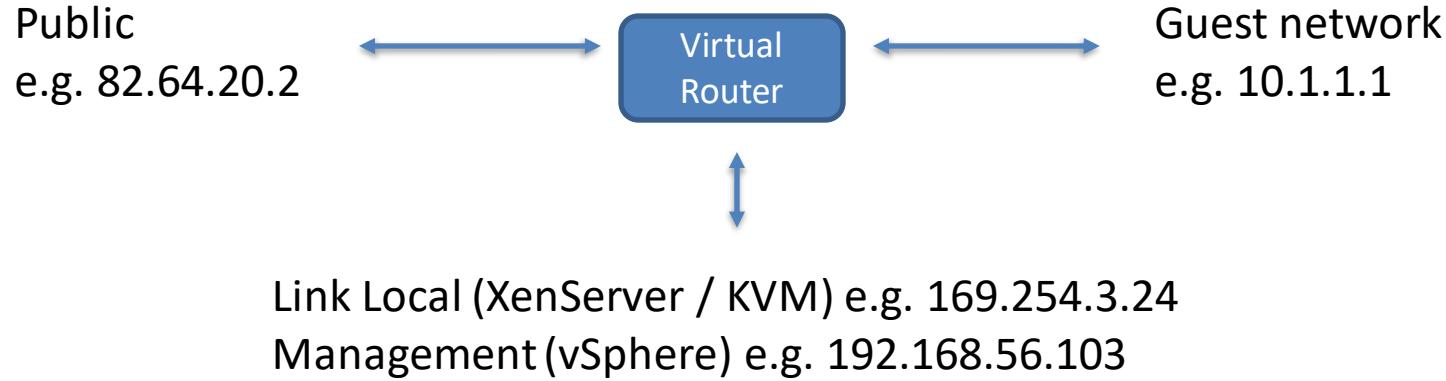
ShapeBlue.com



@ShapeBlue

Virtual Router

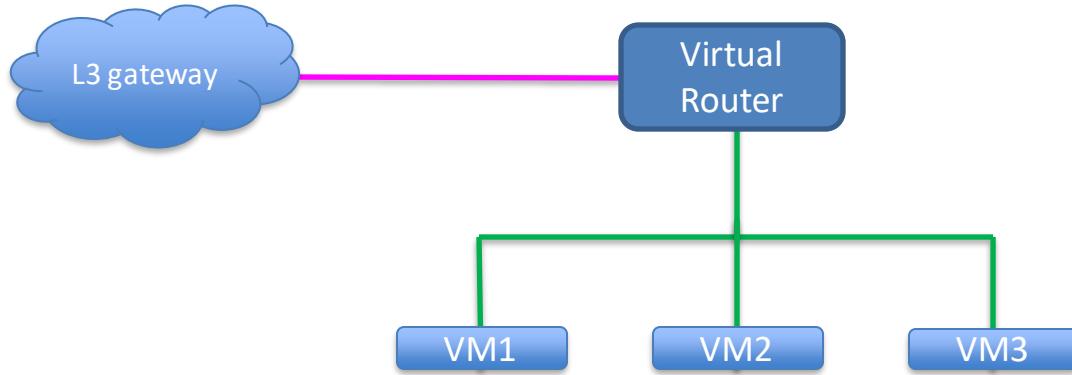
Networking



Virtual Router (Advanced Zone)

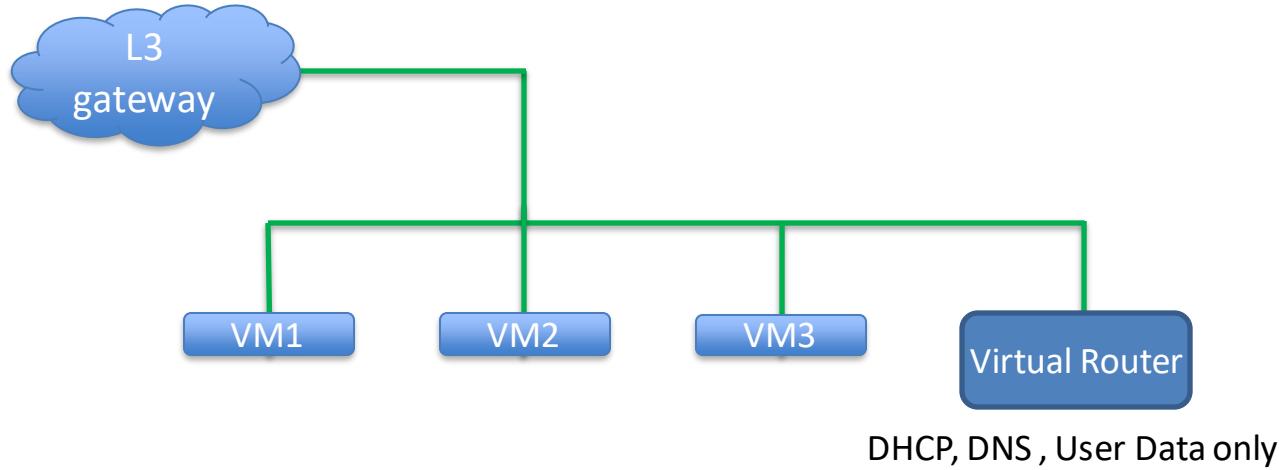
Networking

DHCP, DNS , User Data, Source NAT, Static NAT, VPN, Firewall, Port Forwarding, Load Balancing, Virtual Private Cloud

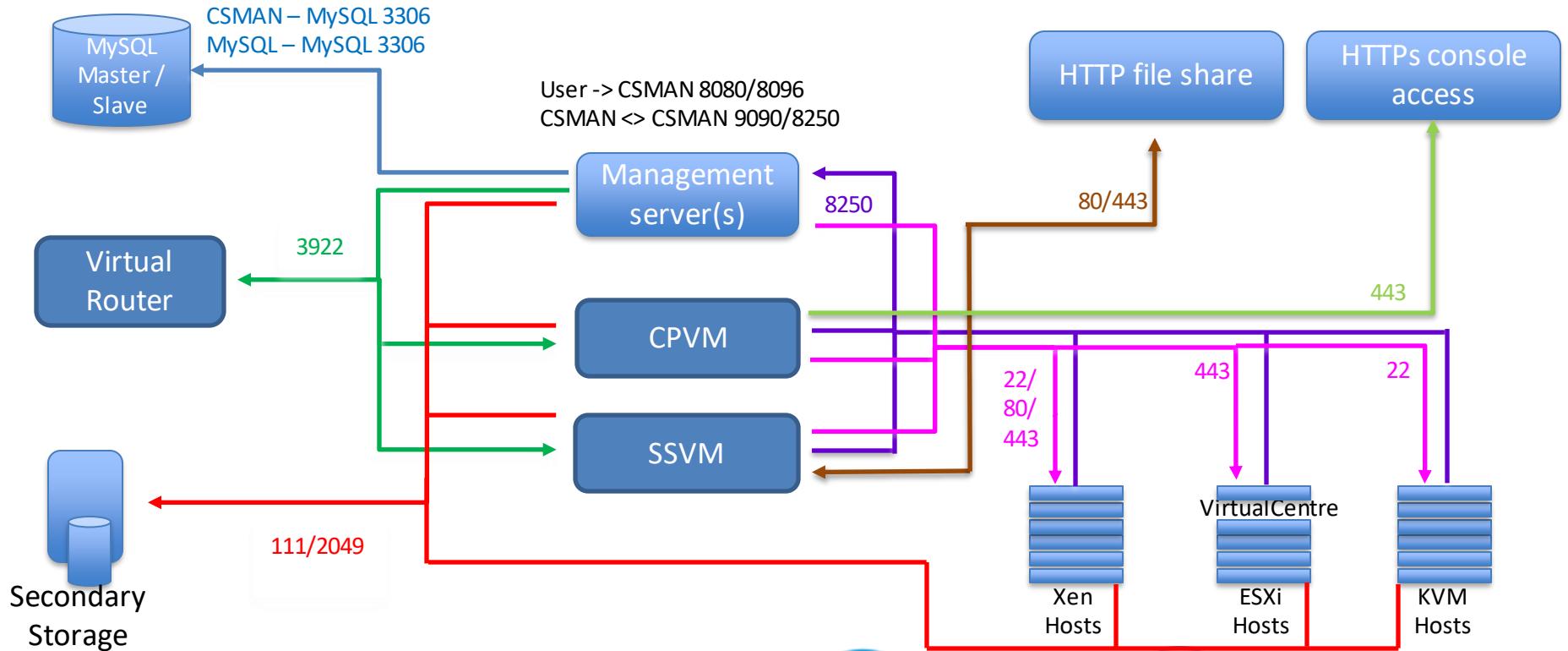


Virtual Router (Basic Zone)

Networking



Communication Ports



Virtual Private Cloud



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

- **Private multi-tiered virtual networks**
- **ACLs between tiers and public networks**
- **Inter-tier routing (layer 2)**
- **Site-2-site VPN**
- **Private gateway**
- **VPC-2-VPC VPN**
- **User VPN**
- **Inter-tier and inbound load balancing**

- **No ‘conserve mode’ so additional unique public IPs required for:**
 - Source NAT
 - Port forwarding
 - Load balancing
- **Redundant VPC now available**

Virtual Private Clouds (VPC)

- **User creates a super CIDR for the VPC**
- **All tiers' subnets must be within the Super CIDR and must not overlap**
- **E.g.**
 - Super CIDR: 10.0.0.0/16
 - Tier subnet1: 10.0.1.0/24
 - Tier subnet2: 10.0.2.0/24



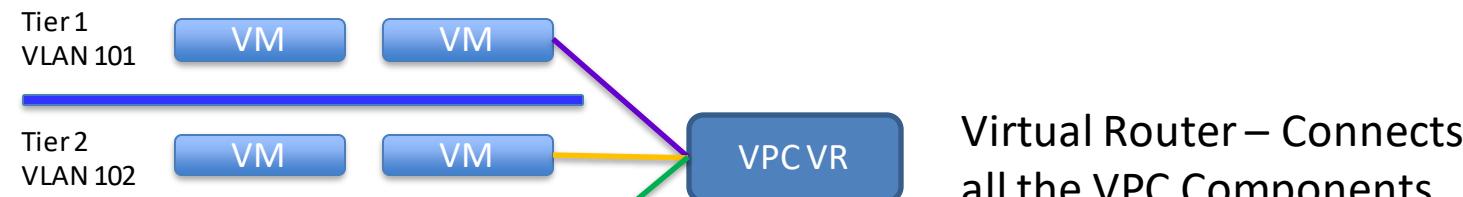
The Cloud Specialists

ShapeBlue.com



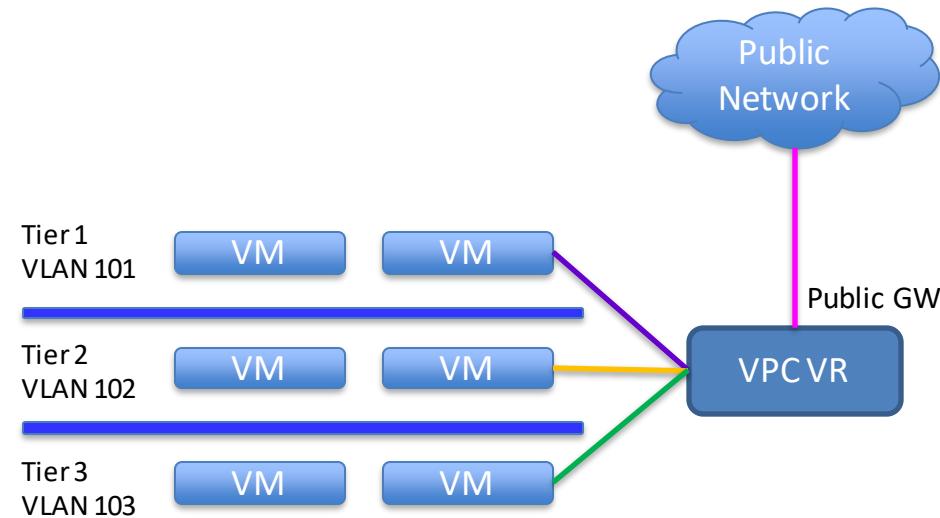
@ShapeBlue

VPC components

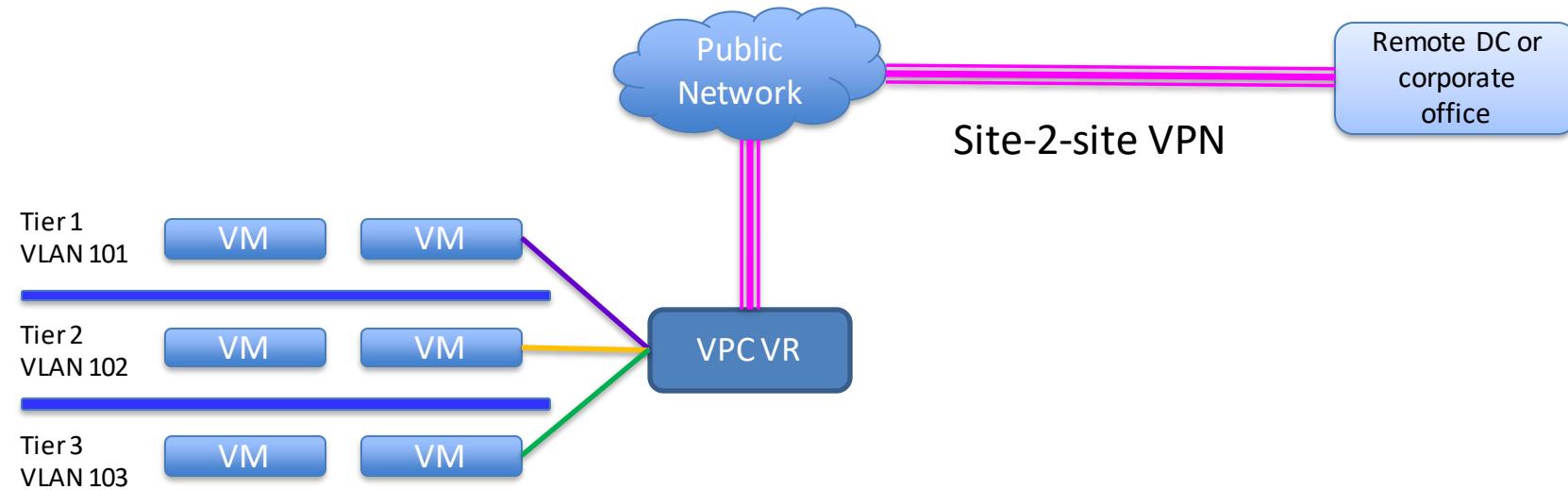


Network tiers – isolated networks, each with unique VLAN and CIDR

VPC components



VPC components



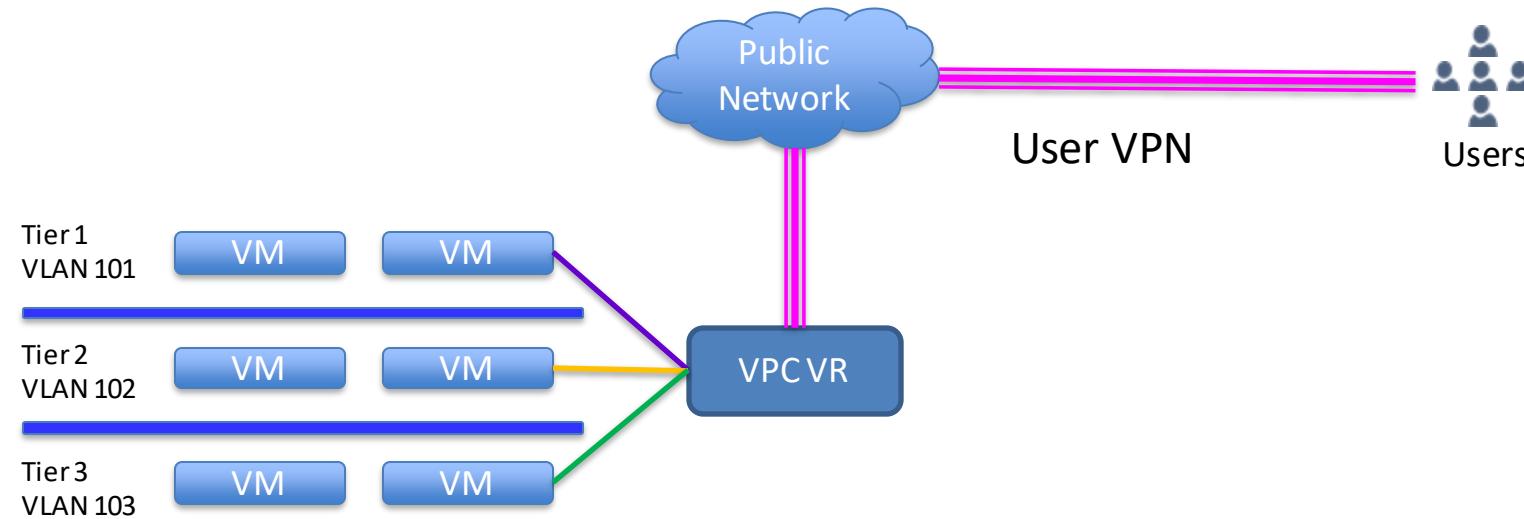
The Cloud Specialists

ShapeBlue.com

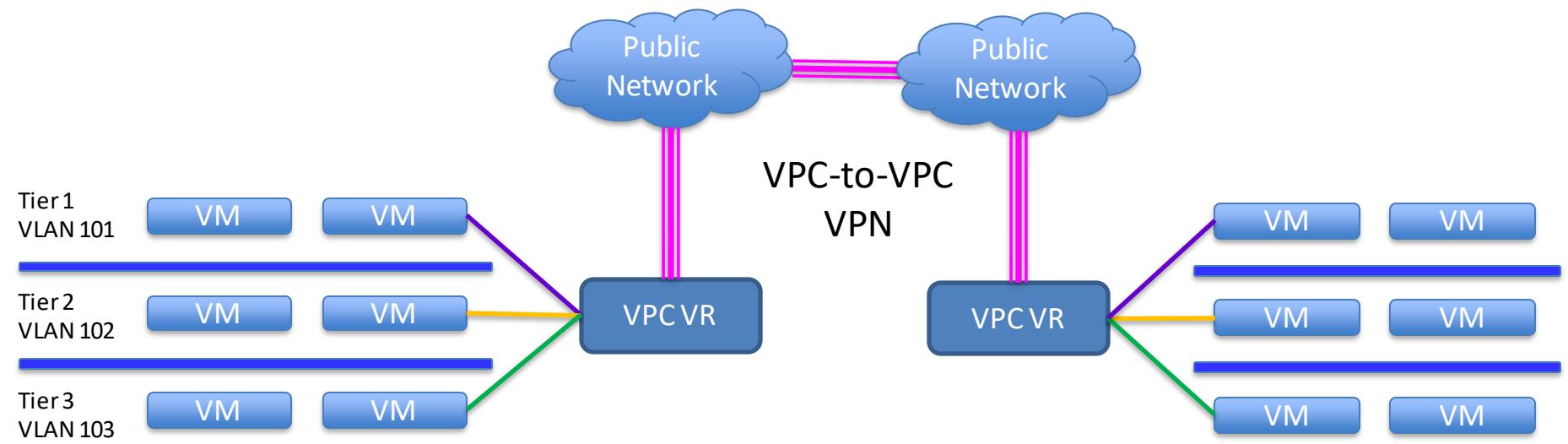


@ShapeBlue

VPC components



VPC components



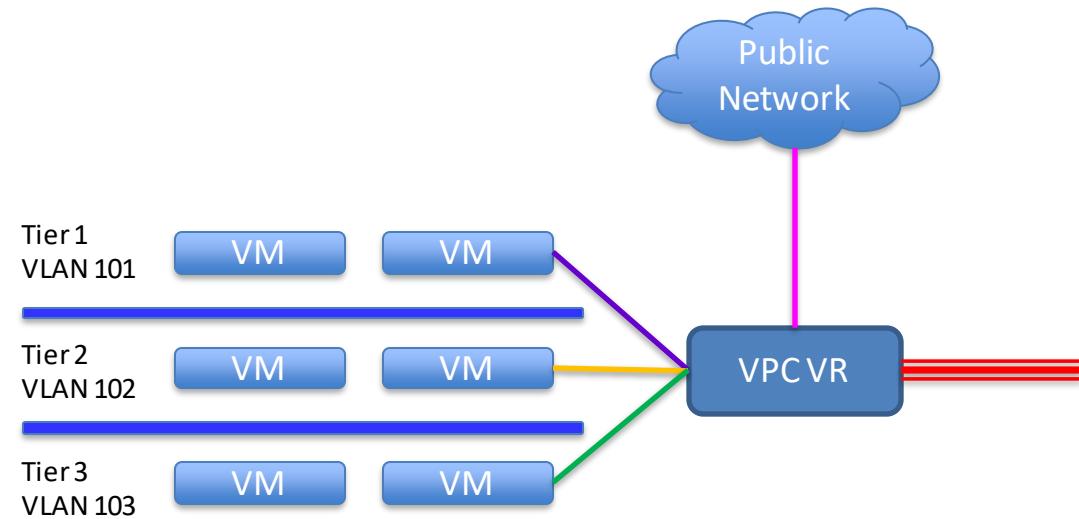
The Cloud Specialists

ShapeBlue.com



@ShapeBlue

VPC Private Gateway



Private Gateway

- Created by root admins
- Static routes must be set on VPC by users



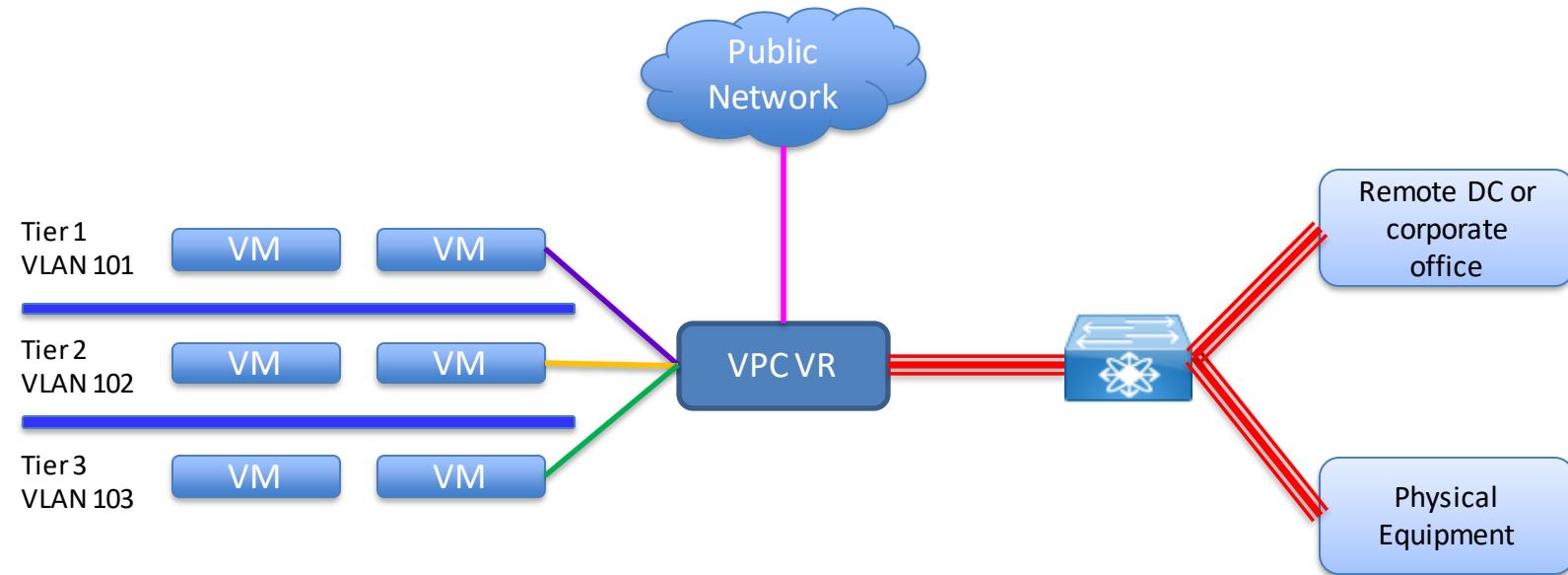
The Cloud Specialists

ShapeBlue.com



@ShapeBlue

VPC Private Gateway



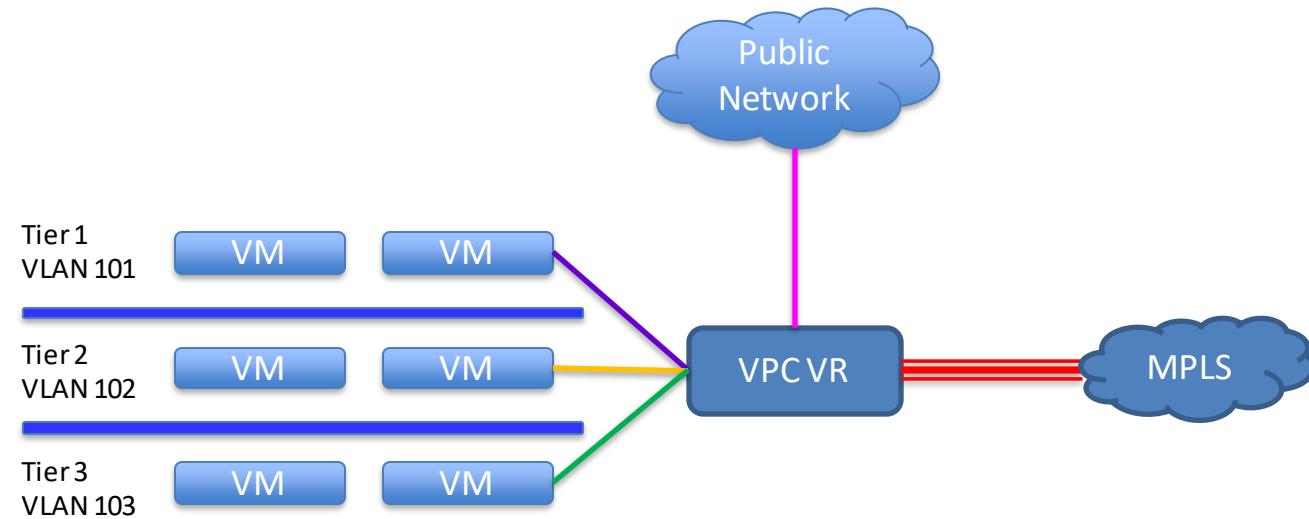
The Cloud Specialists

ShapeBlue.com

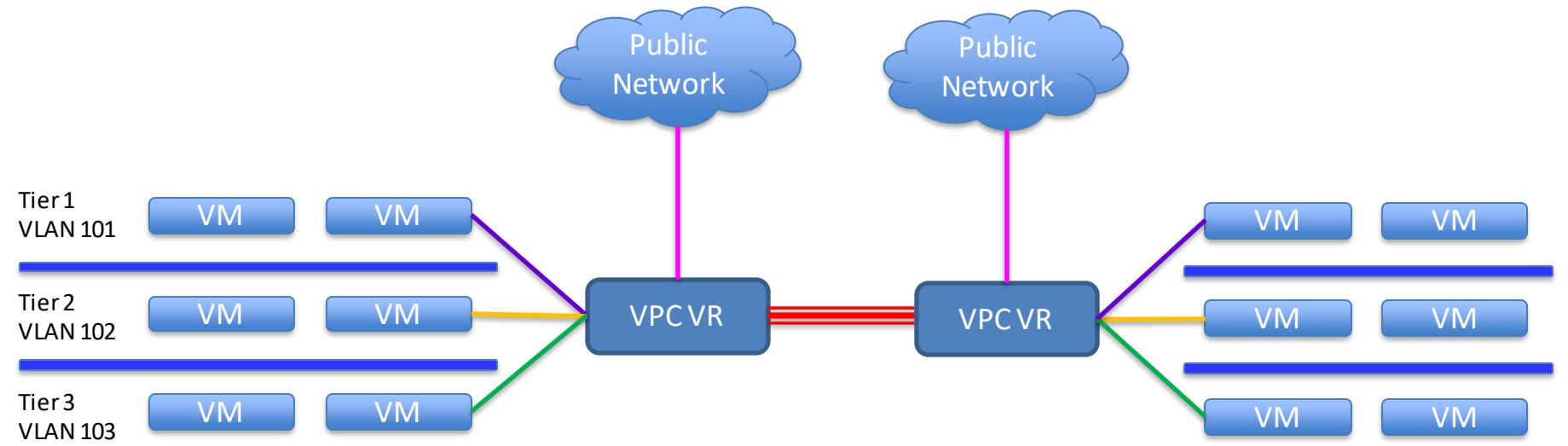


@ShapeBlue

VPC Private Gateway



VPC Private Gateway



Blue The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Bootcamp networking recap

End of part 2 – any questions?



