

# VMS Clusters on the Cloud and Hybrid VMS Clusters

Jonathan Bergdahl & Martin Schneider

Date: 14/5 2025

# Part 1: VMS Clusters on the Cloud

Jonathan Bergdahl & Martin Schneider  
Date: 14/5 2025

# Agenda Part 1

**01**

Introduction

**02**

System Topology

**03**

Clusters in one OCI region

**04**

Local VCN Peering

**05**

Clusters across OCI regions

**06**

Failure Scenarios

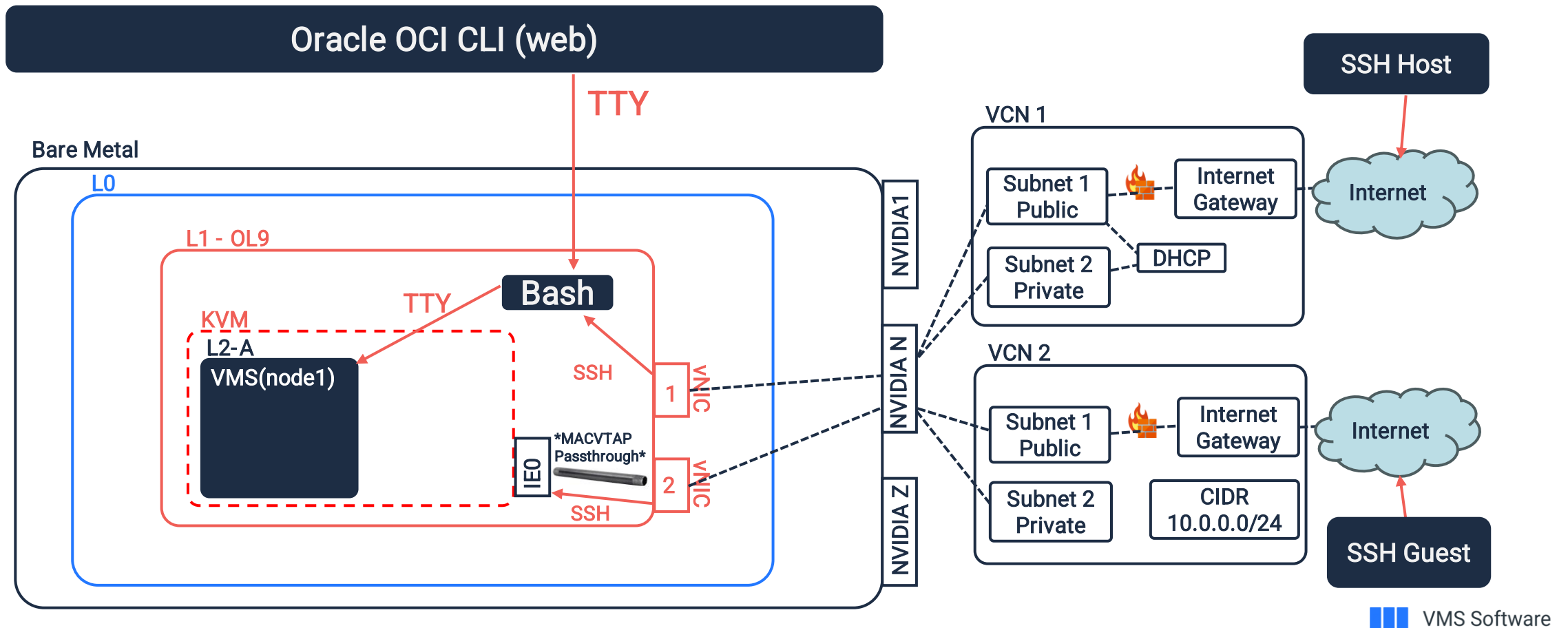
**07**

Summary

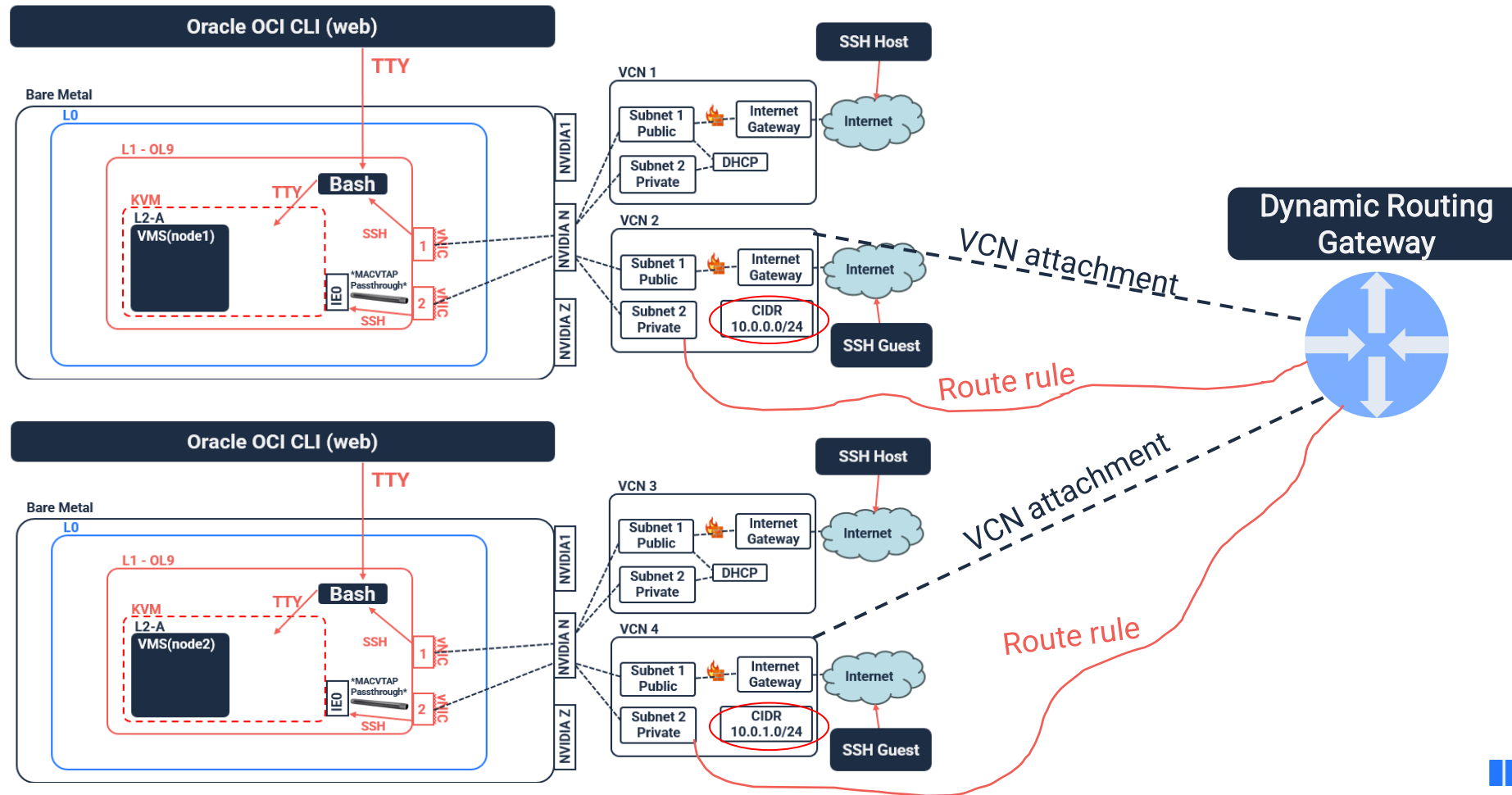
# Presentation Objectives

- This presentation will:
  - Show the process of setting up a VMS cluster on the cloud.
  - OCI Networking and tunnelling
  - Share our learned lessons and roadblocks
  - Motivate you to try it out yourself!
- This presentation assumes you have:
  - Watched the previous webinar:
    - How to set up and configure your OCI instance
    - Details about KVM configuration
    - VMS guest in KVM
- Let's get started!

# System Topology – a quick recap from last webinar



# Cloud Cluster in the same OCI Region



# DRG creation (local peering)

Networking > Customer connectivity > Dynamic routing gateways

Customer connectivity

- Overview
- Site-to-Site VPN
- FastConnect
- Dynamic routing gateway**
- Customer-premises equipment

## Dynamic routing gateways

Dynamic routing gateways (DRGs) are optional virtual routers that you can add to your VCN. They provide a path for private network traffic between your VCN and on-premises network.

Create dynamic routing gateway

Name	Lifecycle state	Oracle redundancy status ⓘ	Created
<a href="#">DRG</a>	Available	—	Thu, Jan 16, 2025, 12:01:46 UTC

Networking > Customer connectivity > Dynamic routing gateways > DRG

## DRG

DRG

AVAILABLE

Resources

- VCN attachments (2)**
- Virtual circuit attachments (0)
- IPSec tunnel attachments (0)
- Remote peering connection attachments (1)
- Loopback attachments (0)

Dynamic routing gateway information

Compartment: vsi (root)/internal/germany/test/Martin OCID: ...nevrh7mqz [Show](#) [Copy](#)