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Part 3 – Network troubleshooting



Some comments before we start

- In this section we will mainly look at the Virtual Router, since this is where most of the CloudStack orchestrated network configuration takes place.
- This section is however just as applicable to other system VMs as well as hypervisor and management hosts – depending on flavour.
- In this session we want to give you the basics of network troubleshooting. We will revisit this later this in more depth.....



The Cloud Specialists

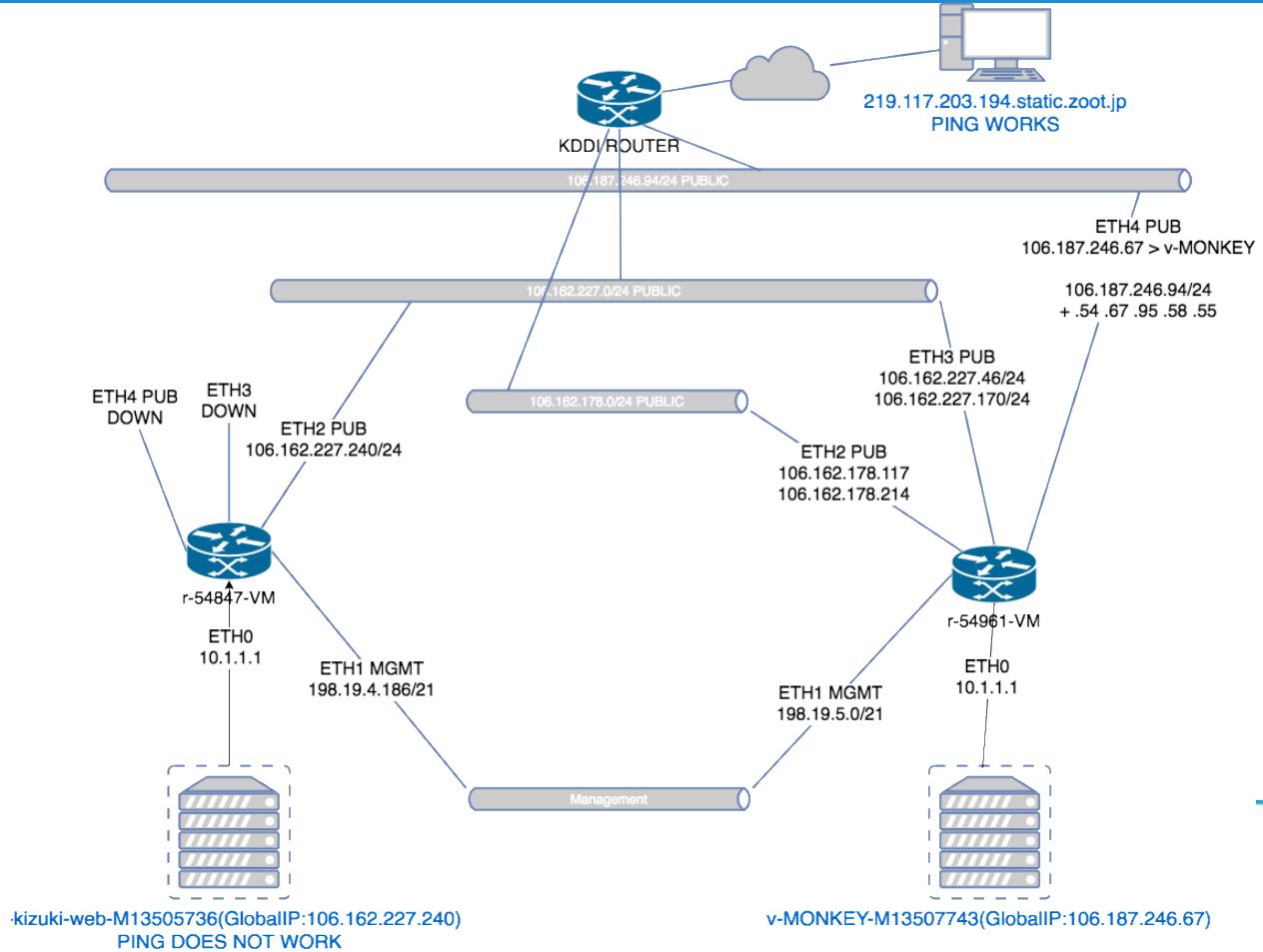
ShapeBlue.com



@ShapeBlue

VR networking – a complex example

- As CloudStack infrastructure age OR users get more advanced more public IP ranges tend to get added
- This increases the amount of public NICs on a VR
- Depending on DC switching and routing this can lead to asynchronous routing scenarios



Network troubleshooting toolbox

Tool	Details
iptables	Handles all system VM firewalling – including ACLs/firewall filter rules, NAT'ing, port forwarding, etc.
tcpdump	Allows us to capture traffic on network interfaces and interpret what is allowed in and out of e.g. a Virtual Router
ping	A simple ICMP tool to check connectivity to an IP address
traceroute	ICMP tool to check the full path a packet takes from source to destination
nmap	Port scanner – allows you to scan a set of ports and protocols on an end target to determine what traffic is allowed/blocked/dropped.
telnet	Simple telnet client which can be used to check Layer7 application response on a per port basis. A simple alternative to a full port scan.
<i>fill in your favourite tools here....</i>	<i>There are hundreds - if not thousands - of various network tools available, all depending on your OS and requirements....</i>

Starting simple – ICMP tools

- **Ping:**
 - ACK connectivity
 - Report packet loss
- **Traceroute**
 - Record path to host
 - Count hops

```
Dags-Mac-mini:~ dag$ ping 10.4.0.9
PING 10.4.0.9 (10.4.0.9): 56 data bytes
64 bytes from 10.4.0.9: icmp_seq=0 ttl=62 time=29.023 ms
64 bytes from 10.4.0.9: icmp_seq=1 ttl=62 time=30.713 ms
^C
--- 10.4.0.9 ping statistics ---
2 packets transmitted, 2 packets received, 0.0% packet loss
round-trip min/avg/max/stddev = 29.023/29.868/30.713/0.845 ms
```

```
Dags-Mac-mini:~ dag$ traceroute 10.4.0.9
traceroute to 10.4.0.9 (10.4.0.9), 64 hops max, 52 byte packets
 1  172.31.100.1 (172.31.100.1)  30.584 ms  26.611 ms  29.200 ms
 2  10.1.31.254 (10.1.31.254)  32.992 ms  27.912 ms  27.452 ms
 3  10.4.0.9 (10.4.0.9)  27.349 ms  28.464 ms  27.671 ms
Dags-Mac-mini:~ dag$
```

Starting simple – telnet

- **Telnet:**
 - Confirms port is responding without having to handshake a connection

```
Dags-Mac-mini:~ dag$ telnet 10.4.0.9 22
Trying 10.4.0.9...
Connected to 10.4.0.9.
Escape character is '^]'.
SSH-2.0-OpenSSH_7.4
^Cb^@^X\
Connection closed by foreign host.
```

```
Dags-Mac-mini:~ dag$ telnet 10.2.2.15 8250
Trying 10.2.2.15...
Connected to 10.2.2.15.
Escape character is '^]'.
Connection closed by foreign host.
Dags-Mac-mini:~ dag$ █
```



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

- A linux kernel firewall implemented as Netfilter modules
- Part of a bigger family of tools
 - Iptables: IPv4
 - Ip6tables: IPv6
 - Arptables: for ARP traffic
 - ebttables: for ethernet frames
- Consists of tables and chains within each table
 - Note the same chains are found in multiple tables
 - Packet traverse chains first – then tables

Iptables man page

Table	Details
filter	<p>This is the default table (if no -t option is passed). It contains the built-in chains:</p> <ul style="list-style-type: none">- INPUT (for packets destined to local sockets),- FORWARD (for packets being routed through the box), and- OUTPUT (for locally-generated packets).
nat	<p>This table is consulted when a packet that creates a new connection is encountered. It consists of three built-ins:</p> <ul style="list-style-type: none">- PREROUTING (for altering packets as soon as they come in),- OUTPUT (for altering locally-generated packets before routing), and- POSTROUTING (for altering packets as they are about to go out).

Iptables man page

Table	Details
mangle	<p>This table is used for specialized packet alteration. Until kernel 2.4.17 it had two built-in chains:</p> <ul style="list-style-type: none">- PREROUTING (for altering incoming packets before routing) and- OUTPUT (for altering locally-generated packets before routing). <p>Since kernel 2.4.18, three other built-in chains are also supported:</p> <ul style="list-style-type: none">- INPUT (for packets coming into the box itself),- FORWARD (for altering packets being routed through the box), and- POSTROUTING (for altering packets as they are about to go out).
raw	<p>This table is used mainly for configuring exemptions from connection tracking in combination with the NOTRACK target. It registers at the netfilter hooks with higher priority and is thus called before ip_conntrack, or any other IP tables. It provides the following built-in chains:</p> <ul style="list-style-type: none">- PRE-ROUTING (for packets arriving via any network interface)- OUTPUT (for packets generated by local processes)

Iptables rule of thumb

- Packet chain traversal:
 - Incoming packets destined for the local system:
PREROUTING -> INPUT
 - *This generally applies to VPN traffic as well as DHCP/DNS traffic on the VR*
 - Incoming packets destined to another host:
PREROUTING -> FORWARD > POSTROUTING
 - *This applies to all NAT'ed and port forwarded traffic which hits the VR and is routed to VMs etc.*
 - Locally generated packets:
OUTPUT -> POSTROUTING
 - *Less used*



The Cloud Specialists

ShapeBlue.com

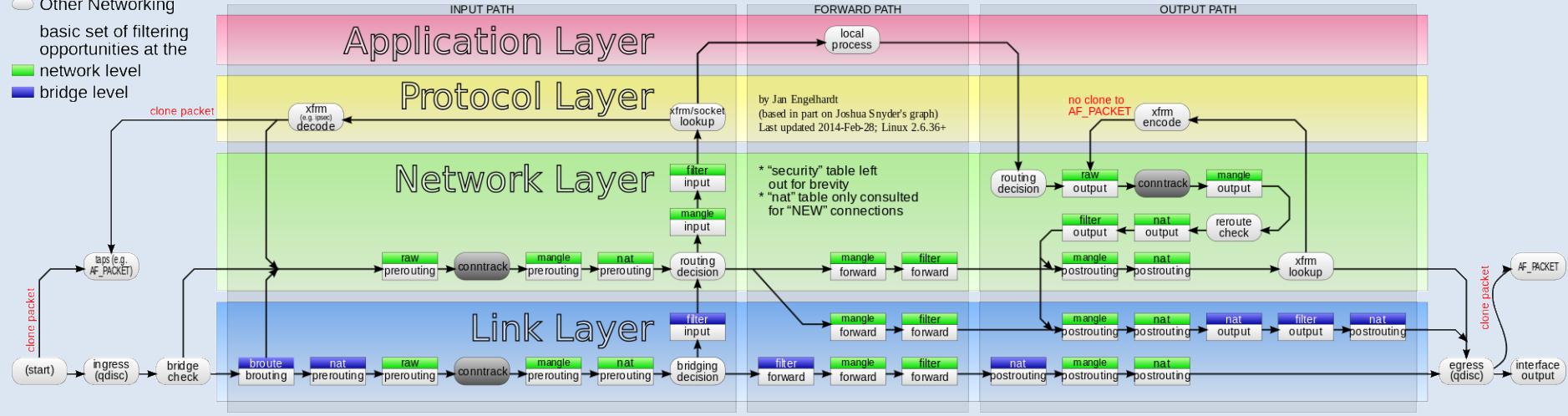


@ShapeBlue

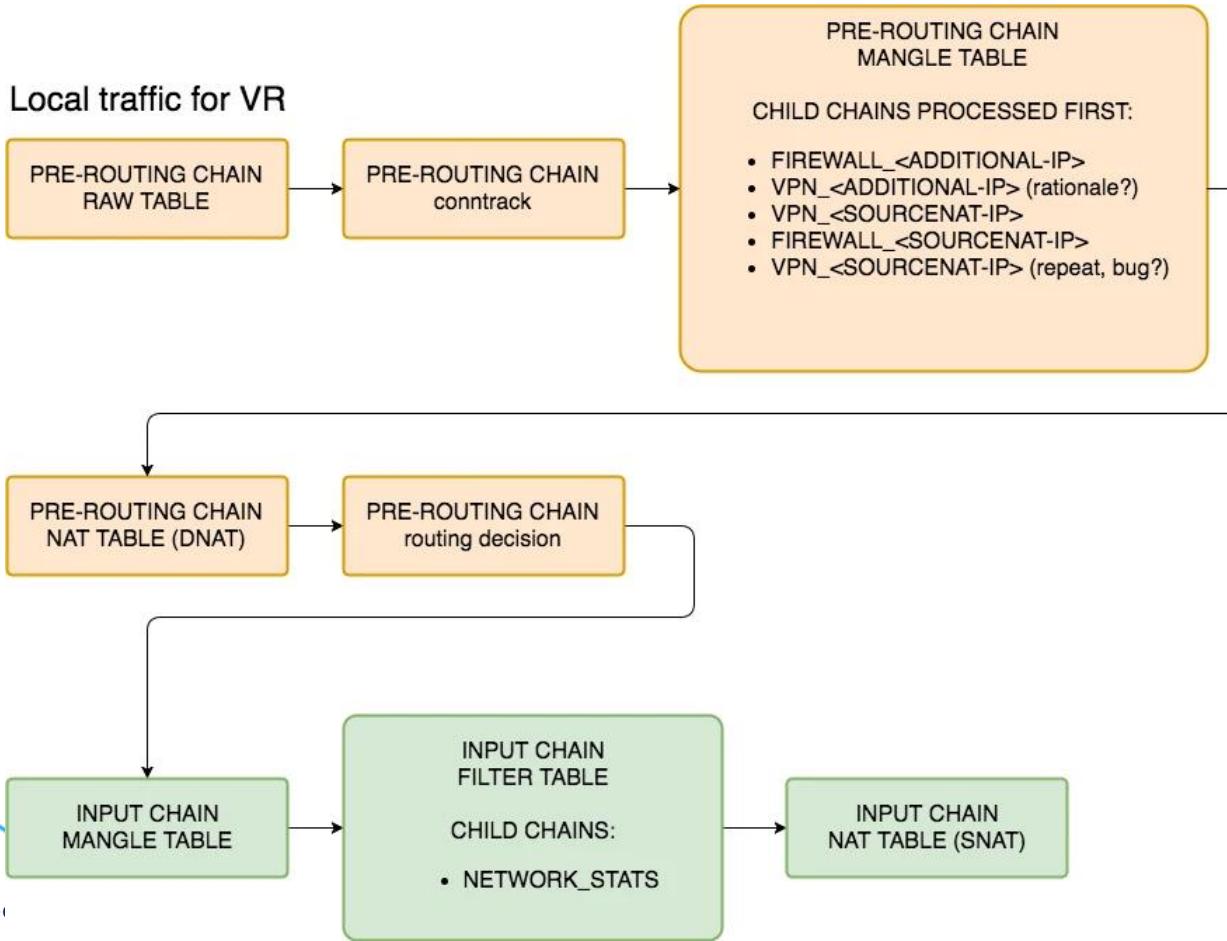
Overall packet flow

Packet flow in Netfilter and General Networking

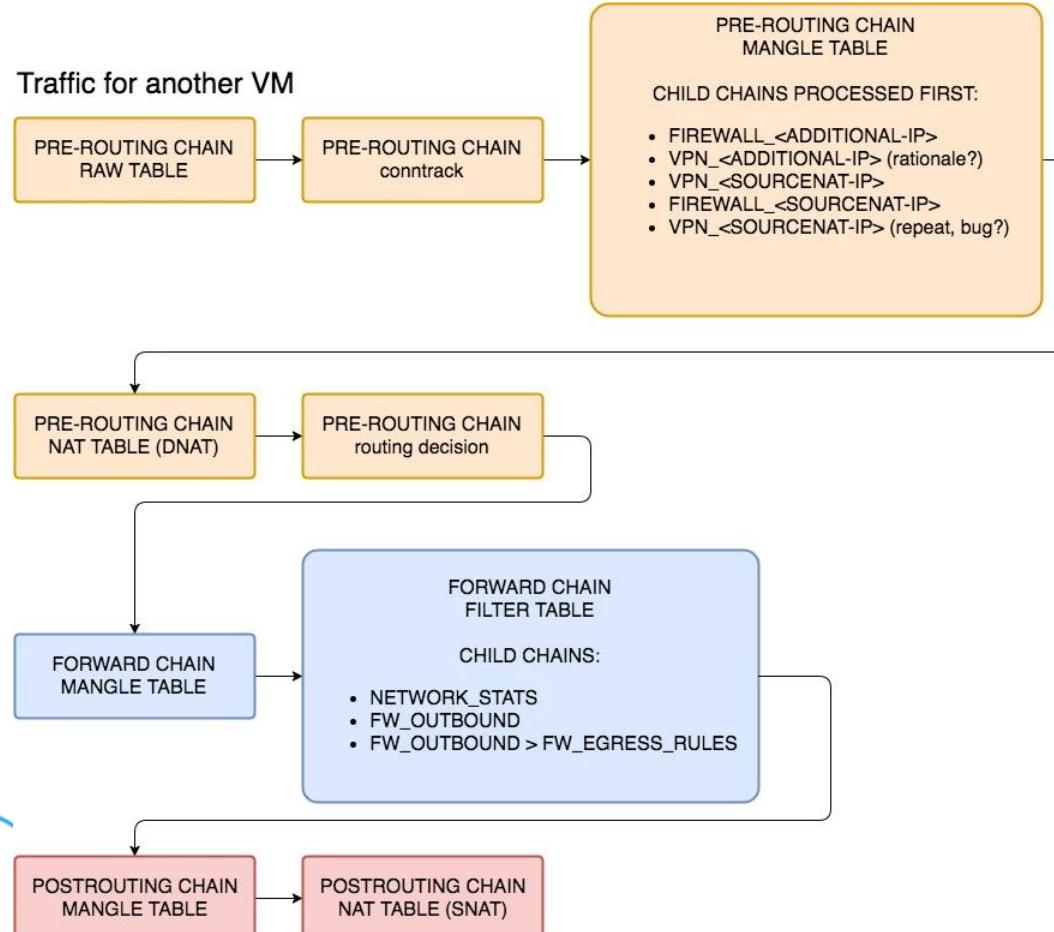
- Other NF parts
- Other Networking
- basic set of filtering opportunities at the
- network level
- bridge level



Traffic for the local VR

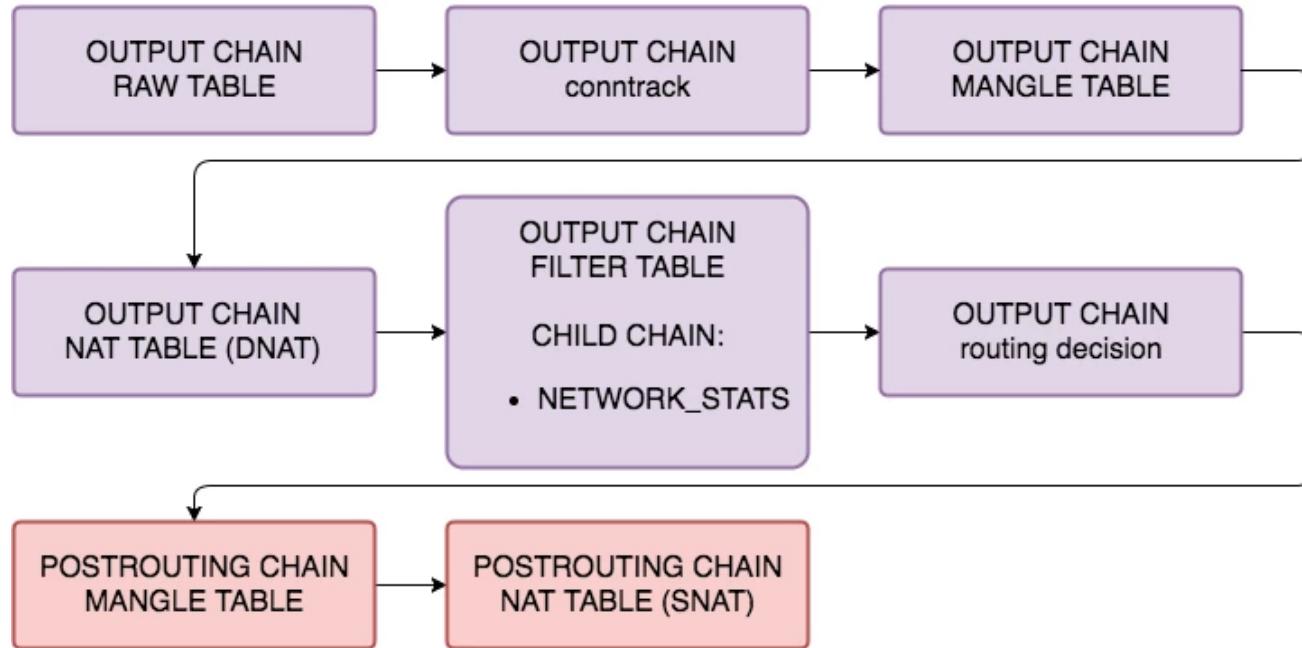


Traffic for another VM



Traffic originating from the VR

Traffic originating from VR



Iptables targets – from the man page

- **Targets determine what to do with a matched packet:**
 - A firewall rule specifies criteria for a packet and a target.
 - If the packet does not match, the next rule in the chain is examined;
 - if it does match, then the next rule is specified by the value of the target, which can be the name of a **user-defined chain** or one of the special values **ACCEPT**, **DROP**, **QUEUE** or **RETURN**:
 - **ACCEPT** means to let the packet through.
 - **DROP** means to drop the packet on the floor.
 - **QUEUE** means to pass the packet to userspace.
 - **RETURN** means stop traversing this chain and resume at the next rule in the previous (calling) chain. If the end of a built-in chain is reached or a rule in a built-in chain with target **RETURN** is matched, the target specified by the chain policy determines the fate of the packet.



The Cloud Specialists

ShapeBlue.com

@ShapeBlue

Iptables table policies

- **Each IPtable table chain also has a default policy which is applied after all rules have been evaluated**
- **On the right:**
 - Anything that isn't accepted during processing in the INPUT chain is dropped
 - Anything that isn't dealt with by the OUTPUT chain is ACCEPTed to continue traversing tables and chains

```
*filter
:INPUT DROP [0:0]
:FORWARD DROP [0:0]
:OUTPUT ACCEPT [987:121726]
:FW_EGRESS_RULES - [0:0]
:FW_OUTBOUND - [0:0]
:NETWORK_STATS - [0:0]
```



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

The anatomy of an Iptables rule

- **-A INPUT -d 10.1.1.1/32 -i eth0 -p tcp -m tcp --dport 443 -m state --state NEW -j ACCEPT**
 - -A: append rule
 - INPUT: chain
 - -d: destination IP range
 - -i: input network interface
 - -p: protocol
 - --dport: destination port
 - --state: state of connection
 - -j: target
- **Translates to: accept new packets destined for 10.1.1.1/32 over port TCP/443, in other words HTTPS traffic to the Virtual Router**

The anatomy of an Iptables rule

- **-A PREROUTING -d 10.1.35.225/32 -j FIREWALL_10.1.35.225**
- **-A FIREWALL_10.1.35.225 -m state --state RELATED,ESTABLISHED -j ACCEPT**
- **-A FIREWALL_10.1.35.225 -j DROP**
- **Translates to:**
 - In prerouting chain, packets destined for 10.1.35.225/32 should be redirected to custom chain FIREWALL_10.1.35.225
 - In FIREWALL_10.1.35.225 all already established connections should be accepted
 - All other packets should be dropped



The Cloud Specialists

ShapeBlue.com

 @ShapeBlue

Working with iptables

- `iptables -t nat|mangle|raw|filter -vnL:`
 - outputs each table in (v)erbose, (n)umeric and (L)ist form
 - Also captures packet counters – used for troubleshooting
- `iptables-save:`
 - outputs all tables in format that can be edited and later `iptables-restore'd`
 - Does not capture any statistics



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Working with Iptables

- “**iptables-save > output_file**” is an easy way to capture all current iptables rules in a reusable format
- These can be edited with an editor and later restored with “**iptables-restore < output_file**”
- This is an easy way to edit and update rules on the fly without having to go back to the CloudStack GUI

```
root@6-VM:~# iptables-save
# Generated by iptables-save v1.6.1 on Thu Apr 19 18:44:56 2018
*raw
:PREROUTING ACCEPT [13303:1089582]
:OUTPUT ACCEPT [4521:1970080]
-A PREROUTING -p icmp -j TRACE
-A OUTPUT -p icmp -j TRACE
COMMIT
# Completed on Thu Apr 19 18:44:56 2018
# Generated by iptables-save v1.6.1 on Thu Apr 19 18:44:56 2018
*nat
:PREROUTING ACCEPT [12064:1011049]
:INPUT ACCEPT [136:9961]
:OUTPUT ACCEPT [883:65408]
:POSTROUTING ACCEPT [3:508]
-A POSTROUTING -o eth2 -j SNAT --to-source 10.1.35.225
COMMIT
# Completed on Thu Apr 19 18:44:56 2018
# Generated by iptables-save v1.6.1 on Thu Apr 19 18:44:56 2018
*mangle
:PREROUTING ACCEPT [17111:1435413]
:INPUT ACCEPT [5540:480843]
:FORWARD ACCEPT [11928:1001088]
:OUTPUT ACCEPT [5634:2105867]
:POSTROUTING ACCEPT [5634:2105867]
:FIREWALL_10.1.35.225 - [0:0]
:VPN_10.1.35.225 - [0:0]
-A PREROUTING -d 10.1.35.225/32 -j FIREWALL_10.1.35.225
-A PREROUTING -d 10.1.35.225/32 -j VPN_10.1.35.225
-A PREROUTING -m state --state RELATED,ESTABLISHED -j CONNMARK --restore-mark --nfmask 0xffffffff --ctmask 0xffffffff
-A PREROUTING -i eth2 -m state --state NEW -j CONNMARK --set-xmark 0x2/0xffffffff
-A POSTROUTING -p udp -m udp --dport 68 -j CHECKSUM --checksum-fill
-A FIREWALL_10.1.35.225 -m state --state RELATED,ESTABLISHED -j ACCEPT
-A FIREWALL_10.1.35.225 -j DROP
-A VPN_10.1.35.225 -m state --state RELATED,ESTABLISHED -j ACCEPT
-A VPN_10.1.35.225 -j RETURN
COMMIT
# Completed on Thu Apr 19 18:44:56 2018
# Generated by iptables-save v1.6.1 on Thu Apr 19 18:44:56 2018
*filter
:INPUT DROP [0:0]
:FORWARD DROP [0:0]
:OUTPUT ACCEPT [5665:2109403]
:FW_EGRESS_RULES - [0:0]
:FW_OUTBOUND - [0:0]
:NETWORK_STATS - [0:0]
-A INPUT -d 10.1.1.1/32 -i eth0 -p tcp --dport 443 -m state --state NEW -j ACCEPT
-A INPUT -d 10.1.1.1/32 -i eth0 -p tcp --dport 80 -m state --state NEW -j ACCEPT
-A INPUT -d 10.1.1.1/32 -i eth0 -p tcp --dport 53 -j ACCEPT
-A INPUT -d 10.1.1.1/32 -i eth0 -p udp -m udp --dport 53 -j ACCEPT
-A INPUT -i eth0 -p udp -m udp --dport 67 -j ACCEPT
-A INPUT -j NETWORK_STATS
-A INPUT -d 224.0.0.18/32 -j ACCEPT
-A INPUT -d 225.0.0.50/32 -j ACCEPT
-A INPUT -i eth2 -m state --state RELATED,ESTABLISHED -i ACCEPT
```