Oracle redundancy status: — Created: Thu, Jan 16, 2025, 12:01:46 UTC

## VCN attachments in Martin Compartment

VCNs are connected to a DRG by an attachment with the VCN type. You can configure all VCNs to use the same route table. [Learn more.](#)

Create virtual cloud network attachment

Attachment name	Lifecycle state	Virtual cloud network	DRG route table	VCN route type	Created
<a href="#">drgattachment20250116120241</a>	Attached	<a href="#">VCN2</a>	<a href="#">Autogenerated Drg Route Table for VCN attachments</a>	Subnet CIDR blocks	Thu, Jan 16, 2025, 12:02:41 UTC

## Create dynamic routing gateway

[Help](#)

Name

Create in compartment

Martin

vsi (root)/internal/germany/test/Martin

[Show Advanced options](#)

Create dynamic routing gateway [Cancel](#)

## Create VCN attachment

[Help](#)

Attachment name *Optional*

Virtual cloud network in **Martin** [\(Change compartment\)](#)

Select a VCN

VCN-W-W

VCN2

WVCCNN

Create VCN attachment [Cancel](#)

# DRG creation (local peering)

- Define each subnet in the route rules

## Route Rules

Traffic within the VCN is handled by the VCN's local routing by default. Intra-VCN routing allows you more control over routing between subnets. [Learn more](#). If you're having problems, use [Network Path Ana](#)

<div>Add Route Rules Edit Remove</div>					
<input type="checkbox"/>	Destination	Target Type	Target	Route Type	De
<input type="checkbox"/>	0.0.0.0/0	Internet Gateway	<a href="#">VCN2-GATEWAY</a>	Static	
<input type="checkbox"/>	10.2.0.0/24	Dynamic Routing Gateways	<a href="#">DRG</a>	Static	
<input type="checkbox"/>	10.6.0.0/24	Dynamic Routing Gateways	<a href="#">DRG</a>	Static	
<input type="checkbox"/>	10.8.0.0/24	Dynamic Routing Gateways	<a href="#">DRG</a>	Static	
0 selected					

CIDR of network  
we want to peer



# Moving it up a notch: stretched cluster

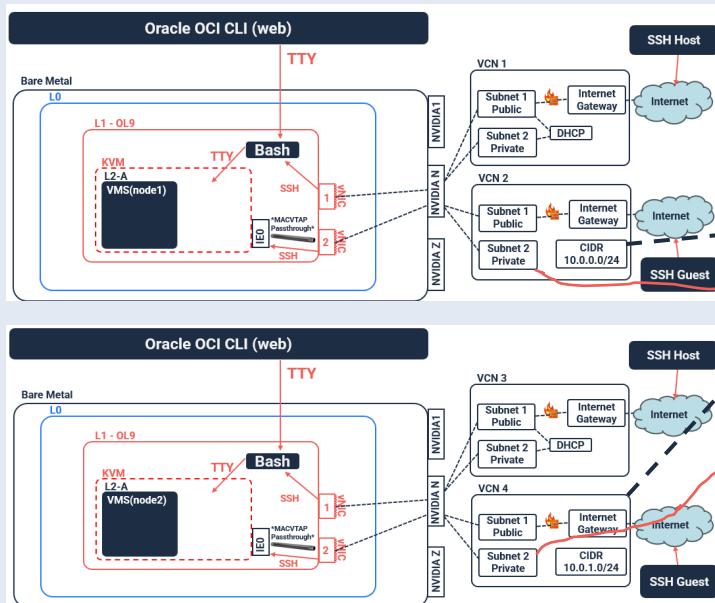
- From 2 nodes in the same Region

To

- 2 nodes in one region, 2 nodes in another region

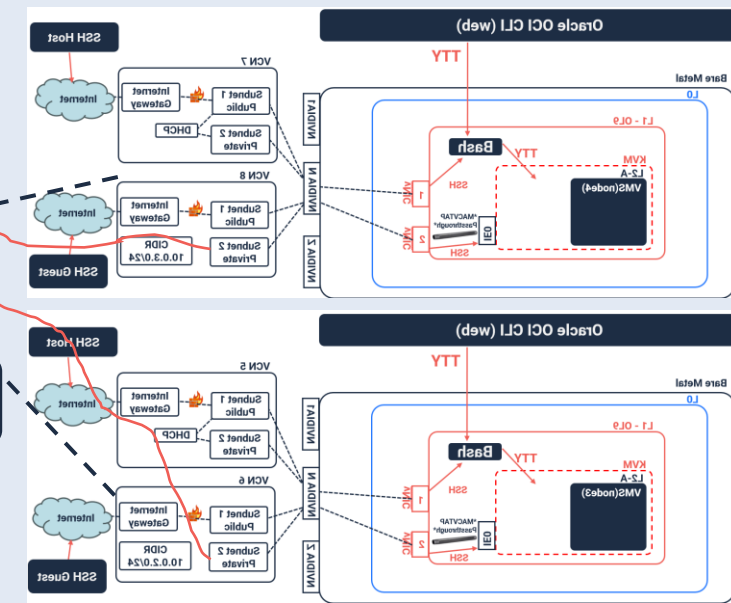
# Cloud Cluster across OCI Regions

## OCI Region: Frankfurt-1



Dynamic Routing Gateway 1

## OCI Region: Chicago-1



Dynamic Routing Gateway 2

Remote Peering Connection (RPC)

# RPC creation (global peering)

Networking > Customer connectivity > Dynamic routing gateways > DRG > Remote peering connection attachments

## DRG

[Edit](#) [Add tags](#) [Move resource](#) [Terminate](#)

**Dynamic routing gateway information** [Tags](#)

**Compartment:** vsi (root)/internal/europe/test/Martin **OCID:** ...nevhrh7mzq [Show](#) [Copy](#)  
**Oracle redundancy status:** — **Created:** Thu, Jan 16, 2025, 12:01:46 UTC

### Resources

VCN attachments (2)  
Virtual circuit attachments (0)  
IPSec tunnel attachments (0)  
**Remote peering connection attachments (0)**  
Loopback attachments (0)

## Remote peering connection attachments in Martin Compartment

Remote peering connection (RPC) attachments are automatically created when an RPC is created. You can't directly create additional attachments for an RPC.

[Create remote peering connection](#)

Attachment name	Lifecycle state	DRG route table	Remote peering connection	Peering status	Created
No items found.					

Showing 0 items < 1 of 1 >

## RPC

[Establish connection](#) [Edit](#) [Terminate](#)

**Remote peering connection Information**

**Compartment:** vsi (root)/internal/europe/test/Martin **OCID:** ...2cezxgvora [Show](#) [Copy](#)  
**DRG OCID:** ...nevhrh7mzq [Show](#) [Copy](#) **Created:** Tue, Feb 11, 2025, 09:02:13 UTC  
**Peer status:** ● New (not peered) **Cross-tenancy:** No  
**Peer region:** — **Peer tenancy OCID:** —  
**Peer connection OCID:** —

## Create remote peering connection

[Help](#)

**Name**  
RPC

**Create in compartment**  
Martin

vsi (root)/internal/europe/test/Martin

**Info** This creates an attachment to the selected DRG. The attachment uses a route table based on the type of resource using the attachment.

[Show Advanced options](#)

[Create remote peering connection](#) [Cancel](#)

## Establish connection

[Help](#)