Tracing iptables processing

- Capturing which iptables rules packets traverse through can be tricky. For this purpose we utilize the TRACE target.
- By appending rules into the RAW table / PREROUTING and OUTPUT chains we can trace which rules packets traverse. This captures all packet paths.
- ***Warning: this is verbose and likely to fill your /var/log/kern.log and /var/log/messages – hence switch it off after use!***
- E.g. to trace all ICMP traffic on all interfaces:
 - # iptables -t raw -A OUTPUT -p icmp -j TRACE
 - # iptables -t raw -A PREROUTING -p icmp -j TRACE
- Turn tracing off with “-D”:
 - # iptables -t raw -D OUTPUT -p icmp -j TRACE
 - # iptables -t raw -D PREROUTING -p icmp -j TRACE



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

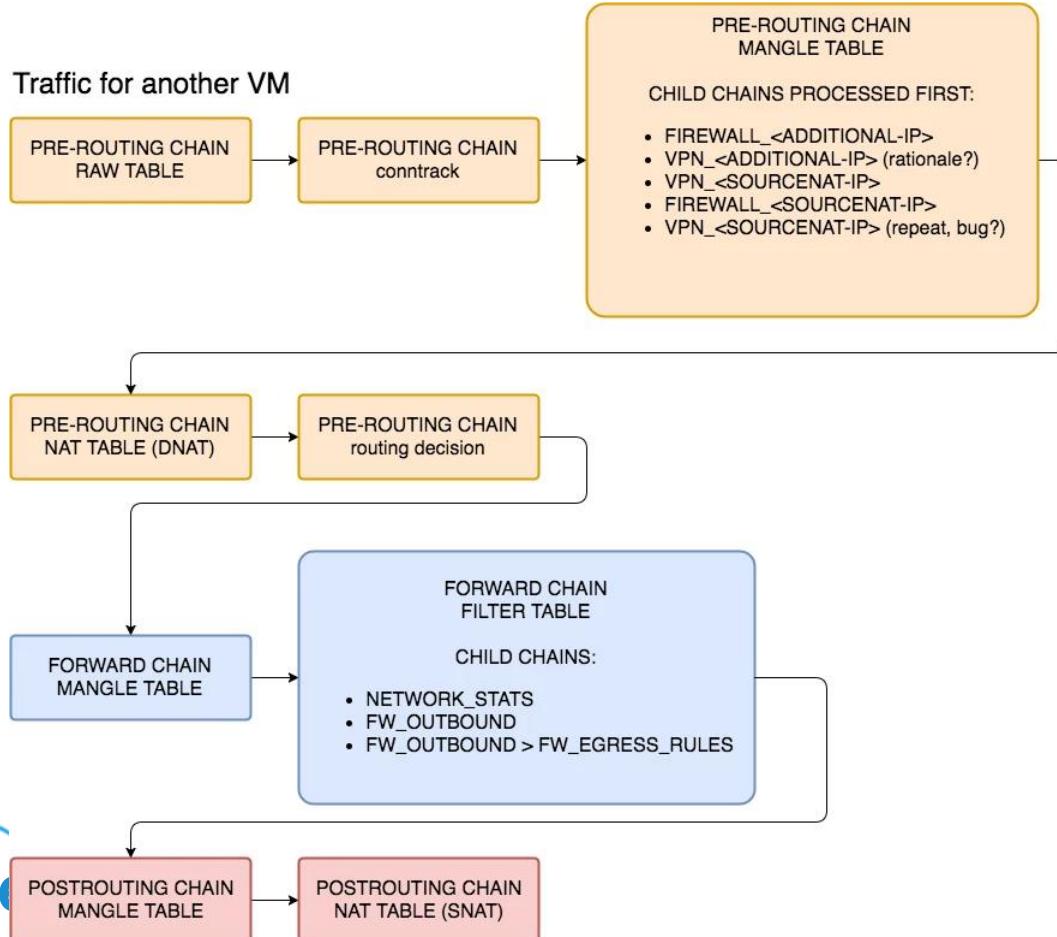
Tracing Iptables processing

```
root@r-6-VM:~# tail -f /var/log/kern.log
Apr 19 18:31:31 r-6-VM kernel: [11809.879940] TRACE: raw:PREROUTING:policy:2 IN=eth0 OUT= MAC=02:00:1c:7f:00:02:02:00:02:c6:00:01:08:00 SRC=10.1.1.91 DST=8.8.8.8 LEN=84 TO S=0x00 PREC=0x00 TTL=64 ID=0 DF PROTO=ICMP TYPE=8 CODE=0 ID=8456 SEQ=11087
Apr 19 18:31:31 r-6-VM kernel: [11809.879974] TRACE: mangle:PREROUTING:policy:5 IN=eth0 OUT= MAC=02:00:1c:7f:00:02:02:00:02:c6:00:01:08:00 SRC=10.1.1.91 DST=8.8.8.8 LEN=84 TOS=0x00 PREC=0x00 TTL=64 ID=0 DF PROTO=ICMP TYPE=8 CODE=0 ID=8456 SEQ=11087
Apr 19 18:31:31 r-6-VM kernel: [11809.879987] TRACE: nat:PREROUTING:policy:1 IN=eth0 OUT= MAC=02:00:1c:7f:00:02:02:00:02:c6:00:01:08:00 SRC=10.1.1.91 DST=8.8.8.8 LEN=84 TO S=0x00 PREC=0x00 TTL=64 ID=0 DF PROTO=ICMP TYPE=8 CODE=0 ID=8456 SEQ=11087
Apr 19 18:31:31 r-6-VM kernel: [11809.880016] TRACE: mangle:FORWARD:policy:1 IN=eth0 OUT=eth2 MAC=02:00:1c:7f:00:02:02:00:02:c6:00:01:08:00 SRC=10.1.1.91 DST=8.8.8.8 LEN=84 TOS=0x00 PREC=0x00 TTL=63 ID=0 DF PROTO=ICMP TYPE=8 CODE=0 ID=8456 SEQ=11087
Apr 19 18:31:31 r-6-VM kernel: [11809.880028] TRACE: filter:FORWARD:rule:1 IN=eth0 OUT=eth2 MAC=02:00:1c:7f:00:02:02:00:02:c6:00:01:08:00 SRC=10.1.1.91 DST=8.8.8.8 LEN=84 TOS=0x00 PREC=0x00 TTL=63 ID=0 DF PROTO=ICMP TYPE=8 CODE=0 ID=8456 SEQ=11087
```

- The /var/log/kern.log file will now show all the TRACE packet traversal steps
- Above we can see:
 - RAW table / PREROUTING chain – policy rule 2
 - MANGLE table / PREROUTING chain – policy rule 5
 - NAT table / PREROUTING chain – policy rule 1
 - MANGLE table / FORWARD chain – policy rule 1
- In addition we see input and output interfaces, source and destination, protocol etc

Tracing Iptables processing

- Based on the flow of traffic from the TRACE log we know this is traffic originating outside the VR and destined to another host



Tracing Iptables processing

- In addition to the TRACE target we can use the packet counters to determine which rules receive / accept / drop packets.
- For this we utilise “`iptables -t <tablename>-vnL`”, i.e. this has to be done on a per table basis – but the filter table is often a good starting point.
- In addition we can utilise the linux “`watch -d`” command, which runs the same command every 2 seconds and highlights changes between screens. This makes it easy to spot changing packet counts.



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Tracing Iptables processing

- This obviously works better when watching the results in the console.....
- However we can spot that we have a large and changing packet count against the DROP target in the FW_EGRESS_RULES chain.
- This give a good indication that's the rule dropping the traffic.

```
Every 2.0s: iptables -t filter -vnL

Chain INPUT (policy DROP 0 packets, 0 bytes)
pkts bytes target    prot opt in   out    source          destination
  0     0  ACCEPT    tcp   --  eth0  *      0.0.0.0/0      10.1.1.1
  0     0  ACCEPT    tcp   --  eth0  *      0.0.0.0/0      10.1.1.1
  0     0  ACCEPT    tcp   --  eth0  *      0.0.0.0/0      10.1.1.1
 16    1149  ACCEPT   udp   --  eth0  *      0.0.0.0/0      10.1.1.1
  2    656  ACCEPT   udp   --  eth0  *      0.0.0.0/0      0.0.0.0/0
5264  464K  NETWORK_STATS all   --  *      *      0.0.0.0/0      0.0.0.0/0
  0     0  ACCEPT    all   --  *      *      0.0.0.0/0      224.0.0.18
  0     0  ACCEPT    all   --  *      *      0.0.0.0/0      225.0.0.50
347  45196  ACCEPT   all   --  eth2  *      0.0.0.0/0      0.0.0.0/0
51    4284  ACCEPT   icmp  --  *      *      0.0.0.0/0      0.0.0.0/0
19    1552  ACCEPT   all   --  lo    *      0.0.0.0/0      0.0.0.0/0
4847  413K  ACCEPT   tcp   --  eth1  *      0.0.0.0/0      0.0.0.0/0
  0     0  ACCEPT    all   --  eth0  *      0.0.0.0/0      0.0.0.0/0
  0     0  ACCEPT    udp   --  eth0  *      0.0.0.0/0      0.0.0.0/0
  0     0  ACCEPT    udp   --  eth0  *      10.1.1.0/24    0.0.0.0/0
  0     0  ACCEPT    tcp   --  eth0  *      10.1.1.0/24    0.0.0.0/0
  0     0  ACCEPT    tcp   --  eth0  *      0.0.0.0/0      0.0.0.0/0
  0     0  ACCEPT    tcp   --  eth0  *      0.0.0.0/0      0.0.0.0/0
Chain FORWARD (policy DROP 0 packets, 0 bytes)
pkts bytes target    prot opt in   out    source          destination
11627  976K  NETWORK_STATS all   --  *      *      0.0.0.0/0      0.0.0.0/0
  0     0  ACCEPT    all   --  eth2  eth0  0.0.0.0/0      0.0.0.0/0
11627  976K  FW_OUTBOUND all   --  eth0  eth2  0.0.0.0/0      0.0.0.0/0
  0     0  ACCEPT    all   --  eth0  eth1  0.0.0.0/0      0.0.0.0/0
  0     0  ACCEPT    all   --  eth0  eth0  0.0.0.0/0      0.0.0.0/0
  0     0  ACCEPT    all   --  eth0  eth0  0.0.0.0/0      0.0.0.0/0
Chain OUTPUT (policy ACCEPT 5385 packets, 2056K bytes)
pkts bytes target    prot opt in   out    source          destination
5385  2056K  NETWORK_STATS all   --  *      *      0.0.0.0/0      0.0.0.0/0
Chain FW_EGRESS_RULES (1 references)
pkts bytes target    prot opt in   out    source          destination
11627  976K  DROP    all   --  *      *      0.0.0.0/0      0.0.0.0/0
Chain FW_OUTBOUND (1 references)
pkts bytes target    prot opt in   out    source          destination
  0     0  ACCEPT    all   --  *      *      0.0.0.0/0      0.0.0.0/0
11627  976K  FW_EGRESS_RULES all   --  *      *      0.0.0.0/0      0.0.0.0/0
Chain NETWORK_STATS (3 references)
pkts bytes target    prot opt in   out    source          destination
11627  976K           all   --  eth0  eth2  0.0.0.0/0      0.0.0.0/0
  0     0           all   --  eth2  eth0  0.0.0.0/0      0.0.0.0/0
  0     0           tcp   -- !eth0 !eth2  0.0.0.0/0      0.0.0.0/0
  0     0           tcp   -- !eth2 !eth0  0.0.0.0/0      0.0.0.0/0
```



Tcpdump is your friend

- Tcpdump is a packet sniffer which allows you to check inbound / outbound traffic on network interfaces.**
- In this example we see inbound ICMP ping packets come in on the VR eth0 private interface, but we do not see the traffic traverse out on the public eth2 interface.**
- We can therefore conclude packets are dropped – refer to previous slides for how to troubleshoot iptables packets.**

```
root@r-6-VM:~# ifconfig | egrep '^eth[linet]\|^lo'
eth0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
      inet 10.1.1.1  netmask 255.255.255.0 broadcast 10.1.1.255
eth1: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
      inet 169.254.0.131  netmask 255.255.0.0 broadcast 169.254.255.255
eth2: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
      inet 10.1.35.225  netmask 255.255.224.0 broadcast 10.1.63.255
lo: flags=73<UP,LOOPBACK,RUNNING>  mtu 65536
      inet 127.0.0.1  netmask 255.0.0.0
root@r-6-VM:~#
root@r-6-VM:~#
root@r-6-VM:~#
root@r-6-VM:~# tcpdump -i eth0 icmp
tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
listening on eth0, link-type EN10MB (Ethernet), capture size 262144 bytes
18:55:45.670781 IP sbconisolated > google-public-dns-a.google.com: ICMP echo request, id 8456, seq 12541, length 64
18:55:46.670870 IP sbconisolated > google-public-dns-a.google.com: ICMP echo request, id 8456, seq 12542, length 64
18:55:47.670822 IP sbconisolated > google-public-dns-a.google.com: ICMP echo request, id 8456, seq 12543, length 64
^C
3 packets captured
3 packets received by filter
0 packets dropped by kernel
root@r-6-VM:~#
root@r-6-VM:~#
root@r-6-VM:~#
```

```
root@r-6-VM:~# tcpdump -i eth2 icmp
tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
listening on eth2, link-type EN10MB (Ethernet), capture size 262144 bytes
^C
0 packets captured
0 packets received by filter
0 packets dropped by kernel
root@r-6-VM:~#
```



Tcpdump is your friend

- So – from what you see here – who can tell my what is happening?
- What traffic is it we're watching being blocked?
- How would you fix it?
(hint – it's not something broken, it's a missing configuration)

```
Chain FW_EGRESS_RULES (1 references)
pkts bytes target      prot opt in     out      source          destination
13526 1135K DROP        all   --  *       *      0.0.0.0/0          0.0.0.0/0
```

```
root@r-6-VM:~# tcpdump -i eth0 icmp
tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
listening on eth0, link-type EN10MB (Ethernet), capture size 262144 bytes
19:12:15.678834 IP sbox1isolated > google-public-dns-a.google.com: ICMP echo request, id 8456, seq 13531, length 64
19:12:16.678778 IP sbox1isolated > google-public-dns-a.google.com: ICMP echo request, id 8456, seq 13532, length 64
^C
2 packets captured
2 packets received by filter
0 packets dropped by kernel
root@r-6-VM:~# tcpdump -i eth2 icmp
tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
listening on eth2, link-type EN10MB (Ethernet), capture size 262144 bytes
^C
0 packets captured
0 packets received by filter
0 packets dropped by kernel
root@r-6-VM:~#
```



The Cloud Specialists

ShapeBlue.com

@ShapeBlue

What have I not covered?

- **Etables work at the Ethernet frame level.**
- **We use etables for security groups in**
 - Basic zones
 - Advanced zones with security groups
- **Rohit will cover these in more detail on Thursday.**



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Network troubleshooting

- This was a high level introduction to some of the techniques we use to troubleshoot network issues in CloudStack
- Rohit will give you a more in-depth session later this week....

Network troubleshooting

**End of part 3 – any questions
apart from “who wants a beer?”**



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

References

- OSI model: https://en.wikipedia.org/wiki/OSI_model
- Teams vs bonds: https://access.redhat.com/documentation/en-us/red_hat_enterprise_linux/7/html/networking_guide/sec-comparison_of_network_taming_to_bonding
- Linux bond modes: <https://wiki.centos.org/TipsAndTricks/BondingInterfaces>
- MAC address vendor list: <http://standards-oui.ieee.org/oui.txt>
- IP subnet classes: <https://tools.ietf.org/html/rfc1918>
- SDN: <https://www.cisco.com/c/en/us/about/press/internet-protocol-journal/back-issues/table-contents-59/161-sdn.html>
- Iptables SB wiki article:
<https://shapeblue.atlassian.net/wiki/spaces/SUPPORT/pages/143687775/Troubleshooting+V+R+IPtables>



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

CloudStack original bootcamp (slides for reference)

Agenda

- What is CloudStack
- CloudStack architecture
- CloudStack networking models
- Using KVM, XenServer & VMware
- Adding zones, PODs, clusters
- System VMs
- Storage
- Service offerings
- Domains, accounts & users
- Projects
- Limits
- Notifications and thresholds
- Virtual machine allocation
- Virtual machine deployment
- Managing templates
- Snapshots
- Virtual Private Clouds
- Using the API
- CloudMonkey
- Troubleshooting
- Working with the databases
- The Apache CloudStack Community



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

IaaS and Cloud services



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

What is “the cloud”?

- **Cloud computing is a general term for anything that involves delivering hosted services over a network.**
- **These services are broadly divided into three categories:**
 - Infrastructure-as-a-Service (IaaS)
 - I want virtual servers all connected to each other via an internal network.
 - Platform-as-a-Service (PaaS)
 - I want to develop / run / manage my own web services / application services / databases without having to worry about infrastructure.
 - Software-as-a-Service (SaaS)
 - I want X number of mailboxes or SharePoint instances.



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

What is “the cloud”?

- **What makes these services ‘Cloudy’?**

- On-demand self-service
- Elasticity
- Scalability
- API integration
- Resource accounting



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Who use Cloud services?

- **Public clouds (SPs/MSPs)**
 - General public can create and manage their own cloud resources and are billed on a per resource/time basis.
 - E.g. AWS, Azure, Google Cloud, Rackspace, Digital Ocean.
- **Private Clouds (Enterprises)**
 - Anyone who wants to be able to orchestrate their own environment through APIs and self service portals.
- **Hybrid Clouds**
 - Cloud models where resources are shared between public / private / community consumption.
 - Utilising combinations of cloud service models or combination of cloud providers.



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Multiple Cloud Strategies

Private Cloud



- ❖ On premise or hosted
- ❖ Dedicated resources
- ❖ Security and higher degree of control
- ❖ SLA bound
- ❖ Internal network
- ❖ Managed by Enterprise or 3rd Party

Public Cloud



- ❖ Mix of shared and dedicated resources
- ❖ Elastic scaling
- ❖ Pay as you go
- ❖ Public internet, VPN access

Introduction to CloudStack



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

What is CloudStack?

- **Secure multi-tenant Cloud Orchestration Platform**
 - Turnkey platform for delivering IaaS clouds.
 - Hypervisor agnostic.
 - Scalable, flexible secure and open.
 - Open source, open standards.
 - Can be used for private or public cloud offerings.



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Historic background

- Launched in May 2010 as Vmops, rebranded to Cloud.com.
- Acquired by Citrix in July 2011. Donated to the Apache Software Foundation in April 2012 and continue being developed as an open source Apache Software Foundation project on the Apache License v2.
- Citrix maintained their own commercial fork of the open source project, which they rebranded Citrix CloudPlatform.
- In January 2016 Citrix announced they were selling off CloudPlatform to Accelerite, which is owned by Persistent Systems group.
- Development of Apache CloudStack continue independently in an active open source community.
- Version at time of writing:
 - Apache CloudStack 4.11.0 released March 2018, with version 4.11.1 and 4.12 in the pipeline.
 - Accelerite CloudPlatform latest release 4.9.

Apache CloudStack vs OpenStack

- **OpenStack relies on a number of components:**
 - Nova: manages compute resources
 - Swift: object storage
 - Cinder: block storage
 - Neutron: networking
 - Horizon: dashboard
 - Keystone: identity service
 - Glance: image service
 - Ironic – bare metal provisioning
- **CloudStack in comparison is a converged solution – with most of these services managed by a single, stable and scalable platform.**



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Apache CloudStack – an open flexible platform

Compute



XenServer VMware OVM KVM Hyper-V UCS Bare metal

Compute primary storage



Local Disk iSCSI Fibre Channel NFS Ceph

User accessible secondary storage



NFS Swift S3

Apache CloudStack – an open flexible platform

Network



Isolation

Advanced – L2

Basic – L3

Services

Routing

Firewall

DHCP

DNS

LB

GSLB

VPN

What can you do with CloudStack?

- **Self service of all resources – compute, storage and networking – with no requirements for highly skilled technical staff.**
- **Automation of all provisioning and management through API.**
- **E.g.**
 - Create Virtual Machines from templates or ISOs
 - All Virtual Machine lifecycle actions: start/stop/delete/storage/networking
 - Create Isolated, Shared and Multi-Tiered Networks
 - Manage firewall and port forwarding rules
 - Manage Network Services such as Load Balancing, Static and Source NAT, VPNs, Global Load Balancing and Autoscaling



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

CloudStack Architecture



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

- **Hierarchical structure enables massive scale.**
 - Region
 - A grouping of Availability Zones within a geographic area
 - Dedicated Management Server infrastructure to manage the Region and all of its Zones
 - Availability Zone
 - Typically one Zone per DC
 - Contains at least 1 POD, 1 Cluster and Secondary Storage
 - Network scope for advanced zones using L2 networking / VLANs.

CloudStack architecture

- **Pod**

- Logical entity, typically a rack containing one or more clusters and networking
- Uses concept of something shared i.e. switch stack or storage array
- Network scope (broadcast domain) for L3 networks in basic zones.