**Region**  
eu-frankfurt-1

**Remote peering connection OCID**  
ocid1.remotepeeringconnection.oc1.eu-frankfurt-1.aaaaaaatd22i

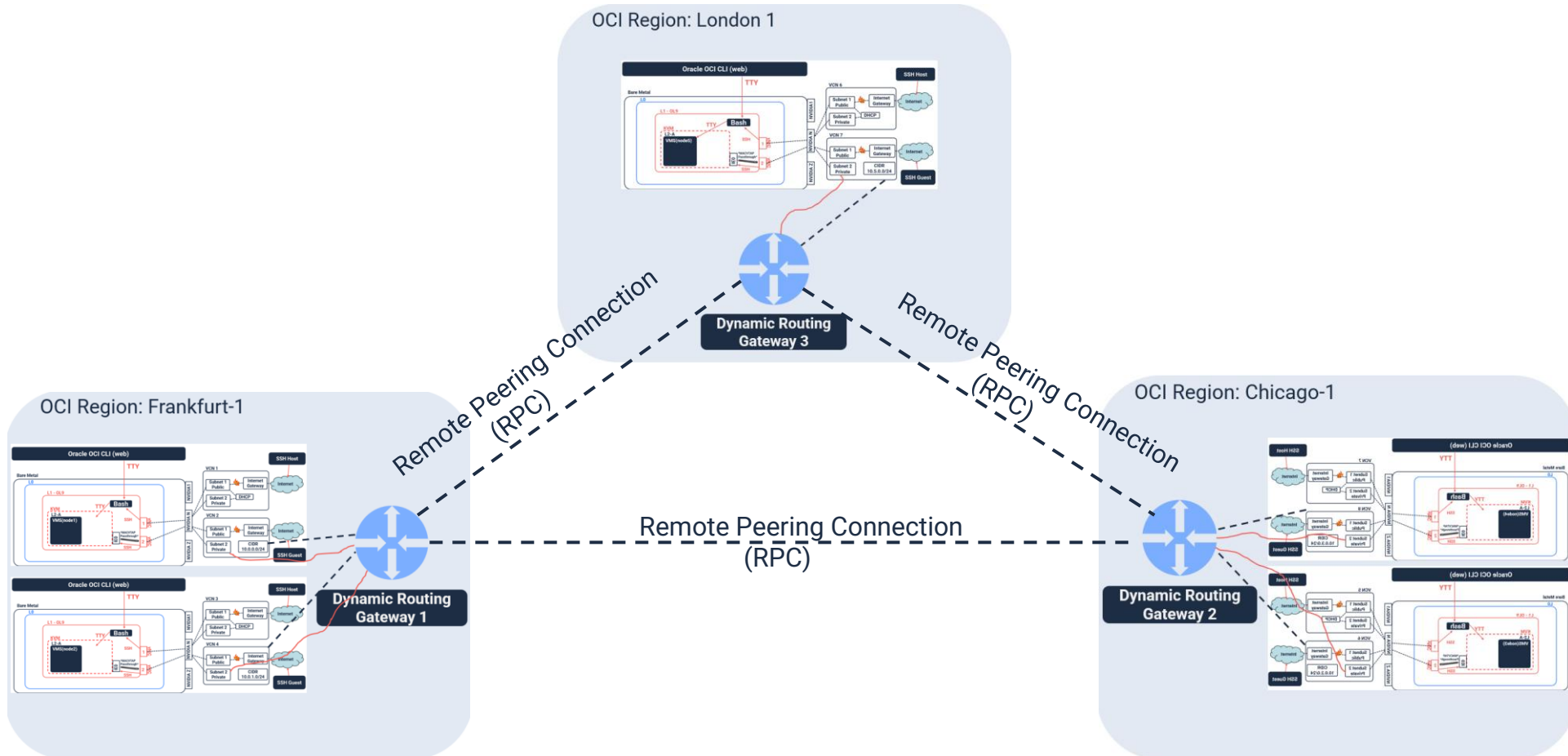
[Establish connection](#) [Cancel](#)

# Improving the cluster

## Issues:

- 2 + 2 cluster
  - Even number total
  - Even number in both sites
  - No quorum disk
- Single point of failure between 2 sites (peering going down)

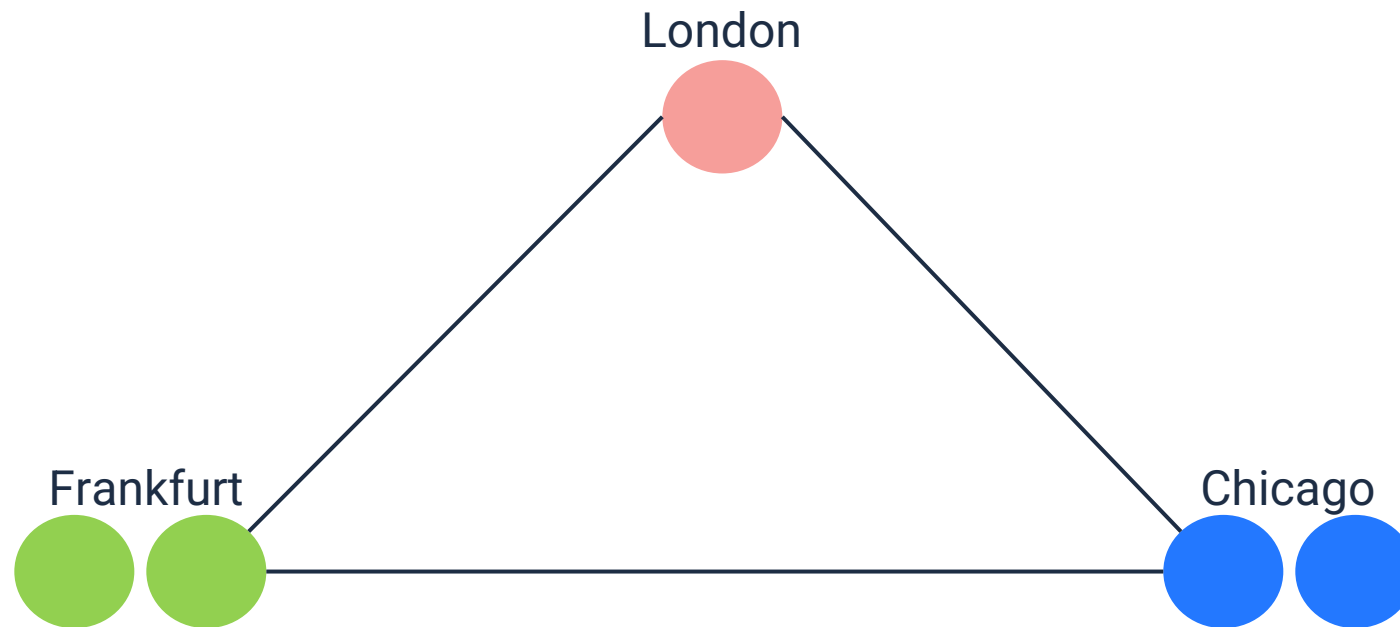
# Third Region, for Cluster improvement



# Cluster config in VMS

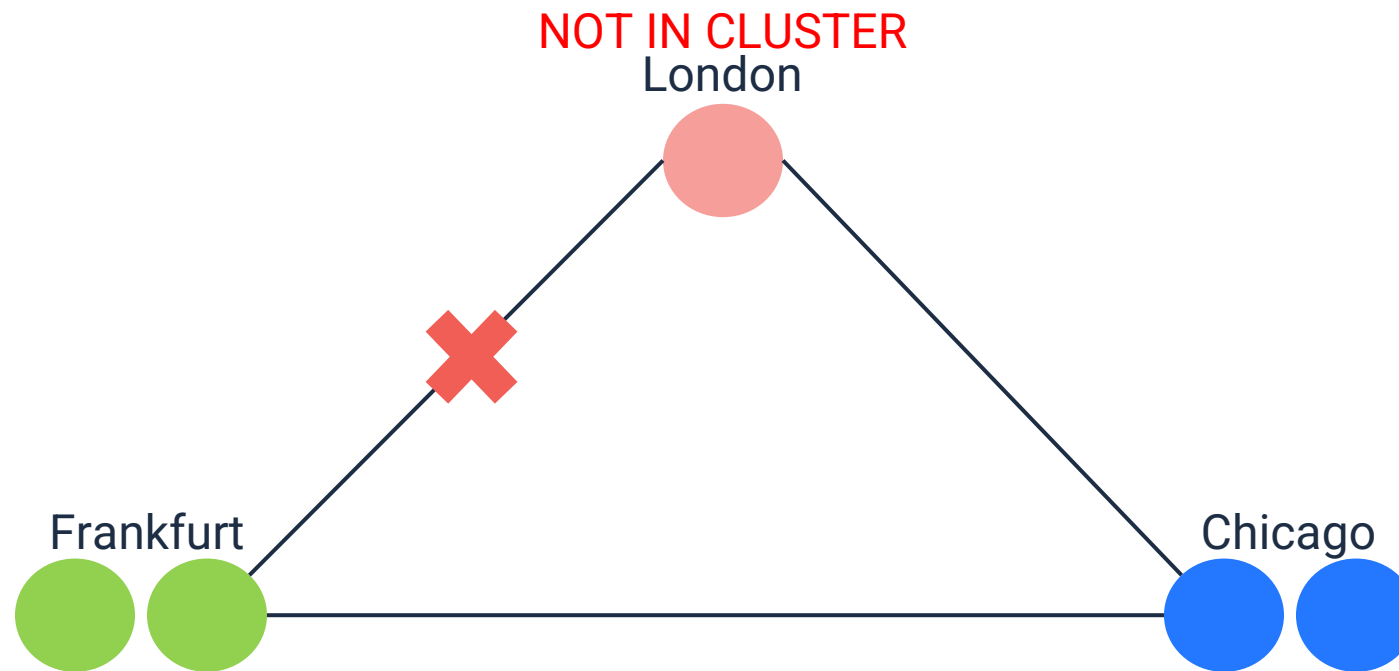
- ▶ EXPECTED\_VOTES is set to 5 (Minimum votes to meet quorum will be 3)
  - Each node contributes with one vote
- ▶ UDP/IP communication is used for clustering (IPCI) – No bare SCS
- ▶ Use Unicast, not Multicast

# Simplified Topology



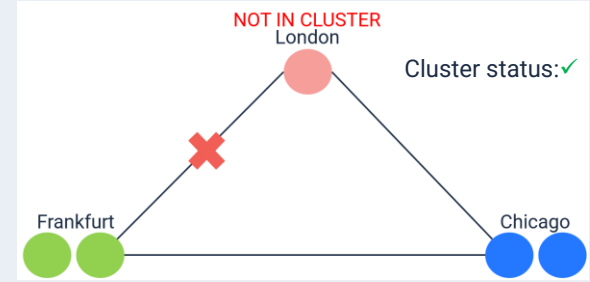
# RPC Failure Scenarios

Cluster status: ✓





# RPC Failure Scenarios



- What happens when one of the RPCs goes down?
- The node(s) in one region will crash
- Cluster will reform, and stay up
  - Frankfurt contributes with more votes
- Behaviour if RPC between Frankfurt and Chicago goes down?