- **Cluster**

- Group of identical hosts running a common hypervisor
- Primary storage



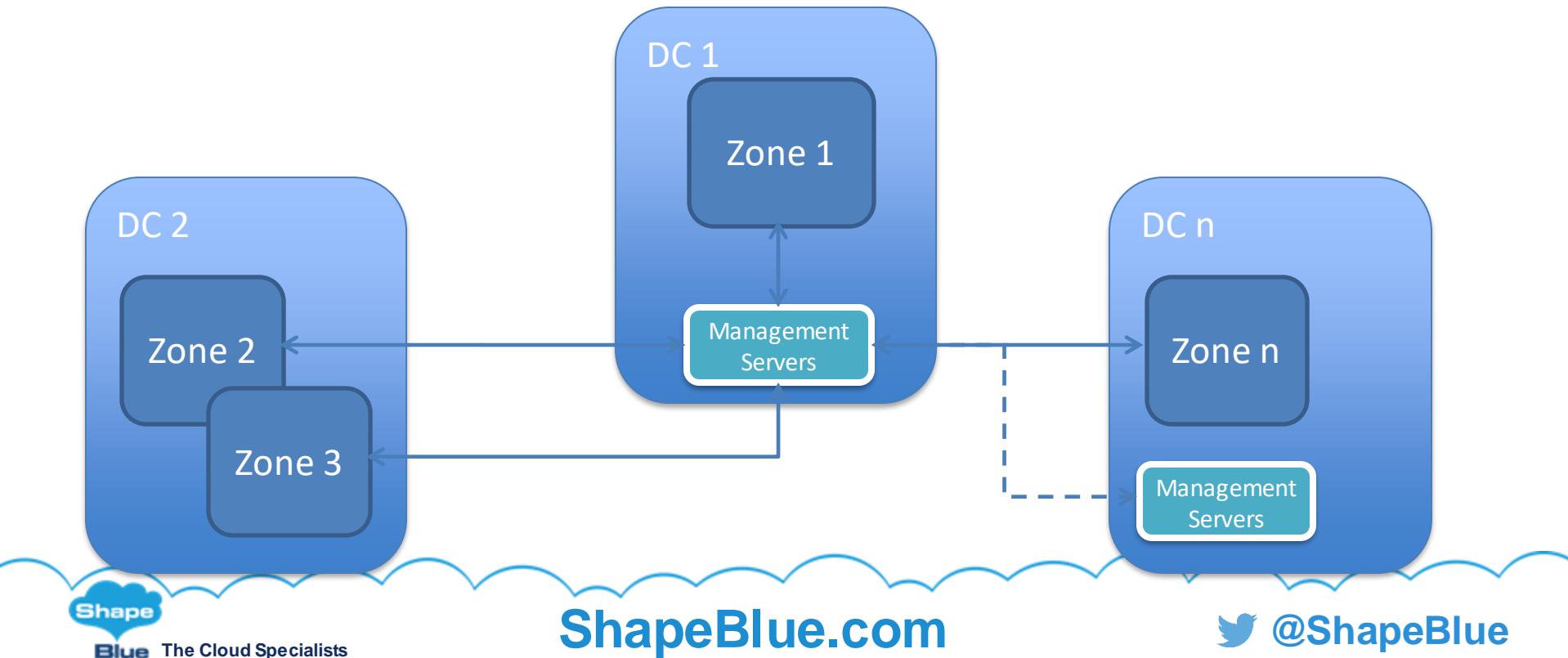
The Cloud Specialists

ShapeBlue.com



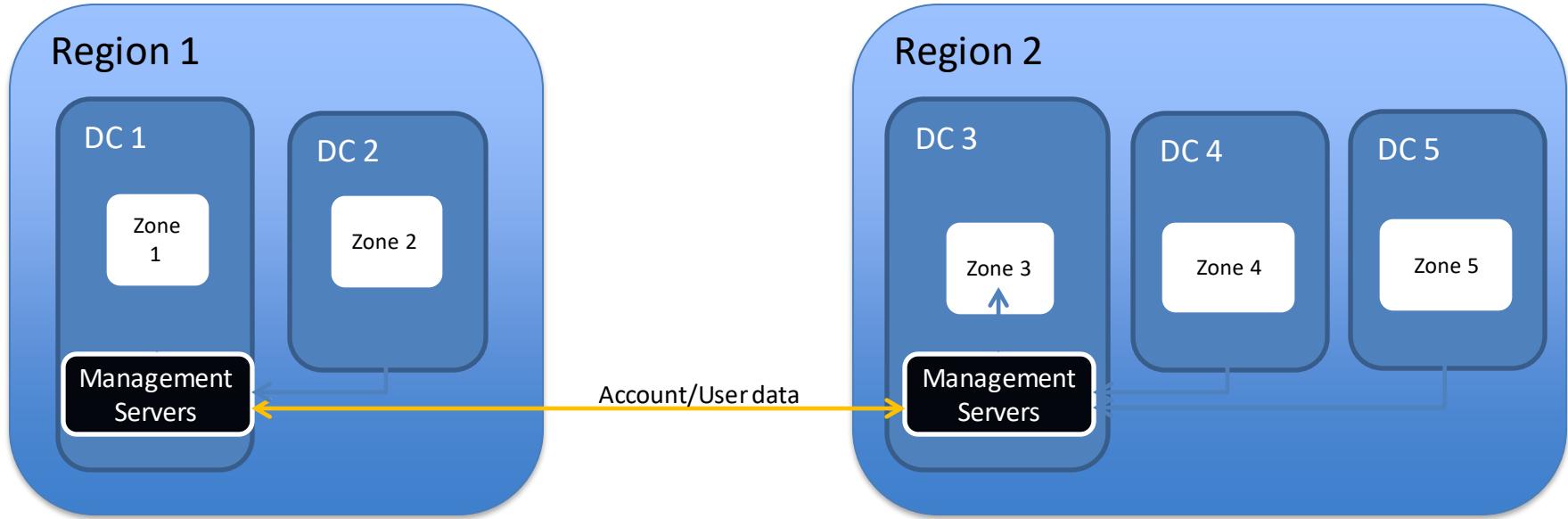
@ShapeBlue

Multiple Availability Zones within a Region

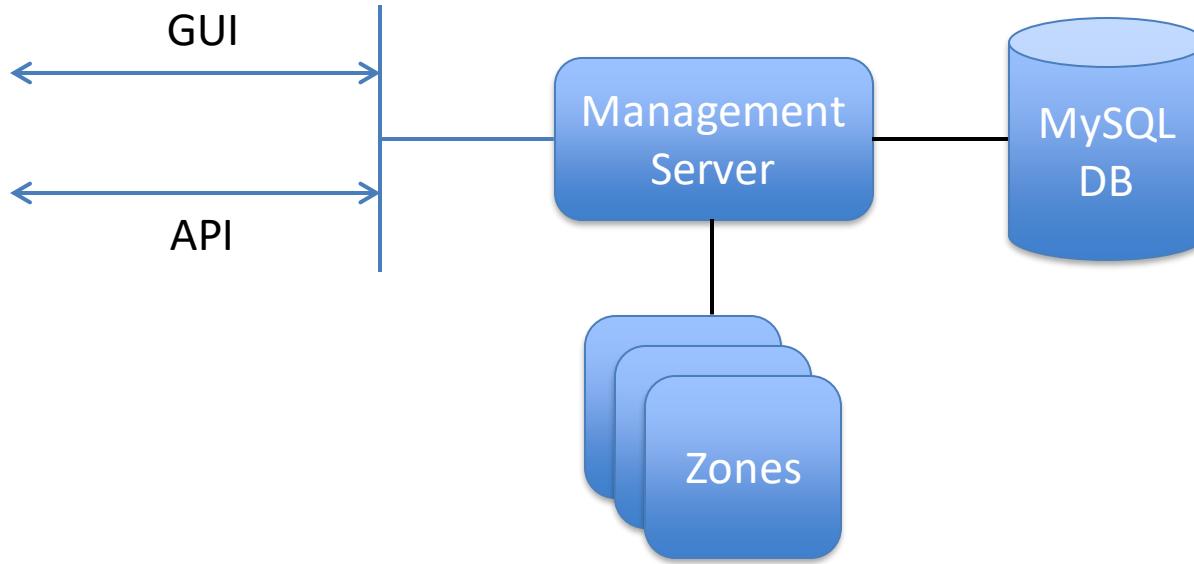


CloudStack Architecture

Multiple Availability Zones within a Region

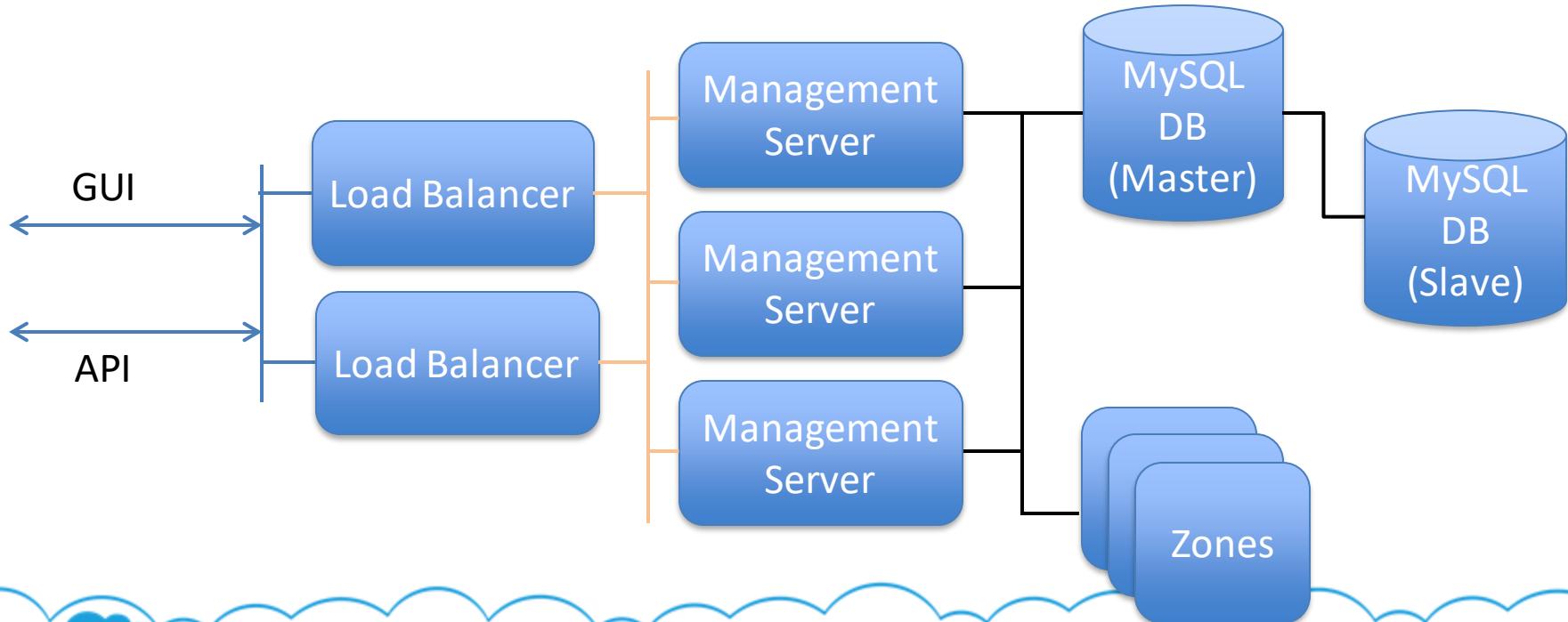


Single-Node Deployment



Management Server Deployment Architectures

- **Multi-Node Deployment**



- Primary Storage
 - Traditional storage backend as supported by each hypervisor.
 - Hosts the guest virtual machine disks and VM snapshots.
 - Traditionally unique to each cluster, but KVM and VMware now support zone-wide primary storage.
- Secondary Storage
 - Stores templates, ISOs and volume snapshots (backups).
 - Zone wide (region wide for S3).
 - NFS + S3 or NFS + Swift for region wide replication.

Questions?



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Lab environment



The Cloud Specialists

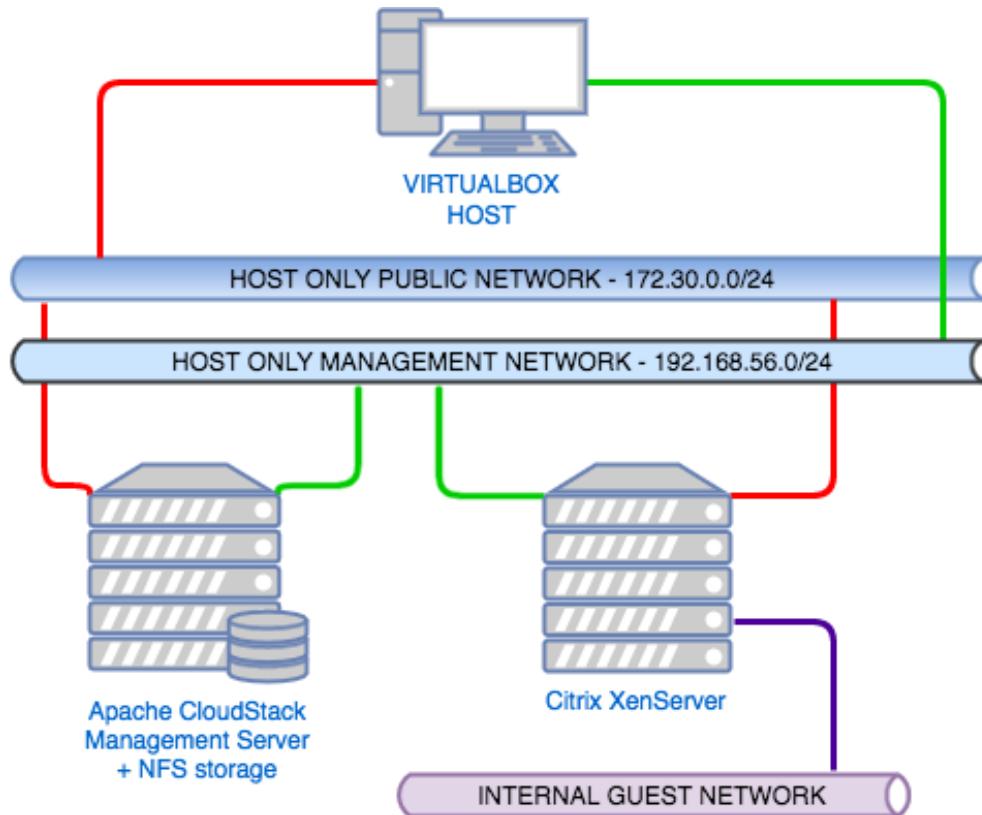
ShapeBlue.com



@ShapeBlue

- **Hands-on labs**
- **Virtual Box based appliances**
 - Preconfigured XenServer
 - Preinstalled Management Server

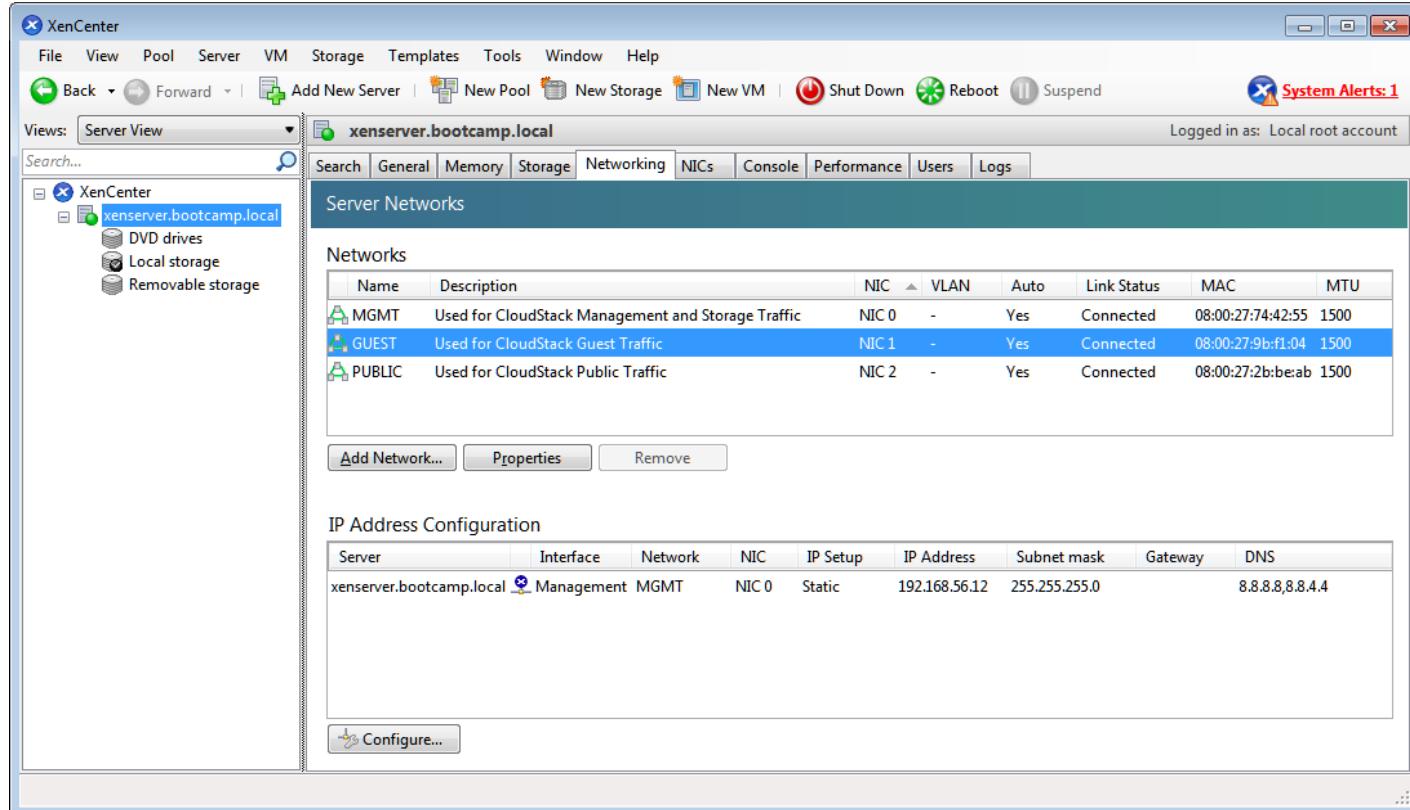
Lab VirtualBox VM Appliances



- **Basic Steps are:**

- yum install cloudstack-management mysql-server ntp
- Optimised MySQL server installation
- Install/setup NFS server
- Add firewall rules to allow NFS etc.
- Start services.
- Deploy databases.
- Seed system VM template to secondary storage.
- Optimise configuration for Bootcamp (reduce memory footprint).

XenServer networks



- ## Document Conventions

- Highlight section of interest 
- Where to 'click' 
- Sequence Number 2
- Text can be copied *Text*
- Additional info *Notes*

- Please follow guide very carefully!

Exercise 1: Import CloudStack Appliance



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

CloudStack networking



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

- **Logical Networking Models**
 - Basic
 - Advanced
 - Advanced with Security Groups

- **Basic zones**

- Guest isolation is provided through layer-3 security groups (IP address source filtering) by the hypervisor host.
- *Note this is only available on XenServer and KVM.*
- No VLANs.



- **Advanced zones:**

- Guest isolation is provided through layer-2 VLANs (or SDN technologies)
- This network model provides the most flexibility in defining guest networks and providing custom network offerings such as firewall, VPN, Load Balancer & VPC functionality.



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

- **Security Groups in Advanced Zones**

- Enables the deployment of multiple ‘Basic’ style networks which use security groups for isolation of VMs, but with each network isolated VLAN (or SDN).



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Physical Networks

- There are 4 ‘physical’ network types:
 - Management
 - Guest
 - Public
 - Storage
- Public and storage networks might not appear in all CloudStack deployments
- There may be multiple guest networks in an advanced zone.

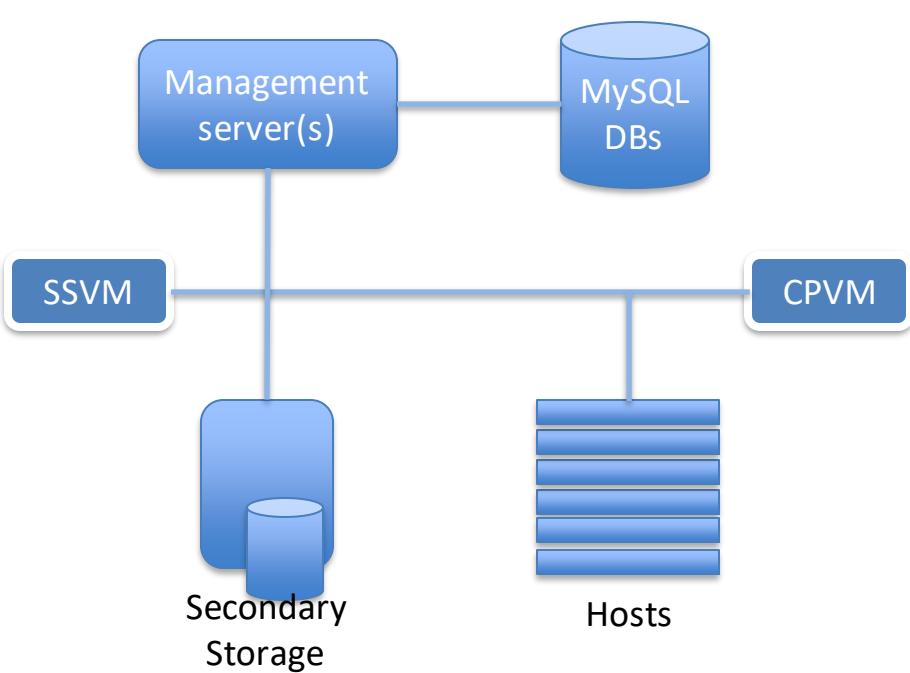


The Cloud Specialists

ShapeBlue.com

@ShapeBlue

Management network

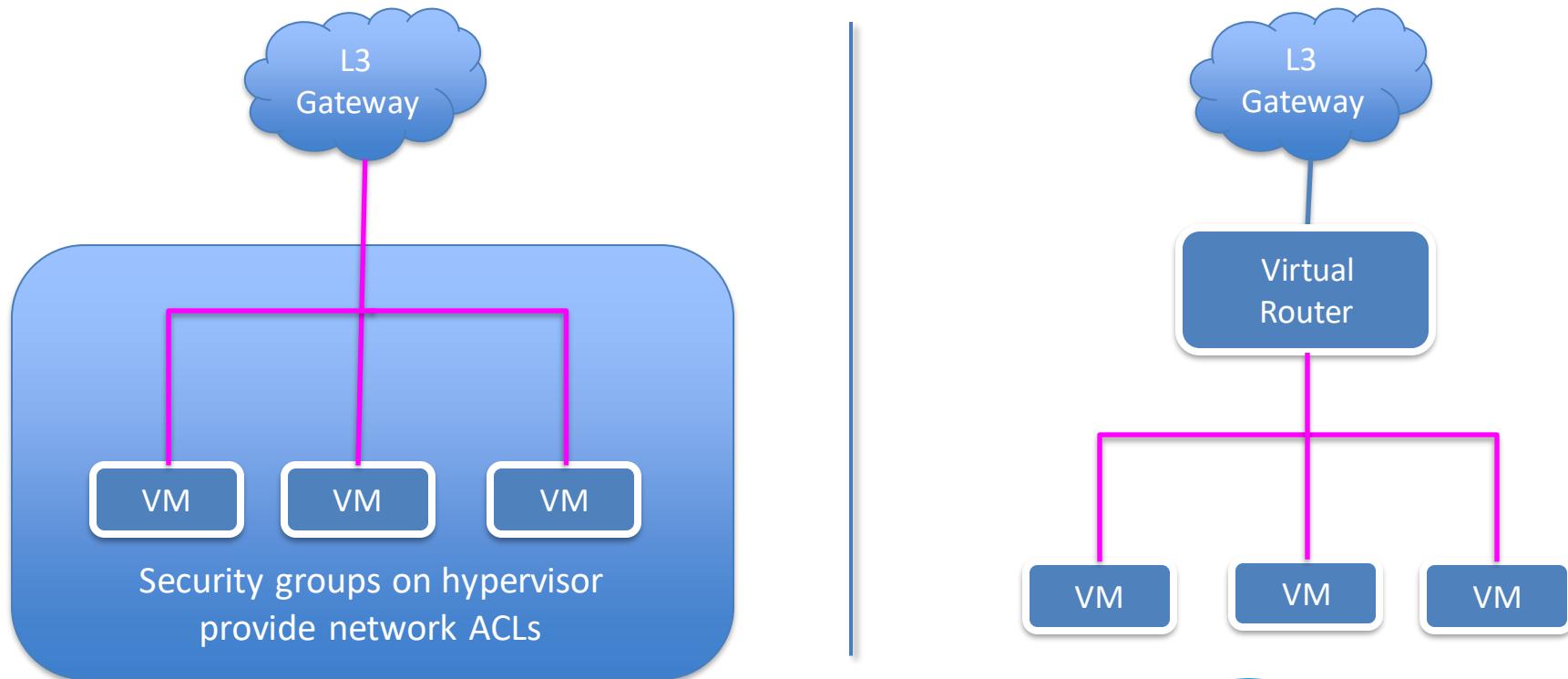


Traffic between CloudStack Management Servers and the various cloud components (Hosts, System VMs, Storage*, vCenter etc)

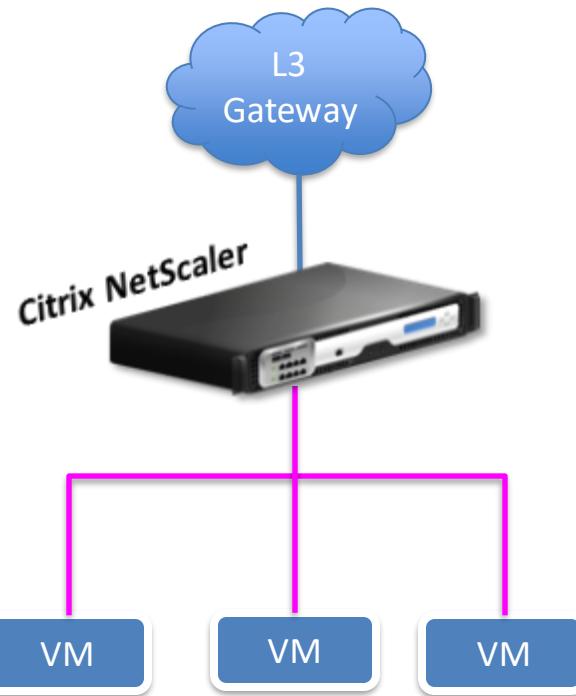
Management network

- **The management network is used for communication between the CloudStack management server and system VMs (Virtual Routers, CPVM and SSVM).**
- **XenServer and KVM assign link-local IP addresses (169.254.0.0/16) to the system VMs. This preserves IP address usage in the management subnet.**
- **VMware does not have this capability. As a result the CIDR used for management in VMware backed zones must have sufficient IP addresses to cover all hypervisor hosts as well as all customer Virtual Routers.**

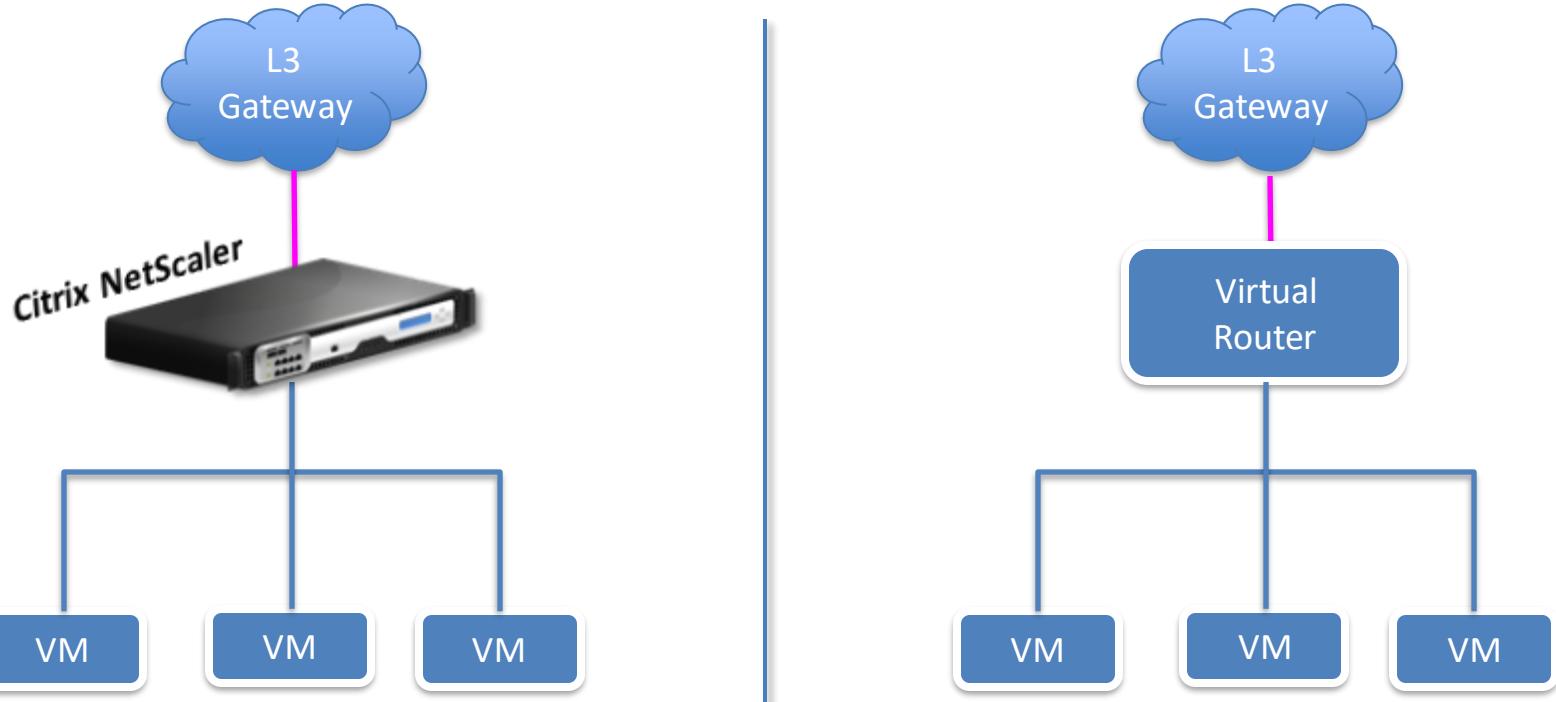
Guest Networks – Basic & Advanced



Guest Network – Basic Zone EIP / ELB

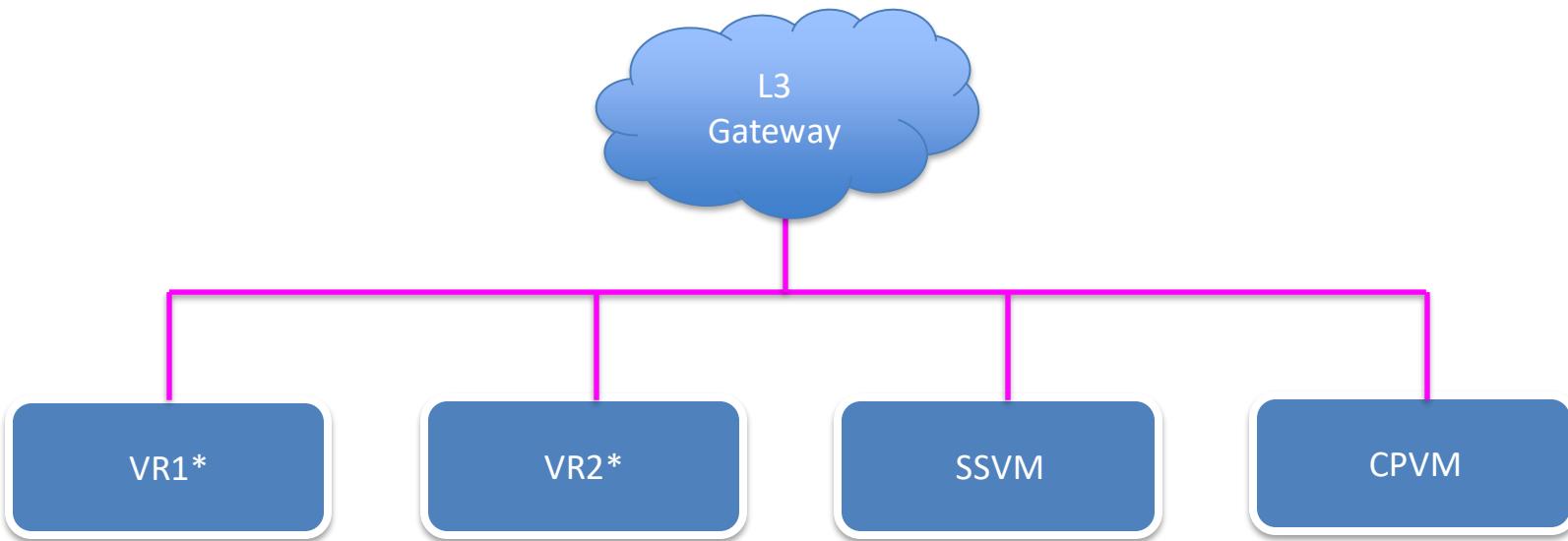


Public Network – Basic & Advanced



- **Public network ranges can be:**
 - Public internet exposed IP ranges.
 - Any other internal company wide private IP network.
- **Basic zone:**
 - End user VMs get assigned IP addresses in the public range (i.e. the guest IP range and the public IP range is the same).
 - When using Netscalers for EIP/ELB the public network is exposed on the northbound interface.
- **Advanced zone:**
 - Public network IP addresses are assigned to the public interface of the customer Virtual Router.

Public Network – System VMs

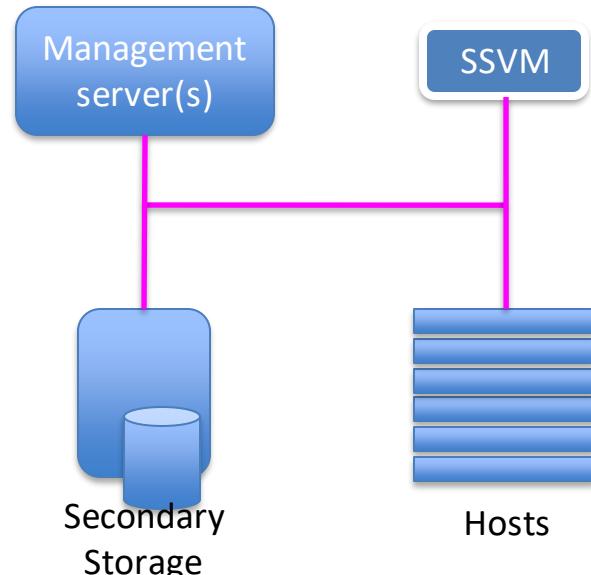


CPVM, SSVM & VRs have a connection to the Public Network

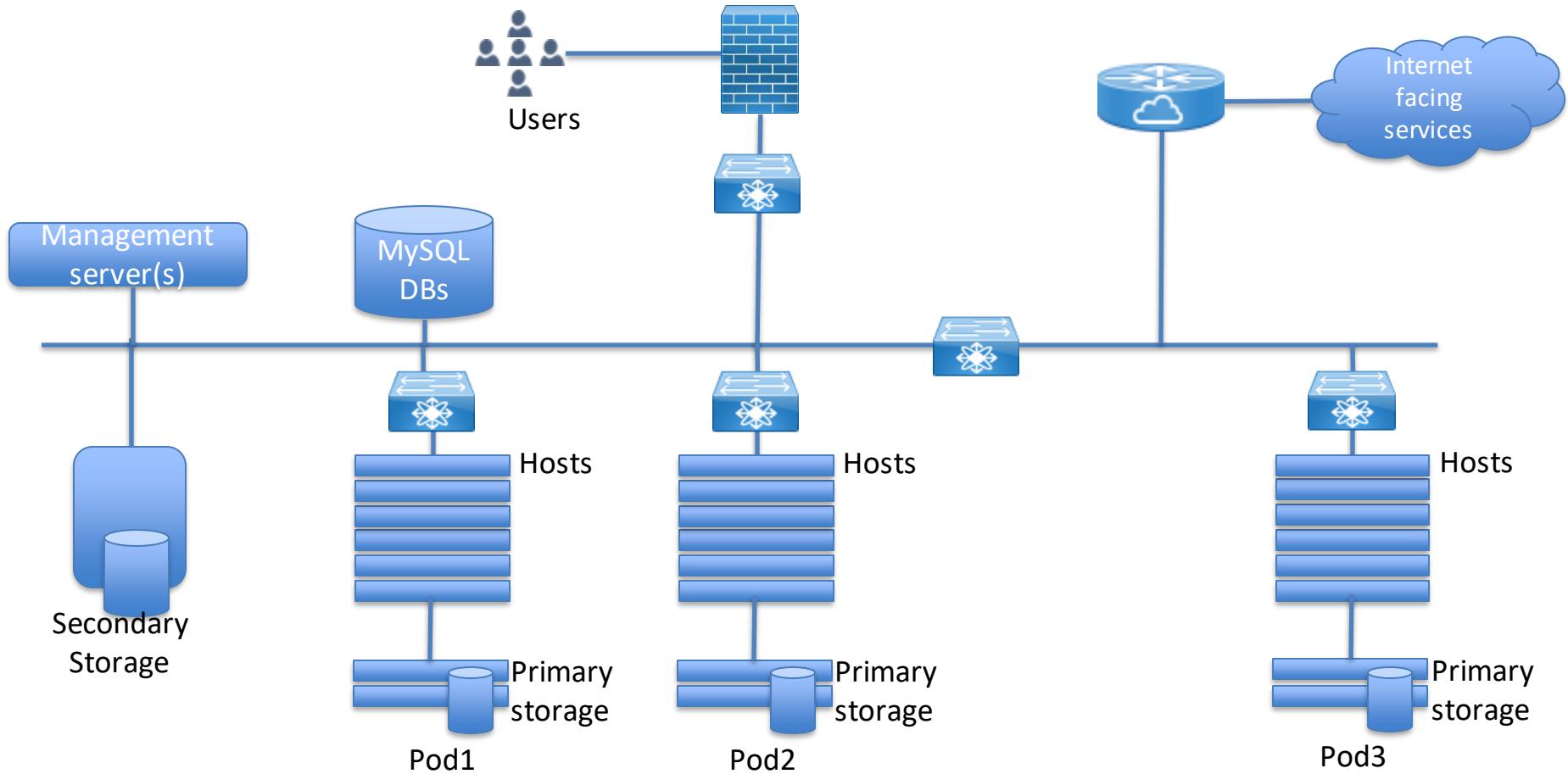
*VRs only have public connection in Advanced Network

Storage network

- Traffic between SSVM, secondary storage and hypervisors
- This is an optional network, traffic will use the management network if not configured.
- If configured, management server must still be able to communicate with ALL secondary storage pools to manage systemvm.iso and secondary storage housekeeping.
- It is NOT used for primary storage traffic.



Physical connectivity



Basic Networking



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Basic networking

- AWS Style L3 isolation – enables massive scale
- Simple routed network for each pod
- Each pod has a unique CIDR (broadcast domain)
- Optional guest isolation via security groups
- Optional NetScaler integration gives elastic IPs and elastic LB
- Optional VMware NSX integration



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Security Groups

- **Isolate traffic between VMs.**
- **Available in both Basic and Advanced zone networking models.**
- **Only supported on XenServer and KVM.**
- **XenServer must use Linux Bridge and not OpenvSwitch**
 - xe-switch-network-backend bridge
 - Edit sysctl to enable *net.bridge.bridge-nf-call-iptables* and *net.bridge.bridge-nf-call-arptables*
 - Must be implemented before adding to CloudStack



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Security Groups

- Must be switched on when the zone is created (Advanced zones). Always enabled in basic zones.
- Uses ingress and egress rules to control traffic flow.
- Default is all outbound traffic allowed, all inbound denied.
- Rules can be mapped to CIDR or another account/security group.

Protocol	Start Port	End Port	CIDR	Add
TCP				<button>Add</button>
TCP	80	80	0.0.0.0/0	<button>X</button>

Protocol	Start Port	End Port	Account, Security group	Add
TCP				<button>Add</button>
TCP	3306	3306	geoff - App...	<button>X</button>

Basic zone with Elastic IP

- Citrix NetScaler can provide Elastic IP & Elastic LB.
- Security groups are enabled.
- A public network IP range is assigned during zone setup.
- The public IP range is assigned to the external Interface of the NetScaler appliance.
- Provides a static NAT (1:1) service to VMs.
- When the VM is powered off the Elastic IP is released.



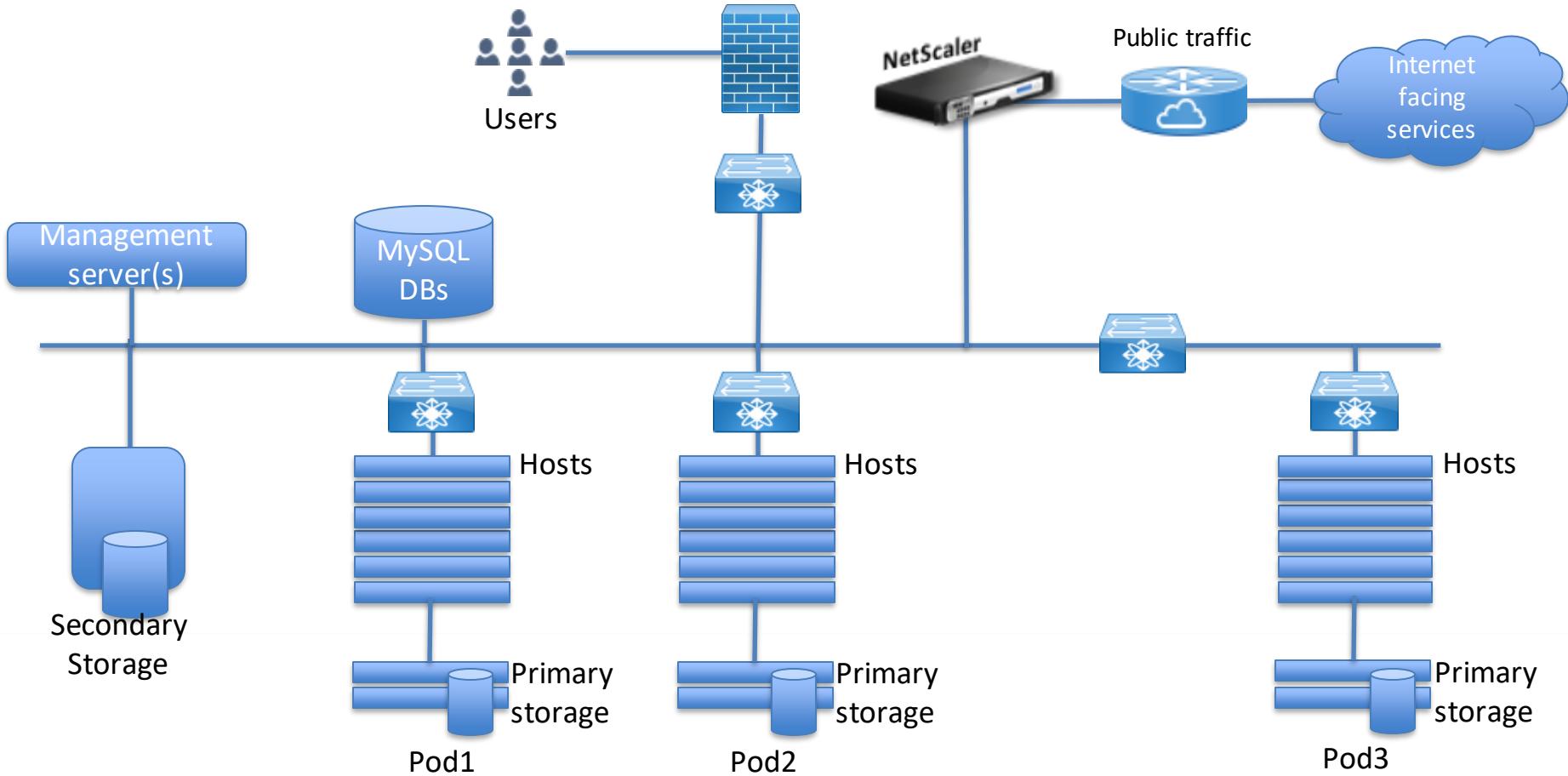
The Cloud Specialists

ShapeBlue.com

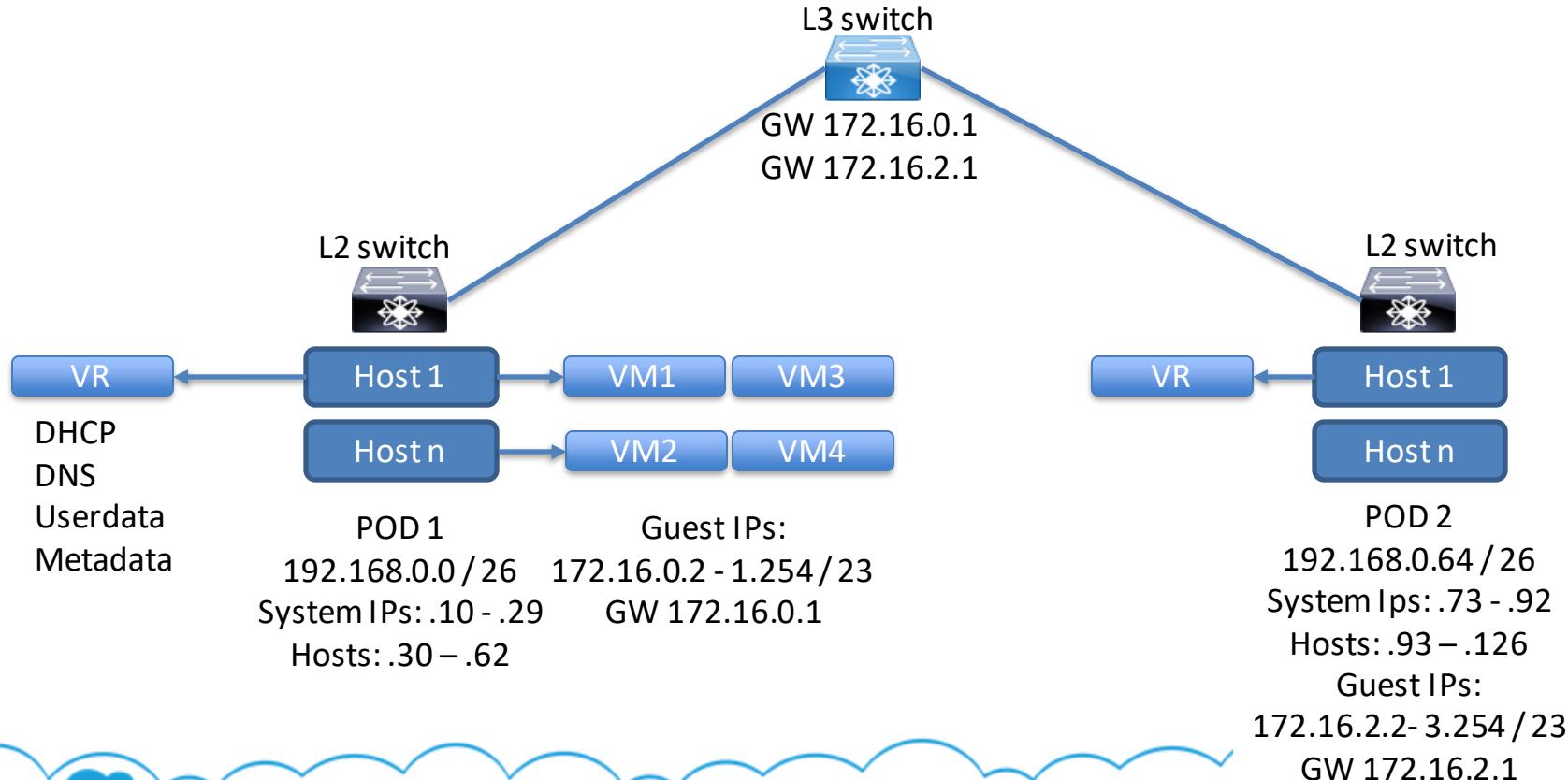


@ShapeBlue

Citrix NetScaler – Elastic IP/LB



Basic zone – example IP schema



Advanced Networking



The Cloud Specialists

ShapeBlue.com

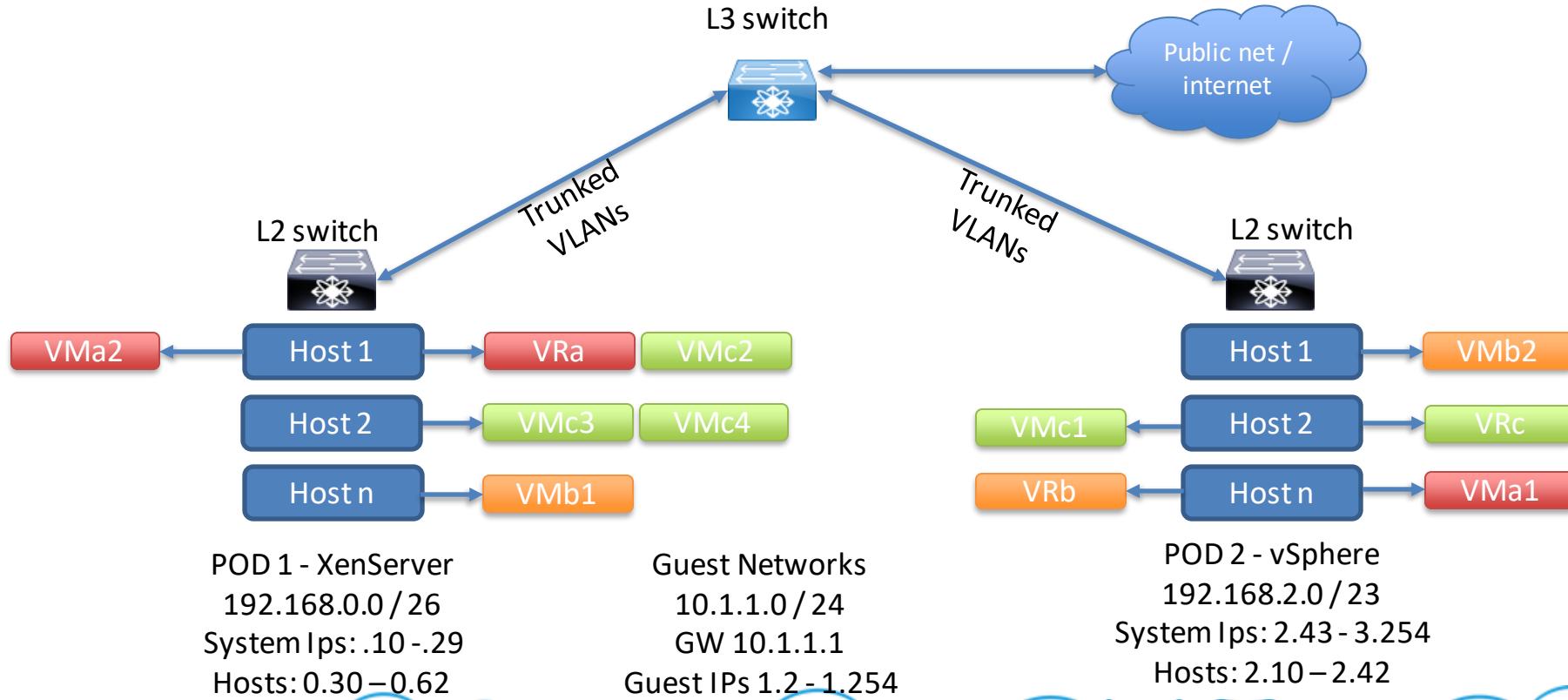


@ShapeBlue

- **Guest networks isolated by VLANs / SDN technologies**
- **Private and shared guest networks**
- **Multiple ‘physical’ guest networks possible**
- **Virtual Router for each network providing any of:**
 - DNS & DHCP
 - Firewall
 - Client VPN
 - Load balancing
 - Source / static NAT
 - Port forwarding



Advanced zone – example IP schema



Network Service Providers

- A hardware or virtual appliance that provide network services to CloudStack e.g.:

Virtual Router
VPC Virtual Router
Internal LBVM
Citrix NetScaler
F5 load balancer
Juniper SRX firewall
VMware NSX (Nicira)

Midokura Midonet
BigSwitch Vns
Cisco VNMC
Baremetal DHCP
Baremetal PXE
Palo Alto
OVS

Questions?



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Storage



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Primary storage

- **Configured at cluster or zone level.**
- **Stores all disk volumes and VM snapshots for VMs.**
- **Options are:**
 - NFS
 - iSCSI
 - PreSetup (i.e. Fibre Channel, OCFS2, CLVM)
 - Local storage
 - Cloudbyte (NFS, iSCSI)
 - Solidfire (iSCSI)
- **Understanding of required performance is critical.**
- **A cluster can have multiple primary storage pools (can be of differing performance and/or technology).**

Primary storage – use cases

- **Local storage:**
 - No requirement for expensive centralised storage
 - Does put limitations on migrations and HA
 - Useful for true cloud stateless instance models
- **Shared storage:**
 - Can be CloudStack managed (NFS, iSCSI) or preSetup.
 - PreSetup allows for more advanced HBA / software based storage connectivity (Fibre Channel, OCFS2, CLVM).
 - Allows for failovers and migrations.



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

How is Primary Storage used?

- Behaves like traditional hypervisor storage.
- Templates are the root volumes of VMs.
- Templates are copied from secondary storage to primary storage by the hypervisor.
- XenServer, vSphere and KVM all create linked clones by default.
- Linked clones are very space efficient, however they create a single point of failure and performance reduces as the chain length increased.
- IOPs requirements often overshadow capacity requirements.

Secondary Storage

- Configured per zone (per region for S3).
- Has to be accessible by management servers, hosts and SSVMs.
- Stores all templates, ISOs and snapshots.
- Has to be NFS (can use S3 but requires an NFS staging area).
- Swift and S3 can be used to replicate data between zones.
- Zones can have more than one secondary storage pool.
- Public templates & ISOs are copied to every store within a zone.



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Storage - questions?



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Using XenServer with CloudStack



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Using XenServer with CloudStack

- **Management server communicates directly with the XenServer pool master.**
- **System VM control is via hosts over link local network.**
- **Physical networking mapped via network labels.**
- **Secondary storage copy jobs are performed by the hosts.**



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Using XenServer with CloudStack

- **Deployment methodology:**

- On nominal master host configure NIC bonding and name each bond inline with 'XenServer Traffic Label' settings used in CloudStack physical networking.
- Create a XenServer pool using XenCenter adding remaining hosts (max 8*), bonds will be created automatically.
- Add 1st Host to CloudStack via GUI/API, all hosts will be added automatically.
- Add NFS/iSCSI primary storage via CloudStack GUI/API.

Other XenServer Considerations

- XenServer uses LinkLocal IP addresses to connect to system VMs and virtual routers.
- HA enabled Guest VMs, shutdown from in-host will auto restart, they need to be shutdown from the CloudStack GUI.
- If connectivity to a storage pool is lost, CloudStack will trigger a forced reboot of the host
(this behaviour can be changed by editing
`/opt/xensource/bin/xenheartbeat.sh` and commenting out each "reboot -f" entry)
- Poolmaster role is protected by the XenServer internal HA mechanism.

Using KVM with CloudStack



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Using KVM with CloudStack

- **CloudStack agent installed on each host.**
- **Relies on Qemu / Libvirt.**
- **System VM control is via link local network.**
- **Physical networking mapped via network bridge names.**
- **Storage can be via NFS or SharedMountPoint (clustered file system, FC)**
- **Clustered files systems are mounted in same location on each host using e.g. OCFS2, CLVM, GFS2.**



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Using KVM with CloudStack

- **Max 16 Hosts per Cluster (set by CloudStack).**
- **Disk format – QCOW2.**
- **Requires network bridges to be configured.**
- **OVS bridges also supported (sensitive to hardware type/drivers).**
- **VMs are suspended during volume snapshots.**



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Using VMware with CloudStack



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

- **Integration via vCenter API - no direct communication with ESXi hosts.**
- **System VM control via CloudStack management network.**
- **All system VMs consume IP addresses in the same CIDR range as management (important design constraint).**
- **Networking mapped to vSwitches (dvSwitches also supported).**

- **Example simple architecture**

- W2K8 R2 MS SQL Server 2008 R2.
- W2K8 R2 vCenter server (connection to SQL via System DSN).
- All Hosts within a cluster must be compatible (16 Max).
- vCenter must be configured to use the standard port 443.
- DRS is supported but CloudStack dictates initial host during deployment.

Using VMware with CloudStack

- **Deployment Methodology**

- Create VMware clusters using vCenter (outside of CloudStack)
- ‘VMware Traffic Label’ within physical networking mapped to vSwitch names.
- vCenter ‘datacentre’ must be associated with the zone.
- The cluster is added into CloudStack which automatically adds the hosts.
- Use VMware Host Profiles to ensure all Hosts are identically configured.
- Add to CloudStack by pointing CloudStack at vCenter.
- Add NFS Primary Storage via CloudStack GUI/API.
- Other storage added as PreSetup.



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Other Vmware considerations

- Pod requires sufficient 'reserved IPs' to allow for system VMs and virtual routers.
- To put a host into maintenance mode:
 - Disable DRS
 - CloudStack maintenance mode
 - vCenter maintenance mode
- Secondary Storage VM (SSVM) must be able to communicate with vCenter.



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Hypervisor comparison

Feature	XenServer	vSphere	KVM
Network Throttling	Yes	Yes	No
Security groups in zones that use basic networking	Yes	No	Yes
iSCSI	Yes	Yes	Yes
FibreChannel	Yes	Yes	Yes
Local Disk	Yes	Yes	Yes
HA	Yes	Yes (Native)	Yes
Snapshots of local disk	Yes	Yes	Yes
Local disk as data disk	Yes	No	Yes
Workload balancing	No	DRS	No
Manual live migration of VMs from host to host	Yes	Yes	Yes

Hypervisor comparison

Primary Storage Type	XenServer	vSphere	KVM
Format for Disks, Templates, and Snapshots	VHD	VMDK	QCOW2
iSCSI support	Yes	VMFS	Yes via Shared Mountpoint
Fiber Channel support	Yes, via existing SR	VMFS	Yes via Shared Mountpoint
NFS support	Yes	Yes	Yes
Local storage support	Yes	Yes	Yes
Storage over-provisioning	NFS	NFS and iSCSI	NFS
SMB/CIFS	No	No	No

Hypervisor integration – questions?



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Adding Zones, PODs and Clusters



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Adding Zones, PODs and Clusters

- **Adding a new zone via the GUI triggers the add zone wizard, which also adds 1 of each of the following:**
 - Pod
 - Cluster
 - Host
 - Primary storage*
 - Secondary storage
- **Zones, pods and clusters can be added independently via GUI or API.**



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Adding Zones, PODs and Clusters

- **The Add Zone Wizard also creates the physical networking and assigns the various traffic types which are:**
 - Public
 - Guest
 - Management
 - Storage
- **Public traffic only applies to advanced zones or basic zones with a NetScaler in place.**
- **Configuring the storage network is optional.**



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Exercise 2: Add a new Zone



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

System VMs



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

CloudStack System VMs overview

- **Based on Debian 9.0 “Stretch” with latest security patches.**
- **Hardened:**
 - No extraneous accounts.
 - Only essential packages installed.
 - SSHd only listens on private/link-local.
 - SSH logins only using keys (keys are generated at install time).
- **Since 4.3 defaulted to 64bit.**
- **Stateless – can be destroyed and recreated (with config re-applied).**
- **High availability option available in advanced zones.**
- **Communicate with management server over management (private) network.**



The Cloud Specialists

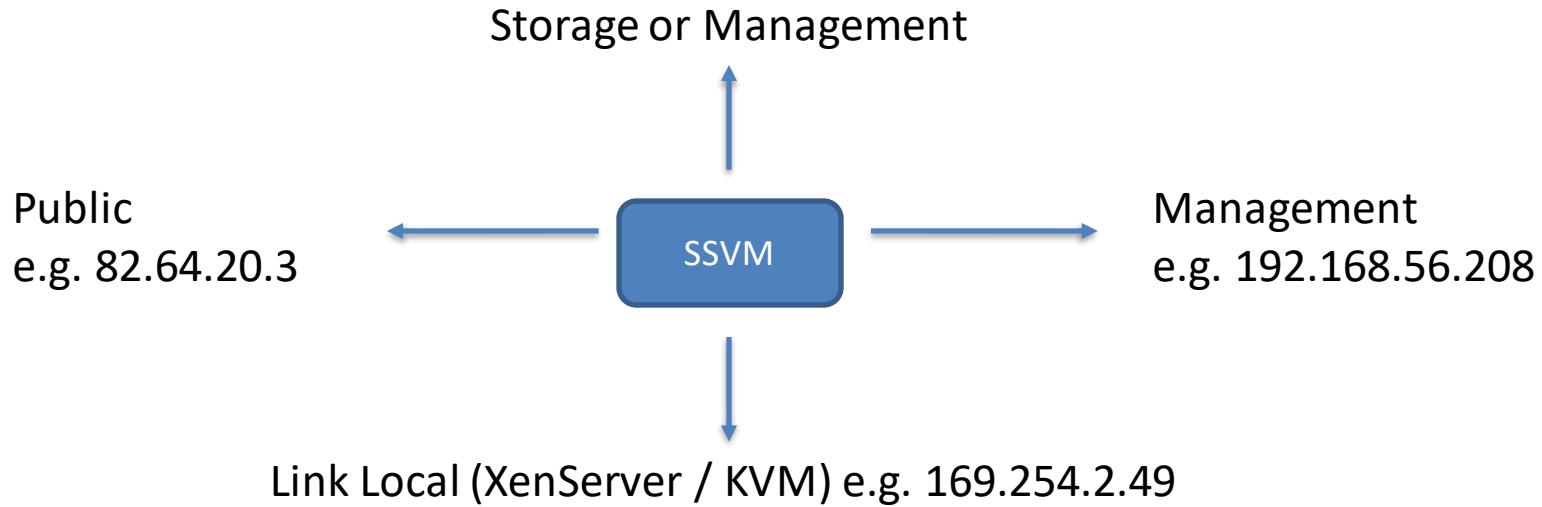
ShapeBlue.com



@ShapeBlue

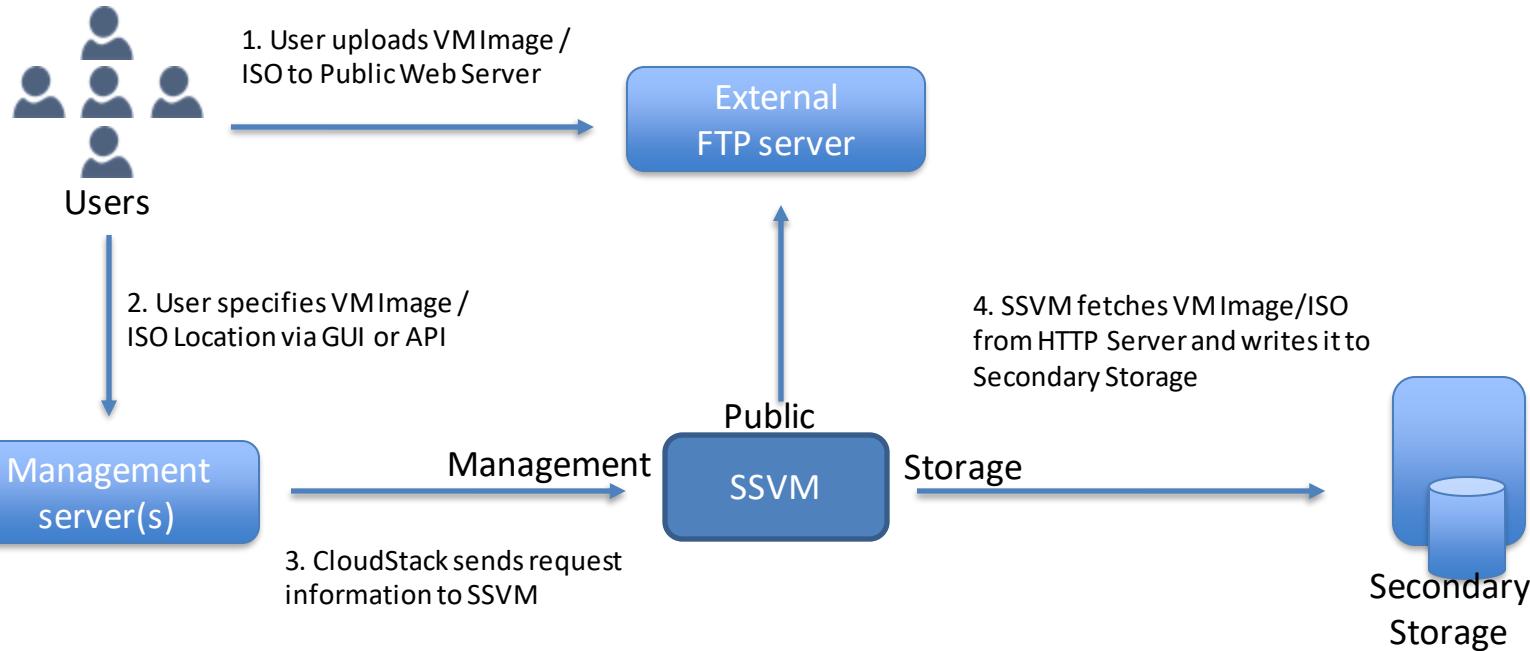
- **Secondary storage virtual machine (SSVM)**
 - Image management:
 - Importing / exporting of templates, ISOs and volumes.
 - Migration of volumes between primary storage pools.
 - Copying OVAs to/from vSphere.
 - Scales out (more spawned) as load increases.
 - Stateless:
 - Can be destroyed, CloudStack will automatically create a new one.