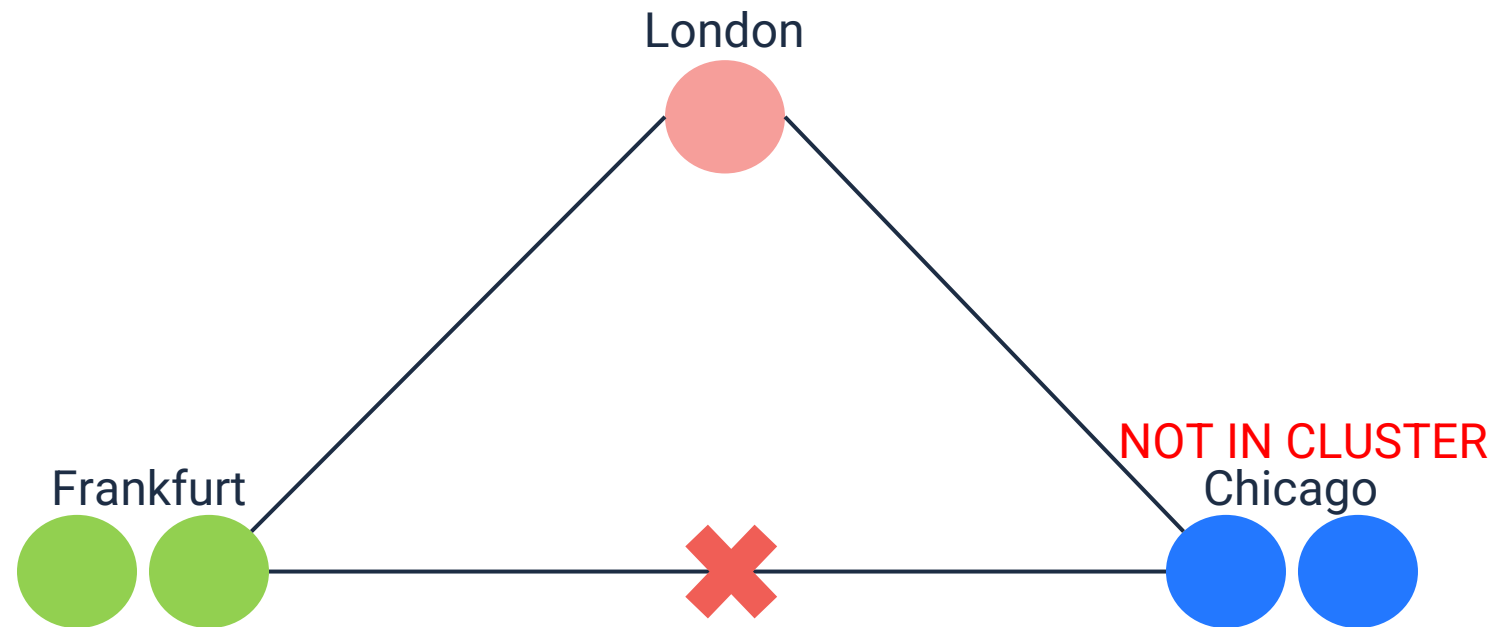
```
View of Cluster from system ID 1025 node: FRANK1 17-FEB-2025 15:40:37
```

SYSTEMS		MEMBERS
NODE	SOFTWARE	STATUS
FRANK1	VMS V9.2-3	MEMBER
FRANK2	VMS V9.2-3	MEMBER
UK1	VMS V9.2-3	BRK_NON
CHIC1	VMS V9.2-3	MEMBER
CHIC2	VMS V9.2-3	MEMBER

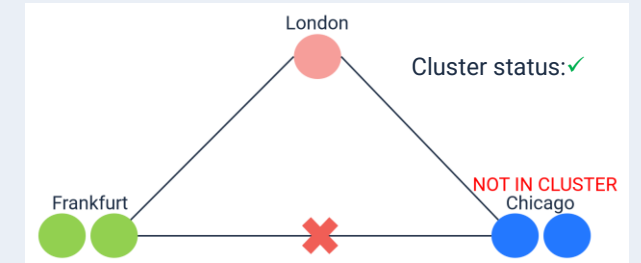
```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
15:40:36.85 Node FRANK1 (csid 00010001) timed-out lost connection to node UK1
```

# RPC Failure Scenarios

Cluster status: ✓



# RPC Failure Scenarios



- What happens when the RPC between Frankfurt and Chicago goes down?
  - 3/5
  - 3/5
- Bringing the RPC up again results in Chicago (re)joining