SSVM networking



Secondary Storage VM

VM Image / ISO FTP Upload Workflow



Secondary Storage VM

VM Image / ISO Local Upload Workflow

1. User prepares VM Image / ISO on local PC



Users

2. User specifies VM Image / ISO Location via GUI or API



Management server(s)

Management

3. CloudStack sends request information to SSVM



Storage

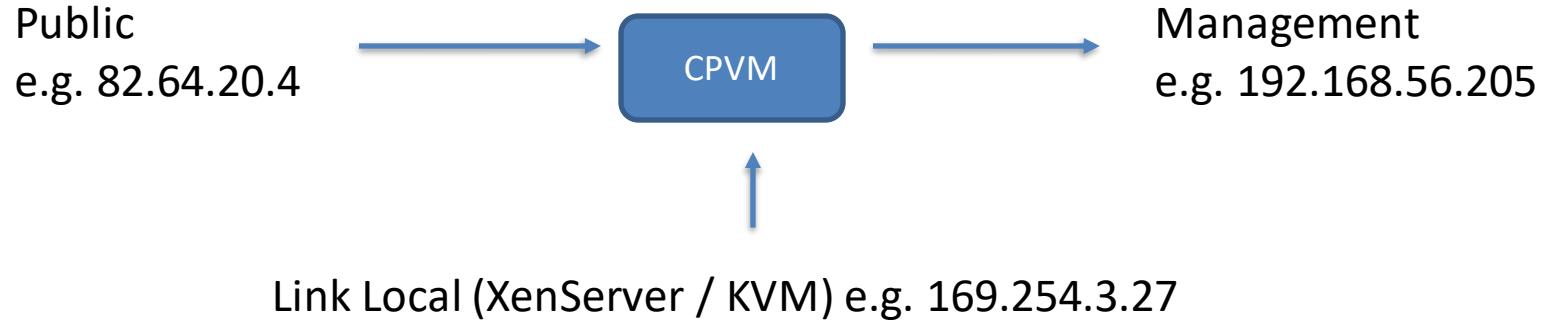
4. SSVM receives VM Image / ISO through session and writes it to Secondary Storagebrowser



Secondary Storage

- **Console Proxy Virtual Machine (CPVM)**
 - Provides AJAX-style HTTPs console viewer
 - Proxies VNC output from hypervisor
 - Scales out (more spawned) as load increases
 - Stateless
 - Can be destroyed, CloudStack will automatically create a new one

Console Proxy VM networking



Note direction of communications



The Cloud Specialists

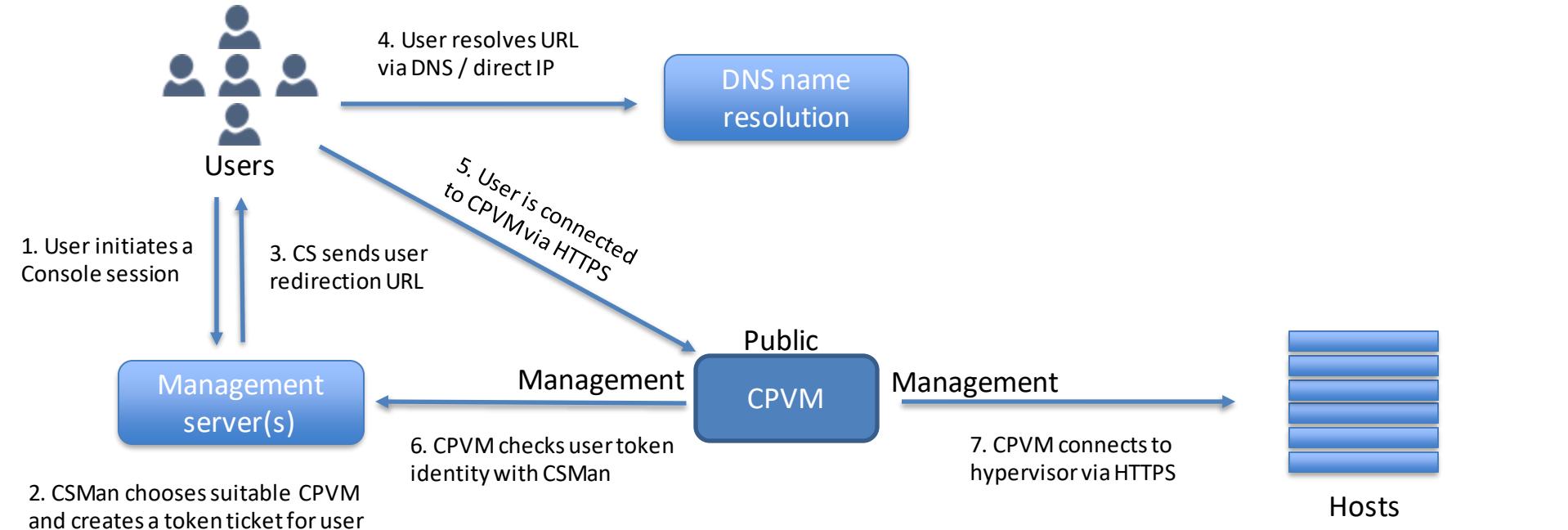
ShapeBlue.com



@ShapeBlue

Console Proxy VM

Remote connection handshake



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

- **Virtual Router (VR):**

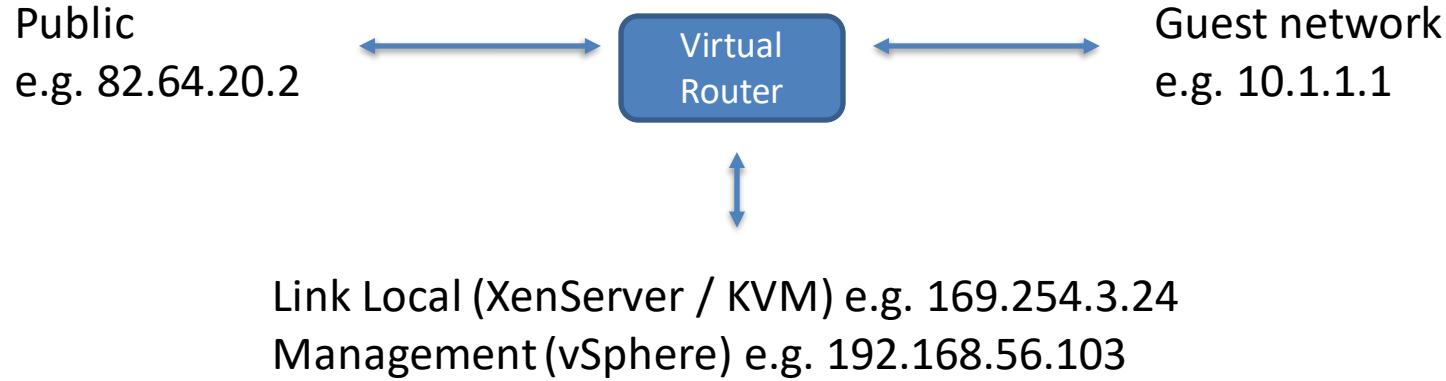
- Provides multiple network services to guest VMs
- DHCP, DNS, routing (static NAT, source NAT) , firewall, port forwarding, VPN
- User-data, meta-data, SSH keys and password change server
- Redundancy via VRRP
- Management server configures VR over SSH (proxied via the hypervisor on XS and KVM)
- Does NOT respond to PING on its public interface

- **Virtual Router (VR)**

- Created when first required – i.e. do not start until the first guest VM instance starts (VPC VRs can run in persistent mode)
- VMs will not start until the network VR is up.
- Advanced zone: one VR per guest network
- Basic zone: one per pod – but provide only basic network services (does not do any actual routing).

Virtual Router

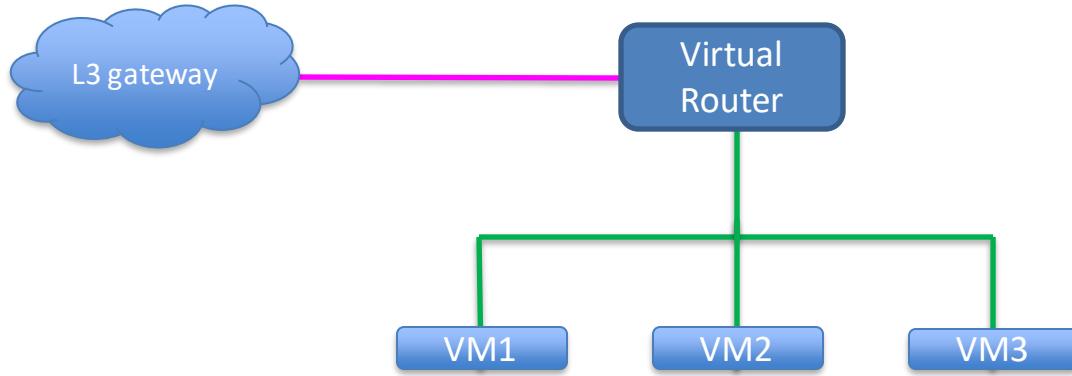
Networking



Virtual Router (Advanced Zone)

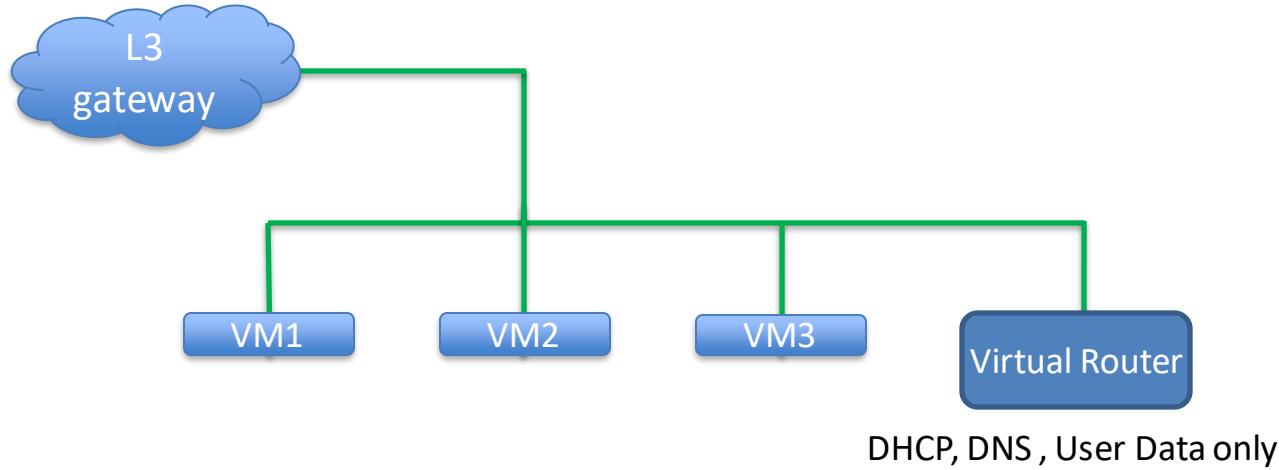
Networking

DHCP, DNS , User Data, Source NAT, Static NAT, VPN, Firewall, Port Forwarding, Load Balancing, Virtual Private Cloud

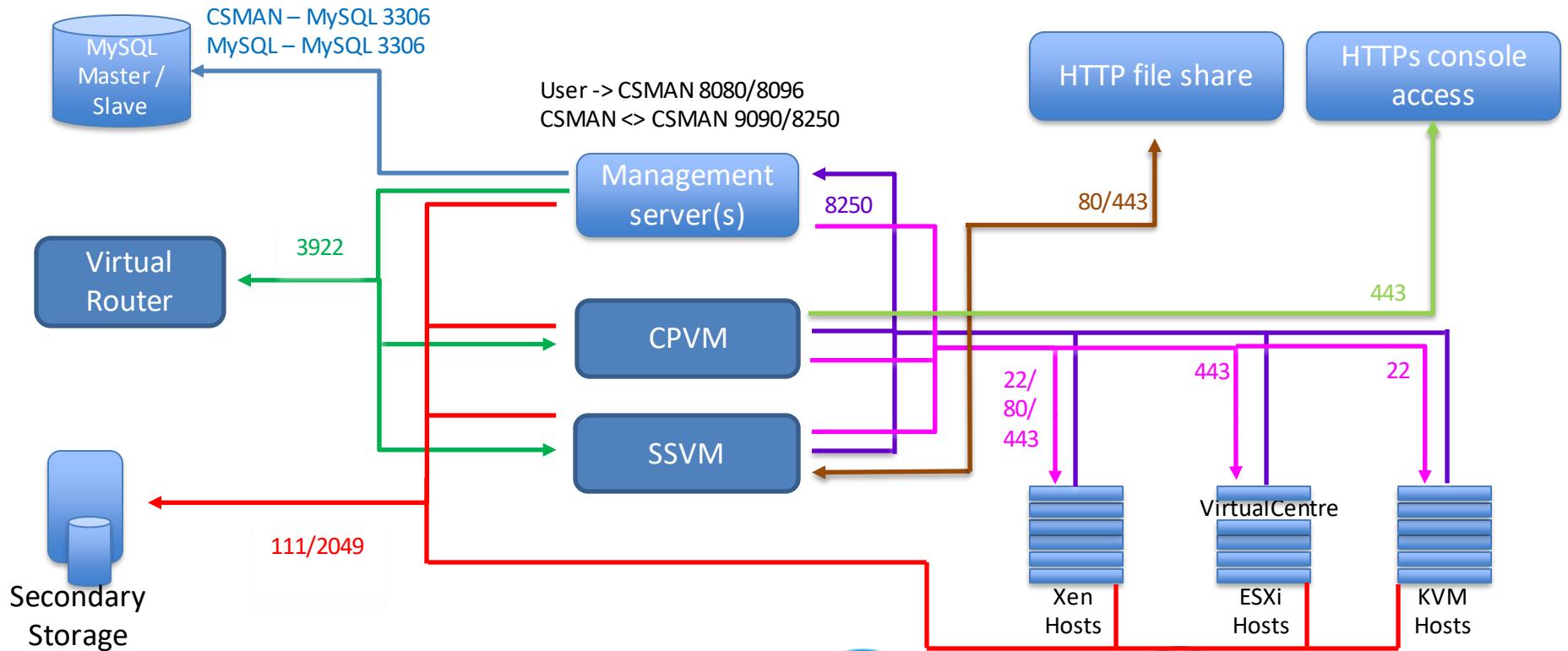


Virtual Router (Basic Zone)

Networking



Communication Ports



Exercise 3: Accessing System VMs



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Service Offerings



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Service Offerings

- When creating a new Instance the end user can choose various characteristics and capabilities, these are controlled by the following three Service Offerings:
 - Compute
 - Disk
 - Network



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Compute Offerings

- Every VM is linked to a Compute Offering, it controls what resources are provisioned

The image shows two overlapping configuration dialogs. The background dialog is titled "Add compute offering" and contains fields for Name, Description, Storage Type (shared), Provisioning Type (thin), Custom checkbox, # of CPU Cores, CPU (in MHz), Memory (in MB), Network Rate (Mb/s), QoS Type, and Offer HA checkbox. The foreground dialog has fields for Storage Tags, Host Tag, CPU Cap (checkbox), Public (checkbox), Volatile (checkbox), Deployment planner, Planner mode, GPU, and Domain (set to ROOT). Both dialogs have "Cancel" and "OK" buttons.

Add compute offering

* Name:

* Description:

Storage Type: shared

Provisioning Type: thin

Custom:

* # of CPU Cores:

* CPU (in MHz):

* Memory (in MB):

Network Rate (Mb/s):

QoS Type:

Offer HA:

Storage Tags:

Host Tag:

CPU Cap:

Public:

Volatile:

Deployment planner:

Planner mode:

GPU:

Domain: ROOT

Cancel OK

- Compute Offerings

- **Cannot be edited once created:**
 - Only name and description can be updated
- **Can be deleted, however:**
 - Existing VMs will continue to use it
 - New VMs will not be able to use it
 - I.e. still exists in the CloudStack database



The Cloud Specialists

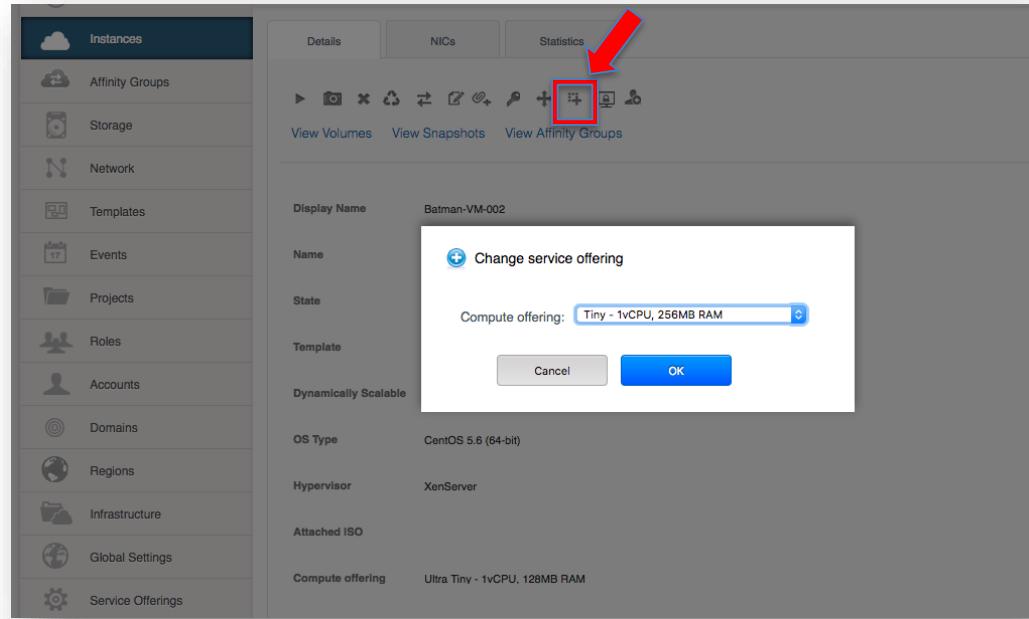
ShapeBlue.com



@ShapeBlue

Compute Offerings

- Users can change the current service offering by selecting the “Change service” button from the instances pane.
- If the OS supports it, a VM can be dynamically up-scaled without a reboot.
- Downscaling always requires a reboot.



System Offerings

- System offerings are almost identical to compute offerings, except they are only used by system VMs.
- They have an extra ‘System VM Type’ field and are lacking the deployment planner and volatile fields.

 Add System Service Offering

* Name:

* Description:

System VM Type: Domain router

Storage Type: shared

Provisioning Type: thin

* # of CPU Cores:

* CPU (in MHz):

* Memory (in MB):

Network Rate (Mb/s):

Disk Read Rate (BPS):

Disk Write Rate (BPS):

Disk Read Rate (IOPS):

Disk Write Rate (IOPS):

Offer HA:

Storage Tags:

Host Tags:

CPU Cap:

Public:

Domain: ROOT



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Disk offerings

- Define the virtual disks that the end users will be able to create when creating a data volume, or deploying a VM using an ISO.
- By using storage tags, the offerings can be mapped to different storage platforms.
- QoS at hypervisor or storage device.
- As well as pre-defined sizes a ‘Custom’ option is available. Global setting limits maximum ‘custom’ size.



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Disk offerings

Add Disk Offering

* Name:

* Description:

Storage Type: shared

Provisioning Type: thin

Custom Disk Size:

* Disk Size (in GB):

QoS Type:

Write-cache Type: No disk cache

Storage Tags:

Public:

Domain: ROOT

Cancel **OK**

Add Disk Offering

* Name:

* Description:

Storage Type: shared

Provisioning Type: thin

Custom Disk Size:

* Disk Size (in GB):

QoS Type: hypervisor

Disk Read Rate (BPS):

Disk Write Rate (BPS):

Disk Read Rate (IOPS):

Disk Write Rate (IOPS):

Write-cache Type: No disk cache

Storage Tags:

Public:

Domain: ROOT

Cancel **OK**

Add Disk Offering

* Name:

* Description:

Storage Type: shared

Provisioning Type: thin

Custom Disk Size:

* Disk Size (in GB):

QoS Type: storage

Custom IOPS:

Min IOPS:

Max IOPS:

Hypervisor Snapshot Reserve:

Write-cache Type: No disk cache

Storage Tags:

Public:

Domain: ROOT

Cancel **OK**



The Cloud Specialists

ShapeBlue.com

@ShapeBlue

Network offerings

- Network offerings define what network services or features will be available to the end user on their guest network.
- A set of default offerings are available and these cover the most common use cases.
- The services and features defined by network offerings are then delivered by a Virtual Router, physical network devices or combinations of these.



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Network offerings

- Network rate
 - Isolated or shared
 - Persistence
 - Specify VLAN/IP
 - VPC
 - Conserve mode
 - Tags
- Services:
 - VPN
 - DHCP
 - DNS
 - Firewall
 - Load balancer
 - User data
 - Source / Static NAT
 - Port forwarding
 - Security groups
 - Network ACL
 - Virtual networking
 - Bare metal PXE

 Add network offering

* Name:

* Description:

Network Rate (Mb/s):

Guest Type: Isolated Shared

Persistent :

Specify VLAN:

VPC:

Supported Services:

VPN: <input type="checkbox"/>
DHCP: <input type="checkbox"/>
DNS: <input type="checkbox"/>
Firewall: <input type="checkbox"/>
Load Balancer: <input type="checkbox"/>
User Data: <input type="checkbox"/>
Source NAT: <input type="checkbox"/>
Static NAT: <input type="checkbox"/>

Conserve mode:

Tags:

Exercise 4a,b,c: Creating Service Offerings



Blue The Cloud Specialists

ShapeBlue.com



@ShapeBlue

End of Day 1

Exercise A1: System Shutdown Procedure



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

**End of day 1
Any questions?**



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

- **What are the 7 layers of a CloudStack cloud?**
 - Regions
 - Zones
 - PODs
 - Clusters
 - Hosts
 - Storage
 - Network

- **Two Networking Models**
 - Basic
 - Advanced
- **Four Physical Networks**
 - Public
 - Guest
 - Management
 - Storage

- **How can Guest isolation be achieved in a Basic Zone?**
 - Security Groups
- **What services does the Virtual Router provide to a Basic Network?**
 - DHCP
 - DNS
 - UserData
 - Security Groups

- **What services does the Virtual Router provide to an Advanced Network?**
 - DNS & DHCP
 - Firewall
 - Client IPSEC VPN
 - Load Balancing
 - Source / Static NAT
 - Port Forwarding
- **What are the three System VMs and what are their functions?**
 - SSVM
 - CPVM
 - Virtual Router

- **How do you put a VMware Host into Maintenance Mode?**
 - Disable DRS, then CloudStack maintenance mode , then vCenter maintenance mode

- **What are the two types of Storage, and what are their functions?**
 - Primary: Running VMs
 - Secondary: Templates, Snapshots, ISOs

- **What are the four types of Service Offerings?**
 - Compute
 - Storage
 - Network
 - System

Domains, Accounts & Users



The Cloud Specialists

ShapeBlue.com



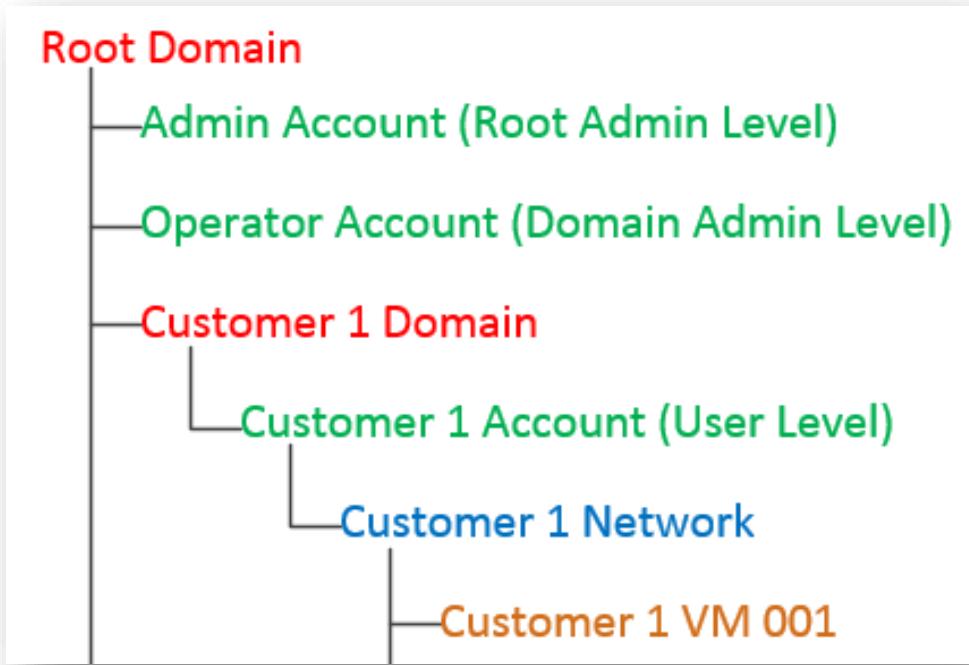
@ShapeBlue

- **A domain is an organisational unit for accounts**
- **An account represents a single tenant**
- **Every account has at least one user**
- **Users within an account share the resources and are not isolated**
- **All resources belong to an account, not a user or domain**

- Domains usually contain accounts that have some form of logical relationship, e.g.
 - all accounts created by a reseller would be under that resellers domain
 - for a private cloud, the account could map to business units
- As some CloudStack resources e.g. compute offerings can be either public or linked to a specific domain, placing each account in its own domain can be beneficial.

Domains, Accounts & Users

- Example of Domain / Account hierarchy



- Domain admins have user level access to all accounts in their domain and associated sub-domains.
- Domain admins can also
 - Create accounts
 - Create users
 - Reset passwords
- Domain admins cannot create new sub-domains under the root domain – only under own domain
- Root admins have full control over all domains and accounts and infrastructure

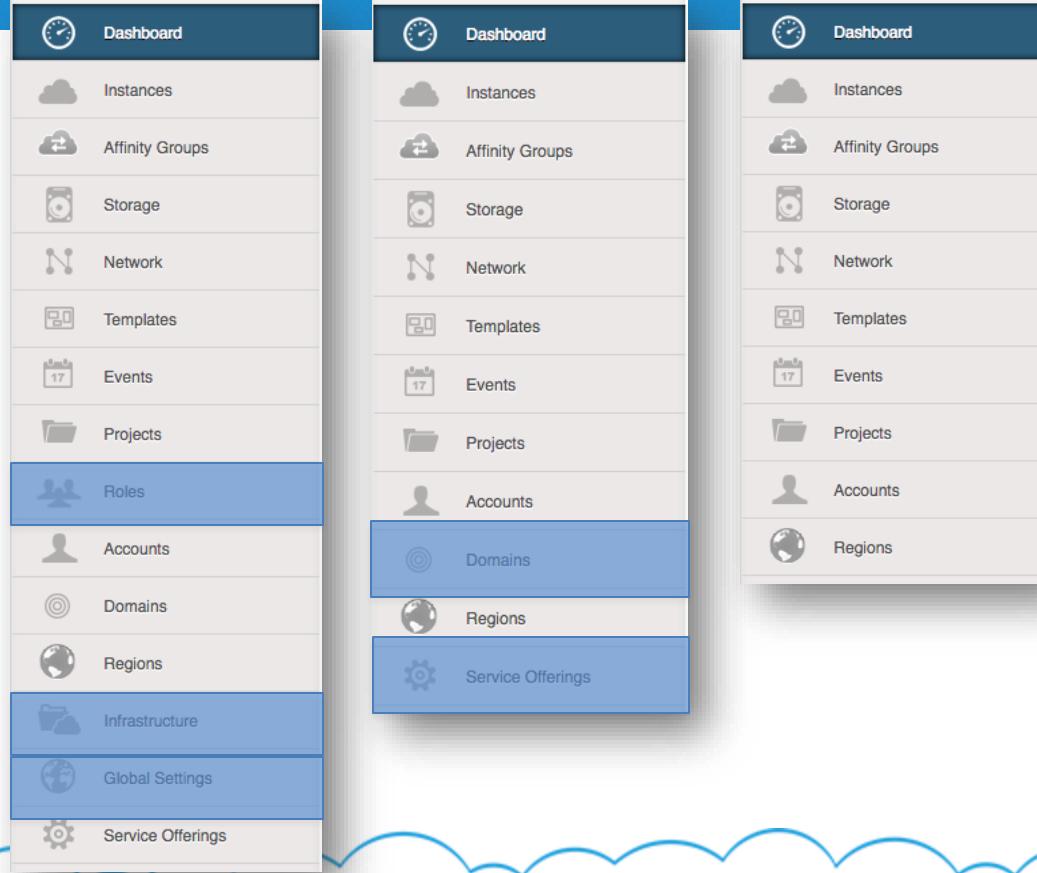
CloudStack dynamic roles

- **CloudStack is since 4.7 using dynamically created roles.**
- **There are four built in roles:**
 - Root admin
 - Resource admin
 - Domain admin
 - User
- **CloudStack allows roles to be added or changed.**

Name	Type	Description	Quickview
Root Admin	Admin	Default root admin role	+
Resource Admin	ResourceAdmin	Default resource admin role	+
Domain Admin	DomainAdmin	Default domain admin role	+
User	User	Default Root Admin role	+

Role comparison

- Root Admin
- Domain Admin
- User



The Cloud Specialists

ShapeBlue.com

@ShapeBlue

Creating new roles

- **Adding / editing roles:**
 - Each role consists of allow/deny rules for each CloudStack API call.
 - Rules can be added in similar fashion to firewall rules.
 - Rules allow for wildcards, e.g. “add*” to include all “add” API calls
 - The built in roles are sufficient for most scenarios.

The image shows two parts of a CloudStack interface. The top part is a modal dialog titled 'Add Role' with fields for 'Name' (MyNewRole), 'Description' (New admin role), and 'Type' (Admin, Domain Admin, User). The bottom part is a table titled 'Details' showing a list of API rules with columns for Rule, Permission, Description, and Action (Add).

	Rule	Permission	Description	Action
		Allow		Add
≡	activateProject	Allow		X
≡	addAccountToProject	Allow		X
≡	addIpToNic	Allow		X
≡	addNicToVirtualMachine	Allow		X
≡	addVpnUser	Allow		X

- **Shared networks can be assigned to:**
 - Single account
 - Domain and associated accounts
 - Domain and all of its sub-domains/accounts

Dedicated Resources

Resource	Domain	Account	Project
Zone	Y	Y	N
Pod	Y	Y	N
Cluster	Y	Y	N
Host	Y	Y	N
Public IP range	N	Y	Y
Guest VLAN Range	N	Y	Y
Compute Offering	Y	N	N
Disk Offering	Y	N	N

Exercise 5a,b: Creating Domains and Accounts



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Projects



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Projects

- **Share resources between accounts *in the same domain***
- **The project owns the resources just like a regular account would**
- **The creator is the project administrator (PA)**
- **PA can add and remove Accounts, and allocate a new PA**
- **Invitations (if enabled) enables invitees to accept before being added to the Project**
- **The following global setting enables invitations**
`project.invite.required`



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Projects

The screenshot shows a cloud management interface for a project named "Bootcamp". The left sidebar contains navigation links: Dashboard, Instances, Affinity Groups, Storage, Network, Templates, Events, Projects (selected), Accounts, and Regions. The main content area has a breadcrumb path: Home > Project dashboard > Dashboard. It features two main sections: "Compute and Storage" and "Users". The "Compute and Storage" section displays metrics: 0 Virtual Machines (Running), 1 Virtual Machine (Stopped), 1 Storage Volumes, and 200 mb/s Bandwidth. The "Users" section lists two users: "User" and "admin". To the right, there's a "Networking and security" summary with 1 IP Addresses, 0 Load balancing policies, and 0 Port forwarding policies, along with a "Manage Resources" button. Below that is an "Events" log with entries from March 22, 2017, and March 21, 2017.

Networking and security	
1	IP Addresses
0	Load balancing policies
0	Port forwarding policies
Manage Resources >	

Events	
03/22/17	user has logged in from IP...
03/21/17	user has logged out
03/21/17	user has logged in from IP...

Domains, accounts, users and projects

Any questions?



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Limits



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Resource limits

- Resource limits can be set per account, project or domain
- They can be applied to:
 - Instances
 - VPCs
 - CPUs
 - Memory
 - Pri / Sec Storage (GB)
 - Volumes
 - Networks
 - Public IPs
 - Snapshots
 - Templates



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

- Domain limits are set to unlimited '-1' (accumulative)
- Global settings control the default values
- There are numerous other limits controlled by global settings, e.g.:
 - Max no of hourly/daily/weekly/monthly volume snapshots
 - Max no of volumes
 - Max / min custom volume size
 - Max no of VPC tiers
 - Template size (import limit)

Notifications and thresholds



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

- **CloudStack provides alerts when the usage of resources approach their maximum limits**
- **Administrators can define disable thresholds for:**
 - Storage pools – allocated and capacity
 - Clusters – CPU and memory
- **Disable means that resources above the threshold will not be considered when a new instance is deployed**
- **The thresholds are defined using global settings**
- **Notification thresholds should be lower than the disable limits**

Configuring Notifications

- **Notifications settings are found in the global settings**
- **SMTP server details need to be entered for mail server in use, and can be different for general alerts.**
- **The alert recipient address is common (can be multiple)**



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Configuring notifications

Name	Description	Value	Actions
alert.smtp.host	SMTP hostname used for sending out email alerts.	smtp.gmail.com	
alert.smtp.password	Password for SMTP authentication (applies only if alert.smtp.useAuth is true).	xxxxxxxxxx	
alert.smtp.port	Port the SMTP server is listening on.	465	
alert.smtp.useAuth	If true, use SMTP authentication when sending emails.	true	
alert.smtp.username	Username for SMTP authentication (applies only if alert.smtp.useAuth is true).	shapeblue.demo@gmail.com	



The Cloud Specialists

ShapeBlue.com

 @ShapeBlue

Configuring notifications

Home > Global Settings >

Select view: Global Settings

Name	Description	Value	Actions
alert.email.addresses	Comma separated list of email addresses used for sending alerts.	alerts@acme.com	<input type="button" value="📝"/>
alert.email.sender	Sender of alert email (will be in the From header of the email).	cloudadmin@acme....	<input type="button" value="📝"/>



Blue The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Virtual Machine Allocation



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Virtual Machine Allocation

- **CloudStack is responsible for determining the best host to deploy a VM instance onto**
- **VMware DRS: CloudStack still chooses the initial host**
- **Admins can configure host preferences for workloads, e.g.:**
 - Host OS preference set to Windows
 - Windows VMs will *favour* these hosts
 - If no suitable ‘Windows’ host is available, a non Windows host will be chosen instead, based on standard capacity metrics



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

- **Global configuration ‘vm.allocation.algorithm’ enables root admins to influence the VM deployment process:**
 - random: randomly chooses a host based with sufficient capacity
 - firstfit: places new VMs on the first host that is found having sufficient capacity to support the VM’s requirements
 - userdispersing: makes the best effort to evenly distribute VMs belonging to the same account on different clusters or pods
 - userconcentratedpod_random: similar to the random algorithm, VMs are however concentrated on the same pod
 - userconcentratedpod_firstfit: similar to the firstfit algorithm, however hosts are chosen within a chosen pod rather than across a zone

Capacity Ordering

- **Capacity ordering is done by looking at the allocated CPU first and then the allocated RAM for the compute resource.**
Controlled by the global setting ‘host.capacityType.to.order.clusters’
Can be set to CPU (default) or RAM.
- **Within a chosen cluster, a particular host and primary storage is chosen by using ‘random’, ‘firstfit’ or ‘userdispersing’ strategy.**
- **The ‘userconcentrated’ algorithm only influences the POD choice**



The Cloud Specialists

ShapeBlue.com

 @ShapeBlue

Virtual Machine Deployment & Management



The Cloud Specialists

ShapeBlue.com

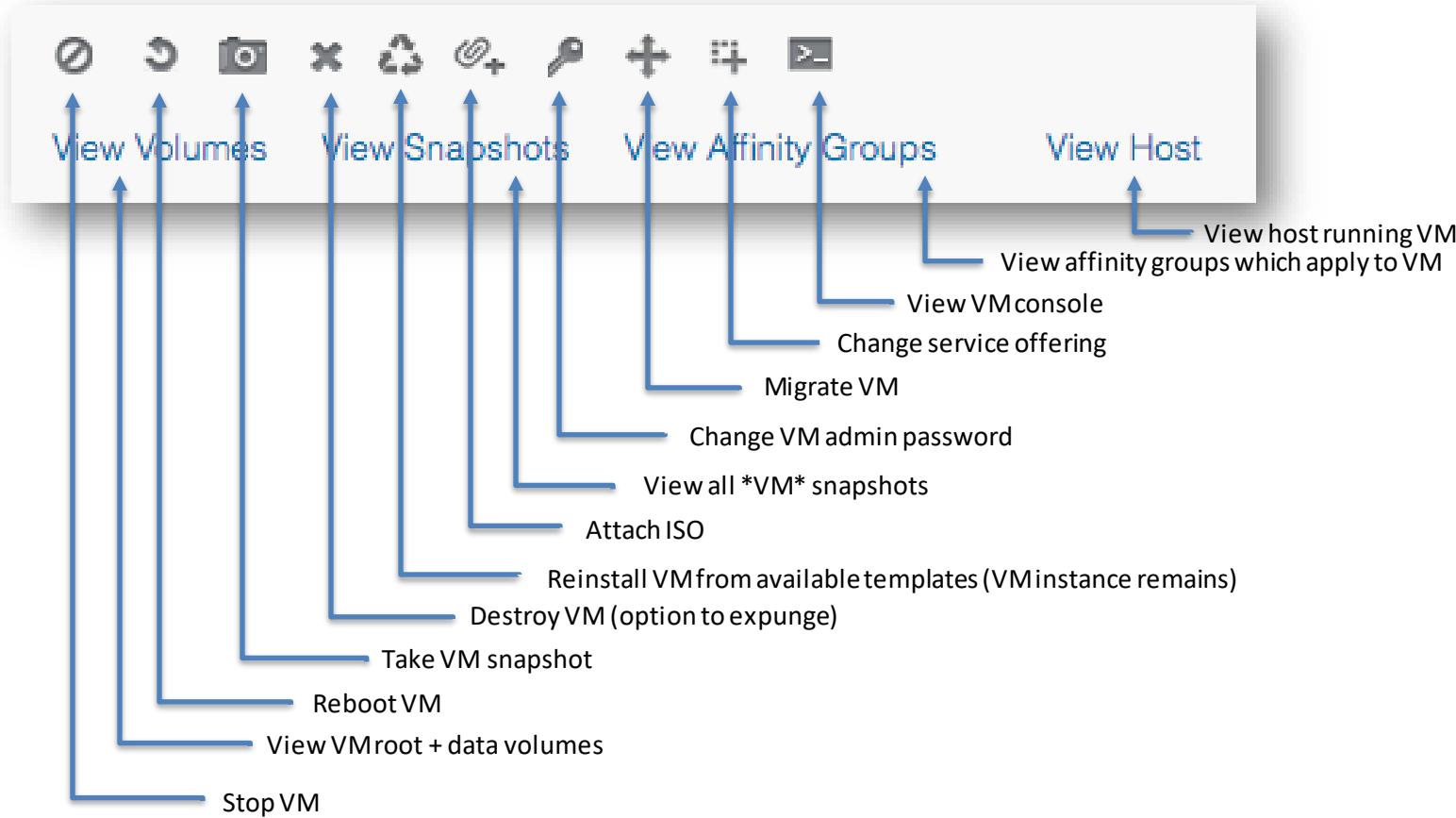


@ShapeBlue

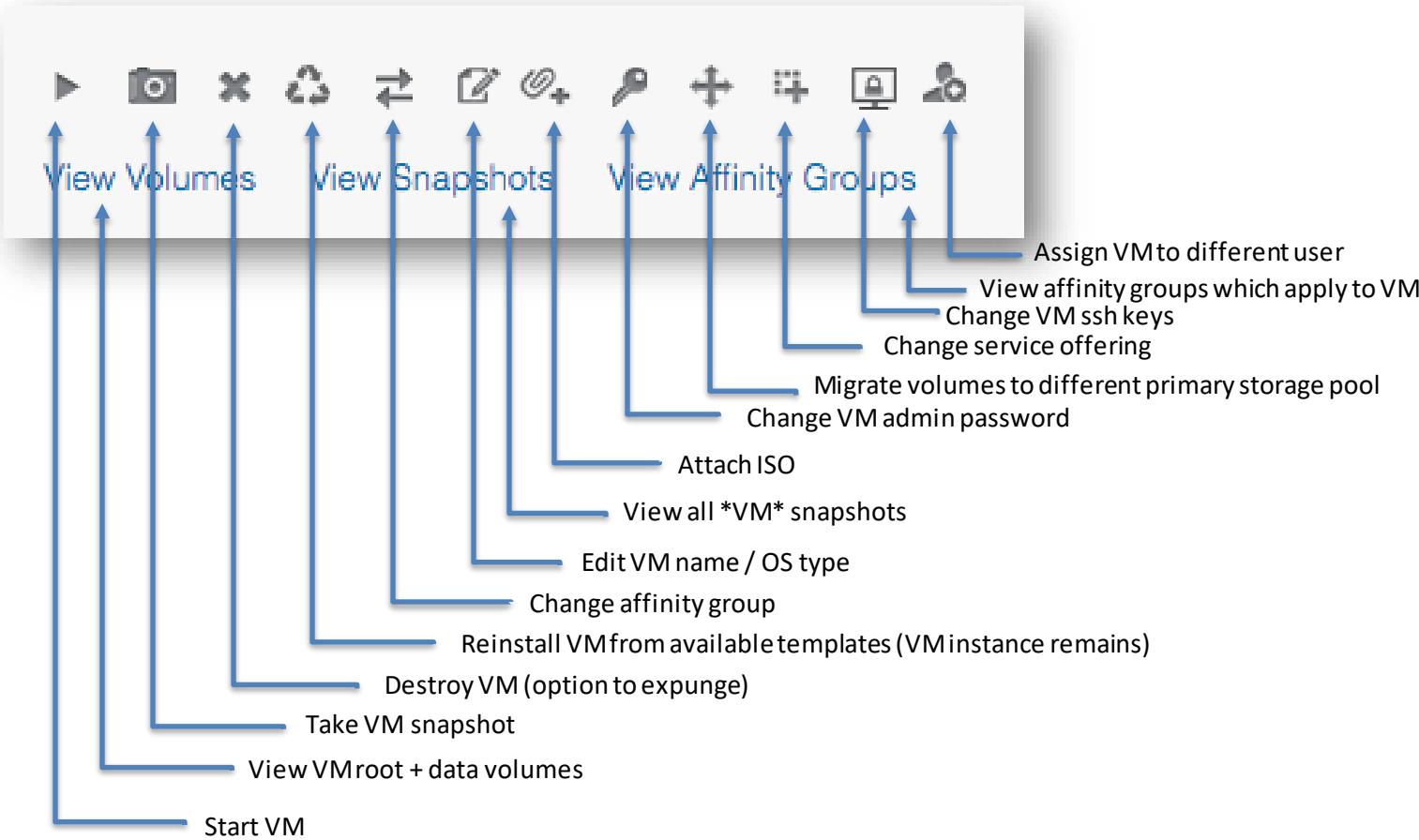
- **Users make choices when deploying a new VM**
 - Region / Zone: which zone to deploy to
 - Template / ISO: create from template or build from ISO
 - Compute Offering: CPU/RAM / storage / network performance etc.
 - Disk Offering:
 - If deploying from template this only applies to additional data disks
 - If deploying from ISO this applies to the size of the root disk
 - Affinity Group: optional (anti-affinity groups only)
 - Network: attach to one or more networks
 - Review: optionally assign a name and put the VM in a group

- **Network is created if not already in existence**
- **Virtual Router is created based on system VM template**
- **POD, cluster, host is chosen first, followed by primary storage**
- **Guest VM template is copied from secondary to primary if it is not already on the chosen primary storage.**
Copying is performed by hypervisor host, not the SSVM (except VMware).
- **Guest VM is created as a linked clone of the guest template.**

VM Operations – Running VM



VM Operations – Stopped VM



System VM naming convention

- **System VMs are named using the following format**
 - r-n-VM
- **r - Role**
 - s = Secondary Storage VM
 - v = Console Proxy VM
 - r = Virtual Router
- **n – Instance ID (starts at 1 then increments for every VM)**
- **e.g.**
 - s-1-VM, v-2-VM, r-4-vm



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

User VM naming convention

- **Guest VMs are named using a similar format:**
 - r-a-n-VM
- **r – Role: always set to “i” to indicate guest instance**
- **a – Account ID (value from ID field in the account table)**
- **n – Instance ID**
- **e.g.**
 - i-5-37-VM



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Exercise 6: Creating and Managing VMs



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Managing VMs – Optional Exercises

- Create a Static NAT for Batman-VM-001 then configure some Firewall Rules (hint – you will need an additional Public IP)
- Create a 2nd Network and move Batman-VM-002 onto it
- Add a Secondary IP to Batman-VM-002



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

- A ‘destroyed’ VM is not immediately purged
- Global settings ‘expunge.delay’ and ‘expunge.interval’ determine when the actual deletion occurs
 - expunge.delay: Determines how old the volume must be
 - expunge.interval: Determines how often the check occurs
 - Both default to 86400s, i.e. 24 hours.

Managing Volumes

- **There are two types of volumes – root and data**
- **Root volumes**
 - The volume where the VM operating system resides
 - Templates are essentially a root volume
 - Stopped VMs, and their root volume can be migrated to alternate primary storage via secondary storage*
 - XenServer storagemotion and VMware storage vMotion can live migrate running VMs to alternative primary storage pool



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

- **Data volumes**

- Data volumes can be added to a VM when deploying from a template
- Data volumes can be created at any time independently from the VM then later attached
- Volumes are only actually created when they are 1st attached to a VM
- Support for hot plug / unplug of data volumes (OS dependent)
- Detached volumes can be migrated to alternative primary storage, or even exported from the system

- **Root volumes will be placed on primary storage attached to the chosen host during the deployment process.**
- **Data volumes will typically be placed on the same primary storage as the root volume of the VM it is attached to, unless the disk offering dictates otherwise**
- **Local storage**
 - Both root and data volumes can be placed on local storage, but they must be on the same host

Exercise 7: Creating and Managing Volumes



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Managing Templates



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

- **Templates are pre-configured guest instances.**
- **They contain the VM root volume and cannot include additional disks.**
- **Templates can be:**
 - Single or multi Zone
 - Extractable (this can be disabled)
 - Password enabled (contain a password reset script)
 - Public or private (ability for users to create public templates can be disabled via global settings)

Managing Templates

- **Templates are imported into CloudStack via the ‘Register Template’ function or uploaded from the users local PC:**
 - Register template downloads the template from a public facing http/ftp site.
 - Upload from local PC copies the template from the local browser session
- **The SSVM is responsible for fetching the template image from the webserver which is hosting it, and transferring it to secondary storage**
- **Public templates are copied to every secondary storage pool within the zone for resilience**
- **Private templates are copied to one random secondary storage pool**

Managing Templates

- **VMware templates require some additional settings when registering them with CloudStack:**
 - Root disk controller type – SCSI or IDE
 - NIC adapter type – E1000/PCNet32/Vmxnet2/Vmxnet3
 - Keyboard type – US / Japanese
 - Format - OVA



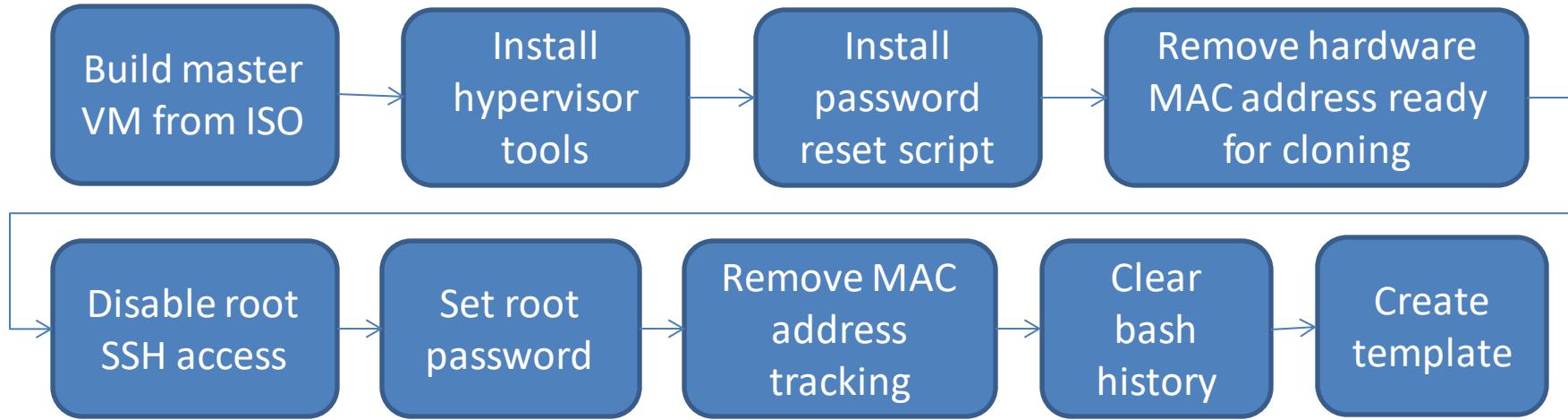
The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Creating Linux Templates



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Windows Sysprep

- Windows templates need to be syspre'ed prior to deployment
- Due to limitations on the amount of times an install can be sys-prepped, 'master' non sys-prepped images should be maintained
- A sysprep answer file needs to be created to enable zero touch deployments



The Cloud Specialists

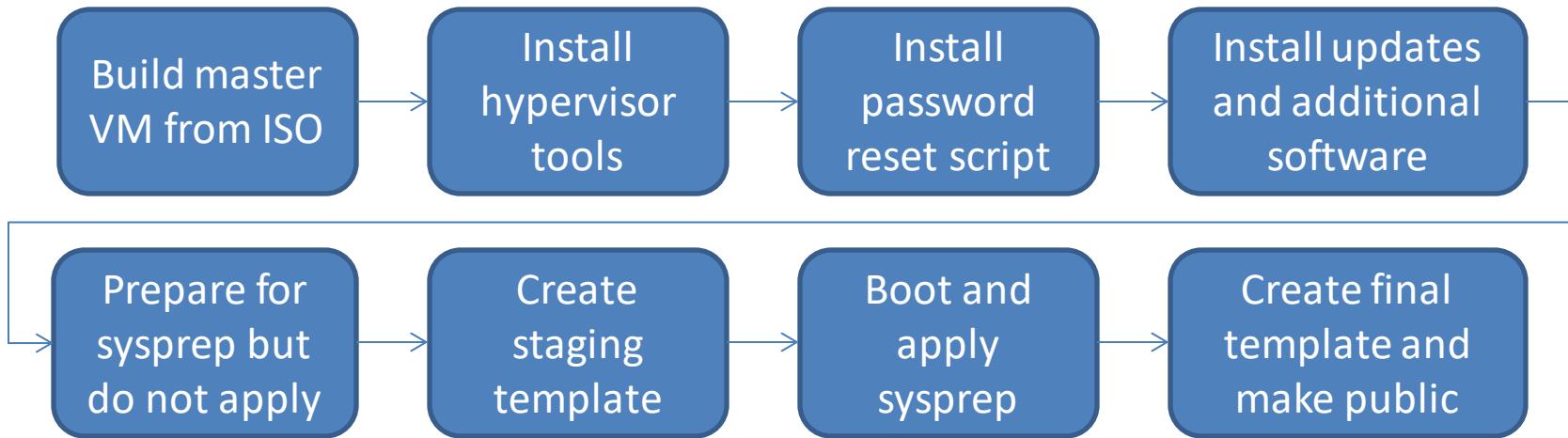
ShapeBlue.com



@ShapeBlue

- **The sysprep answer file should contain at least the following:**
 - Regional settings – eg UK keyboard, etc.
 - SPLA product key to enable auto activation.
 - Persistent drivers – required for VMware to prevent VMware virtual drivers from being replaced with Microsoft drivers.
 - Post-sysprep cleanup script – removes answer file once completed

Creating Windows Templates



Running Sysprep

- **Copy the unattend.xml file to c:\windows\system32\sysprep**
- **Copy the setupcomplete.cmd file into %WINDIR%\Setup\scripts**
- **From a command prompt at 'c:\Windows\System32\sysprep' run the following command**
 - sysprep.exe /generalize /oobe /shutdown /unattend:unattend.xml
- **The VM will shut down once sysprep has completed:**
 - ensure the VM is not running on a HA Compute offering
 - do not manually start the VM
- **Now create the final template from the root volume**



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Snapshots



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

VM Snapshots

- Traditional hypervisor snapshot
- Can be used for quickly reverting back to a previous state
- Can optionally capture memory
- Quiesce VM available on VMware (with tools)
- Not available for KVM
- XenServer pauses the VM to snapshot memory
- Remains on primary storage (as opposed to volume snapshots)
- Can not:
 - download a VM snapshot
 - create a template from a VM snapshot



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Volume Snapshots

- In reality simply a volume backup
- VMs with multiple volumes will need multiple snapshots configured
- Users can initiate a single snapshot, or schedule recurring snapshots
- To recover a root volume from snapshot either attach it as a data volume to an existing VM, or create a template from the snapshot and create new VM
- Get copied to secondary storage



The Cloud Specialists

ShapeBlue.com

 @ShapeBlue

Exercise 8: Snapshots



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Snapshots – Optional Exercises

- **Restore a VM from a Volume Snapshot (this will also require you to create a new Template)**

Managing CloudStack

Any questions?



Blue The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Virtual Private Cloud



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

- **Private multi-tiered virtual networks**
- **ACLs between tiers and public networks**
- **Inter-tier routing (layer 2)**
- **Site-2-site VPN**
- **Private gateway**
- **VPC-2-VPC VPN**
- **User VPN**
- **Inter-tier and inbound load balancing**

- **No ‘conserve mode’ so additional unique public IPs required for:**
 - Source NAT
 - Port forwarding
 - Load balancing
- **Redundant VPC now available**

Virtual Private Clouds (VPC)

- **User creates a super CIDR for the VPC**
- **All tiers' subnets must be within the Super CIDR and must not overlap**
- **E.g.**
 - Super CIDR: 10.0.0.0/16
 - Tier subnet1: 10.0.1.0/24
 - Tier subnet2: 10.0.2.0/24



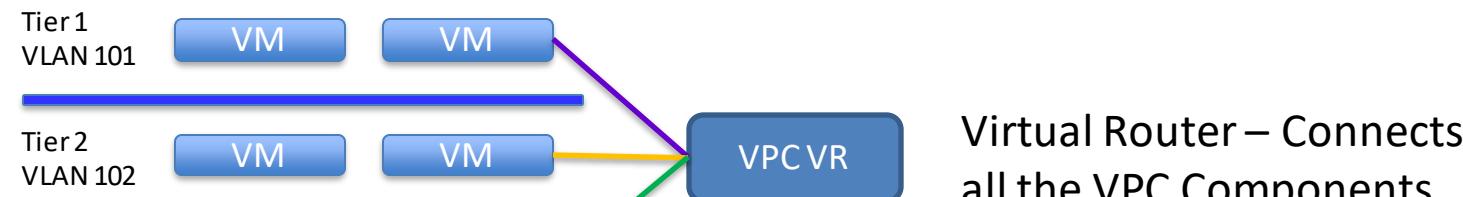
The Cloud Specialists

ShapeBlue.com



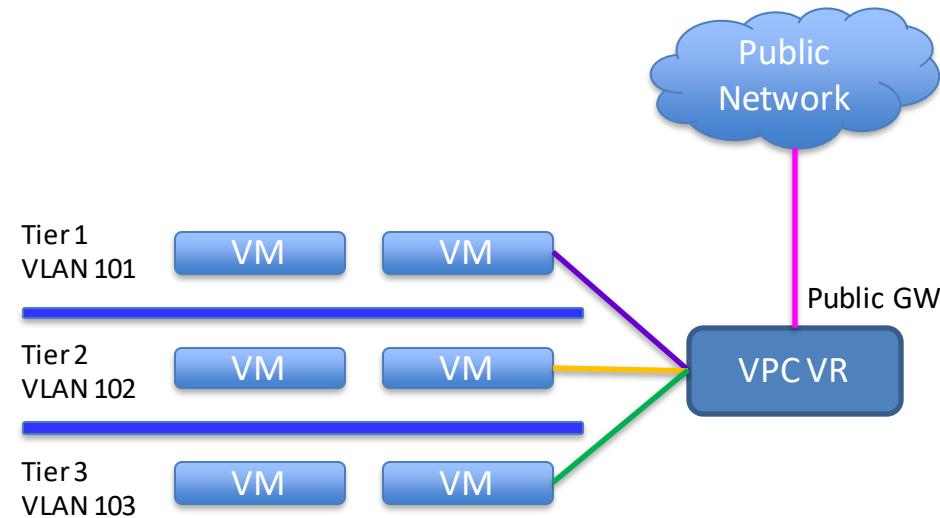
@ShapeBlue

VPC components

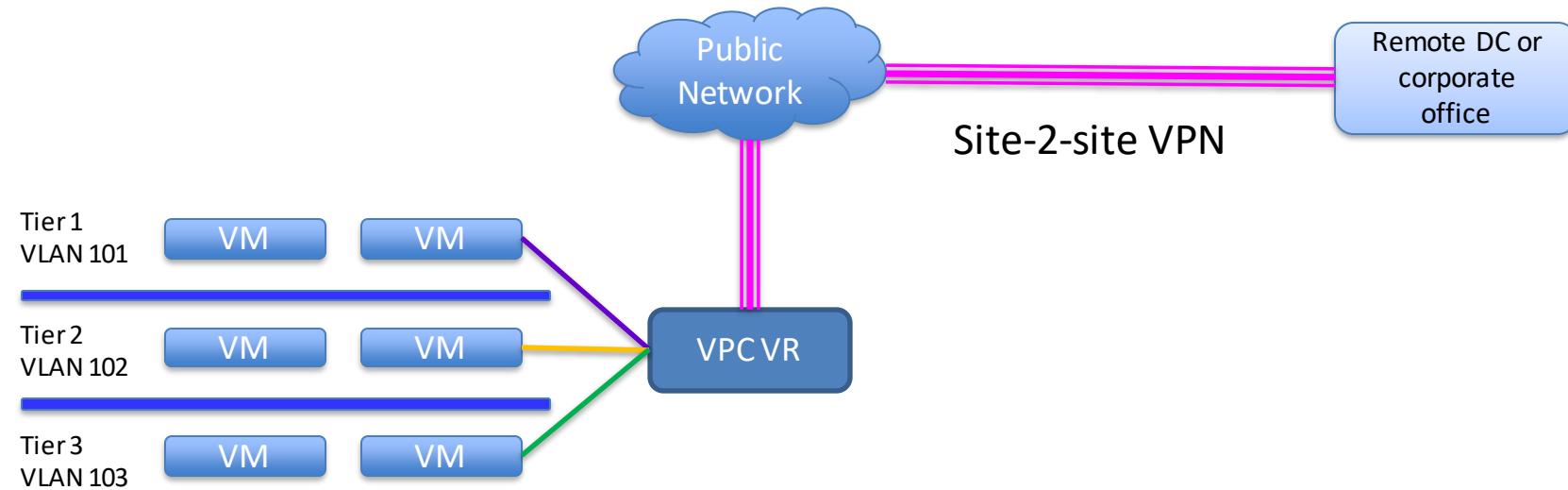


Network tiers – isolated networks,
each with unique VLAN and CIDR

VPC components



VPC components



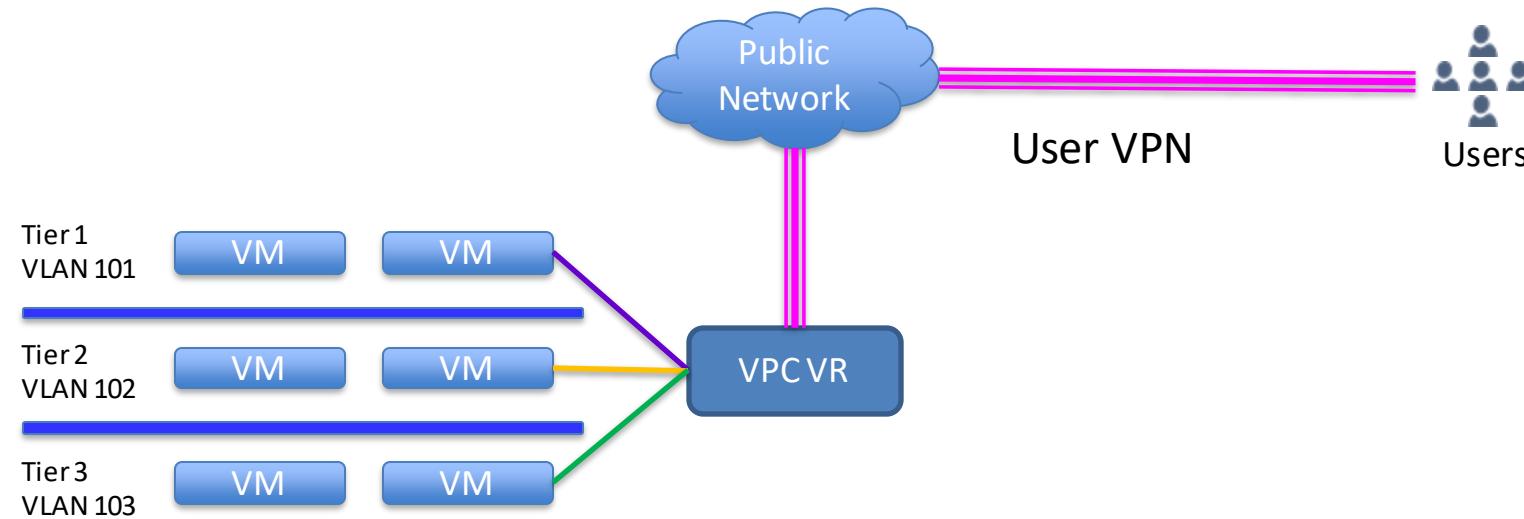
The Cloud Specialists

ShapeBlue.com

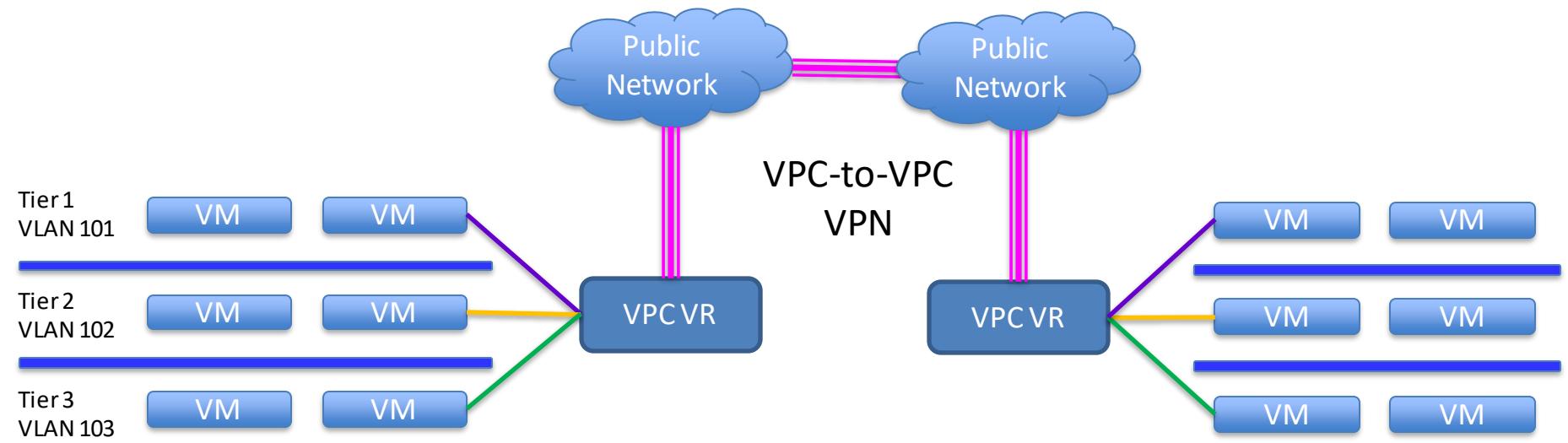


@ShapeBlue

VPC components



VPC components



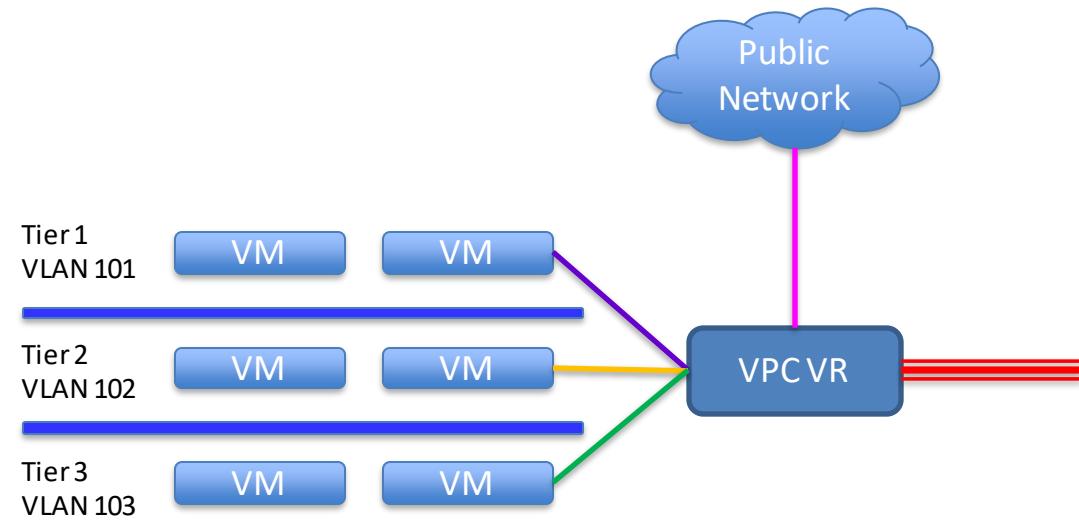
The Cloud Specialists

ShapeBlue.com



@ShapeBlue

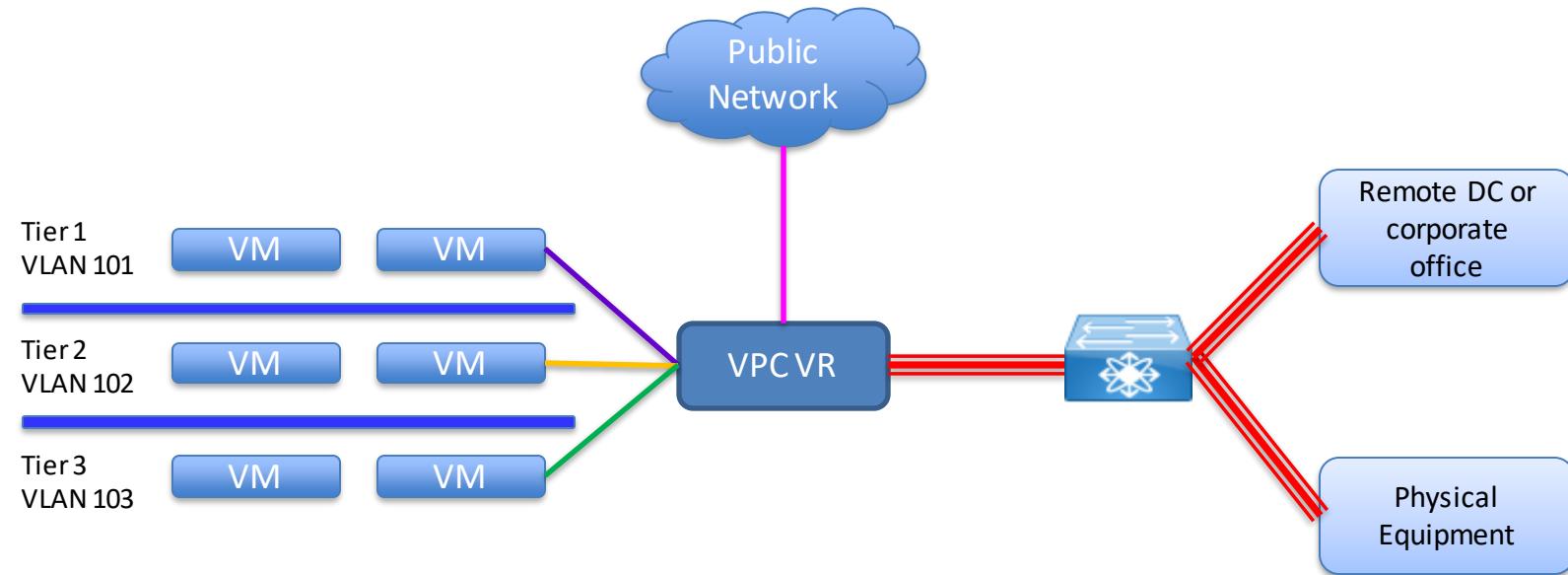
VPC Private Gateway



Private Gateway

- Created by root admins
- Static routes must be set on VPC by users

VPC Private Gateway



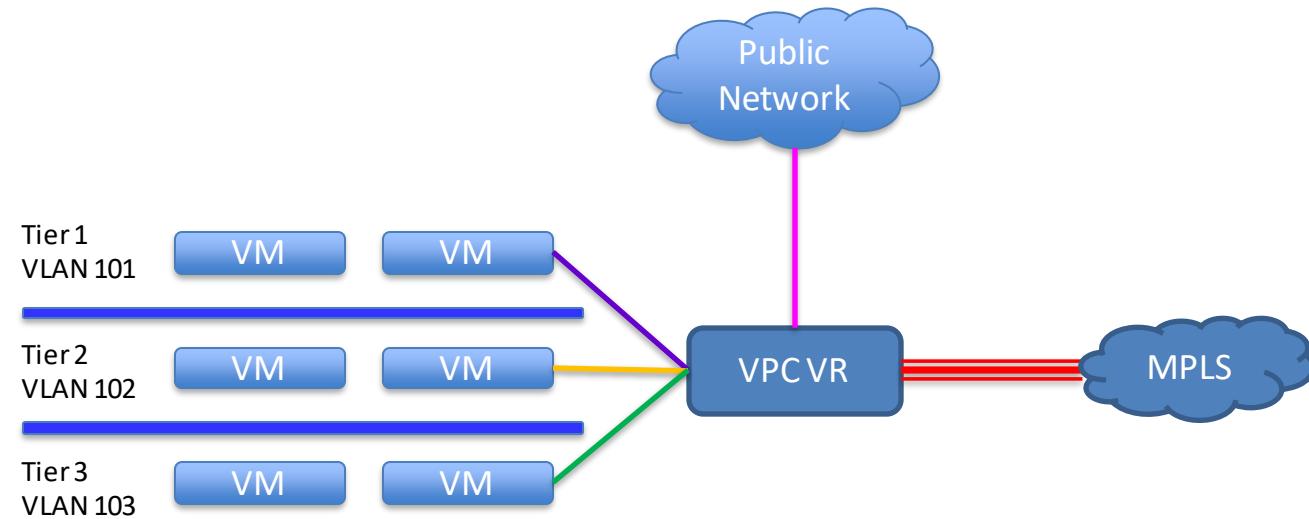
The Cloud Specialists

ShapeBlue.com

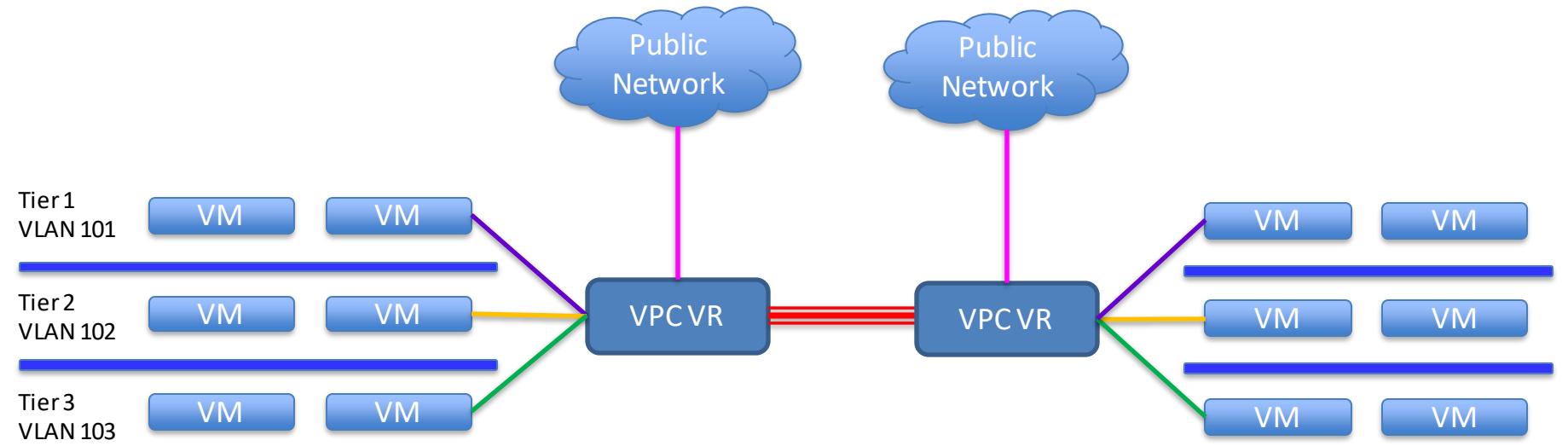


@ShapeBlue

VPC Private Gateway



VPC Private Gateway



Blue The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Exercise 9: Virtual Private Clouds



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

VPC – Optional Exercises

- **Create a VPC Tier with Internal LB and configure load balancing**
- **Create a Private Gateway and then setup some Static Routes**



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

VPC

Any questions?



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Using the API



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

About the API

- **CloudStack API is a RESTlike query API**
- **Uses http(s) get commands**
- **Authentication using APIkey and SecretKey to create a signature**
- **XML or JSON responses**
- **The Little CloudStack Book (Sebastien Goasguen):**
<https://github.com/runseb/cloudstack-books/blob/master/en/clients.markdown>
- **>15 Clients written for CloudStack API, in: Java, Python, Ruby, C#, php, Clojure**

- **The CloudStack GUI is built on the API**
- **Use the API to release the full power of CloudStack**
- **GUI does not expose all of the features of the API**



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Example Admin Use Cases

- **Access to full feature set**
- **Automation of complex tasks**
 - Testing new releases
- **Setup auto scaling**
- **Add additional physical networks**
- **Bulk actions such as create multiple VMs, client on-boarding**
- **Lock accounts**
- **Some things can only be done via the API:**
 - Create a VM for an alternate account
 - Specify host



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

- API key and secret key – unique to each user
- Signature hash
- Unauthenticated API Port (often set to 8096):
- NEVER expose the unauthenticated API Port to the Internet

Simple Structure of API Commands

- **http://URL:PORT/client/api?command=**
- **deployVirtualMachine**
- **serviceofferingid=nnnn**
- **templateid=nnnn**
- **zoneid=nnnn**
- **http://URL:PORT/client/api?command=deployVirtual
Machine&serviceofferingid=nnnn&templateid=nnnn&
zoneid=nnnn**

Asynchronous Commands

- Identified in the API Reference by an (A)
- Immediately return a job ID
- Typically API calls which take a long time to complete, e.g. `deployVirtualMachine`
- API call `queryAsyncJobResult` and pass the job ID to check status



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

- **API commands:**

<http://cloudstack.apache.org/docs/api/>

- **Developer guide:**

<http://docs.cloudstack.apache.org/en/latest/>

CloudMonkey



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

- **Cloudmonkey is a command line interface (CLI) tool for CloudStack written in Python**
- **Interactive shell or command line tool**
- **Supports CloudStack 4.0 and above**
- **Simpler lowercase commands**
- **Auto-completion**
- **Reverse Search**
- **Debug log**
- **Optional JSON, tabular and CSV output**

- **Argument Passing**
 - <cmd> <verb> key1=value1
 - [root@xenserver72 ~]# xe vm-listuuid (RO) : b2833cba-d7bd-2e8a-2423-d26705274e95 name-label (RW): s-1086-VM power-state (RO): runninguuid (RO) : e71db4ef-7c0f-df44-b7ec-6ba0a4ffc031 name-label (RW): v-1100-VM power-state (RO): runninguuid (RO) : 2667bc63-7f9a-4143-b952-e6f6f199e403 name-label (RW): Control domain on host: xenserver72 power-state (RO): running name=demo

- **Text processing using pipes**
- **Async jobs**

- Can progress straight onto next job without waiting

Exercise 10 (Optional): Using the API

Exercise 11: Cloudmonkey Familiarisation



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Troubleshooting



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

CloudStack Logs

- **The following logs are all used by CloudStack and are located in:**
 - /var/log/cloudstack/management
- **management-server.log**
 - Main CloudStack log – tracks all actions undertaken by this management server.
- **apilog.log**
 - Records every API and the resulting output
- **catalina.out**
 - Tomcat log



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Querying The Logs

- **Querying the Management Log for historical issues**
 - `grep -i -E 'exception|unable|fail|invalid|leak|warn|error'`
`/var/log/cloudstack/management/management-server.log`
 - `grep 'job-66'`
`/var/log/cloudstack/management/management-server.log`
- **Querying the Management Log for live issues**
 - `tail -f /var/log/cloudstack/management/management-server.log`



The Cloud Specialists

ShapeBlue.com

 @ShapeBlue

Accessing System VMs

- **XenServer / KVM (via host over the link local network)**
 - ssh onto the host which the VM is running on
 - ssh -i /root/.ssh/id_rsa.cloud -p 3922 root@169.254.n.n
-
- **VMware (via CloudStack management server)**
 - ssh onto the management Server
 - ssh -i /usr/share/cloudstack-common/scripts/vm/systemvm/id_rsa.cloud -p 3922 root@private-ip



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

- **System VMs are ‘stateless’ and load their configuration at boot time**
- **A restart does not load new settings so always stop/start**
- **If symptoms persist you can safely destroy any system VM and CloudStack will deploy a new instance with the latest configuration**

- **Implementing 3rd party monitoring solutions will be a significant improvement over the standard logging:**
 - Zenoss
 - ScienceLogic
 - Nagios
- **Pushing the management logs out to a Syslog server will make interpreting them a lot easier, especially when using multiple servers**

Working with CloudStack Databases



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Working with CloudStack Databases

- ***Manipulating the databases should only be done under the direction of experienced support personnel.***
- **When working with any database directly always make sure a full SQL export is taken before manipulating any of the data.**
 - `mysqldump -u cloud -p cloud > /tmp/cloud-$(date +%Y-%m-%d-%H.%M.%S).sql.bz2`
 - `mysqldump -u cloud -p cloud_usage > /tmp/cloud_usage-$(date +%Y-%m-%d-%H.%M.%S).sql.bz2`



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Demo: Working with CloudStack Database



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Exercise 12: Working with CloudStack Database



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

API, databases and troubleshooting

Any questions?



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

The Apache CloudStack Community



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

- **Contributor -> Committer -> PMC Member -> VP**
- **Developer**
 - Bug fixing
 - New feature Development
 - Testing releases
 - Voting

- **Contributor -> Committer -> PMC Member -> VP**
- **Non-Developer**
 - Testing
 - Bug reporting
 - Helping on user mailing list
 - Voting



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

- **Contributor -> Committer -> PMC Member -> VP**
- **Non-Technical**
 - Documentation
 - Marketing
 - Events



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Get Involved

- **Events**
 - CloudStack Collaboration Conference
 - Meetups
 - Build-a-cloud days
 - User groups

Resources

- **Website**
 - <http://cloudstack.apache.org/>
- **Wiki**
 - <https://cwiki.apache.org/confluence/display/CLOUDSTACK/Home>
 - Use with caution – sections easily get out of date
 - Contains design documentation
- **Docs**
 - <http://docs.cloudstack.apache.org/en/latest/>
 - <http://cloudstack.apache.org/docs/api/>

- **Mailing lists**
 - <http://cloudstack.apache.org/mailing-lists.html>
 - Announce, Users, Dev, Commits, Issues, Marketing
 - *If it's not on the mailing list – it didn't happen*
- **IRC**
 - On Freenode
 - #cloudstack: This channel is for general cloudstack discussion and support.
 - #cloudstack-dev: This channel is for developer discussions and support.

Resources

- **Jira**

<https://issues.apache.org/jira/browse/CLOUDSTACK/>

- Bug reporting and tracking

- **Jenkins**

<http://jenkins.buildacloud.org/>

- Automated building and testing of code



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Resources

- **Developers**

<http://cloudstack.apache.org/developers.html>

- **Non-Developers**

<http://cloudstack.apache.org/contribute.html>

- **Meet-ups and Events**

<http://lanyrd.com/topics/apache-cloudstack/>



The Cloud Specialists

ShapeBlue.com



@ShapeBlue

Feedback

- Please provide feedback, its very important to help us improve the course
<https://www.surveymonkey.com/s/YPWM5QY>
- This Bootcamp trainer was: Dag Sonstebo
- No feedback = No certificate ☺

Apache CloudStack Bootcamp

Dag Sonstebo

Cloud Architect

dag.sonstebo@shapeblue.com

Twitter: [@dagsonstebo](https://twitter.com/dagsonstebo)