View of Cluster from system ID 1025 node: FRANK1

SYSTEMS		MEMBERS
NODE	SOFTWARE	STATUS
FRANK1	VMS V9.2-3	MEMBER
FRANK2	VMS V9.2-3	MEMBER
UK1	VMS V9.2-3	MEMBER
CHIC1	VMS V9.2-3	BRK_NON
CHIC2	VMS V9.2-3	BRK_NON

View of Cluster from system ID 1029 node: UK1

SYSTEMS		MEMBERS
NODE	SOFTWARE	STATUS
UK1	VMS V9.2-3	MEMBER
CHIC2	VMS V9.2-3	NEW
CHIC1	VMS V9.2-3	NEW
FRANK1	VMS V9.2-3	MEMBER
FRANK2	VMS V9.2-3	MEMBER

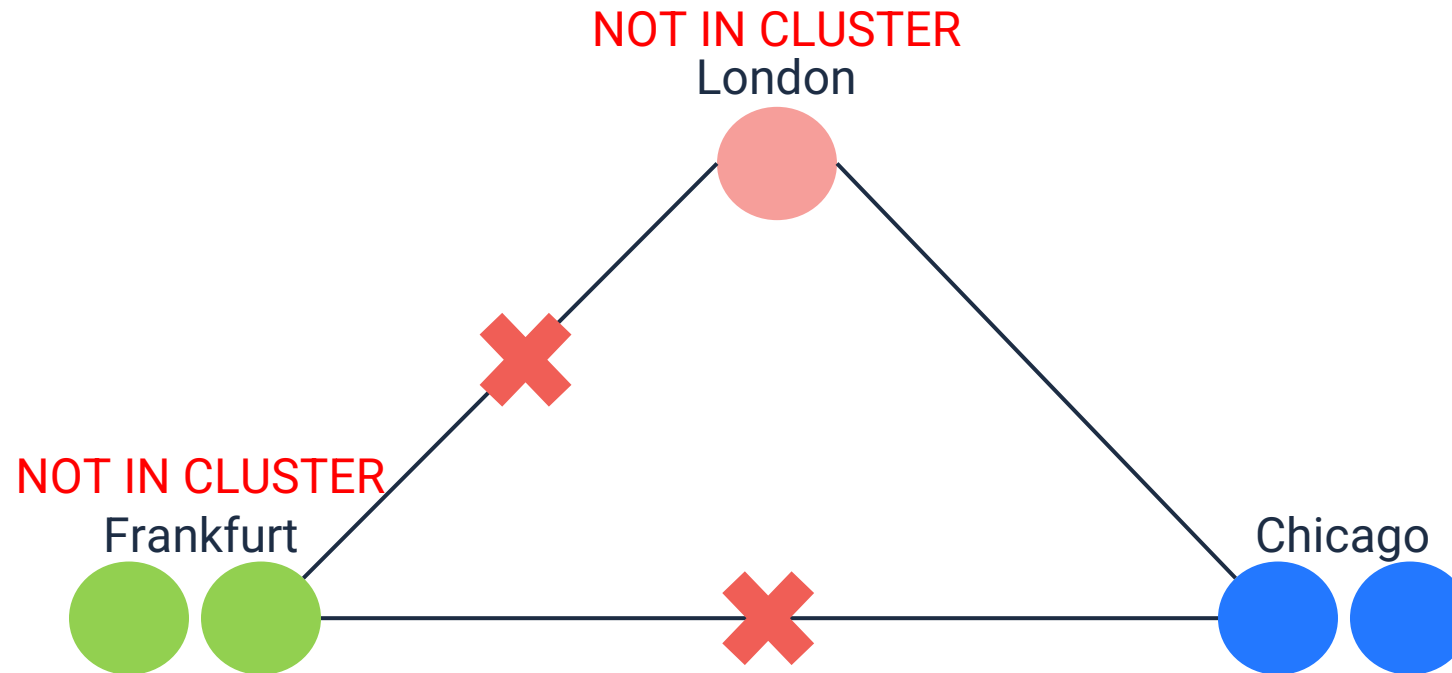
When RPC is down (Frankfurt – Chicago)

When RPC is brought back up

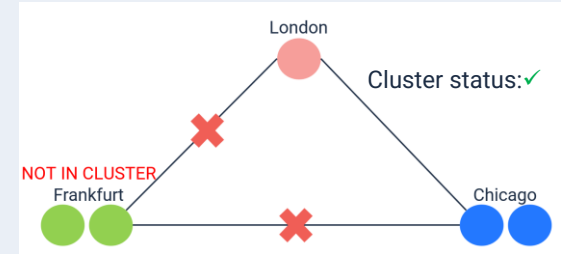
```
%PEA0, Cluster communication successfully initialized on IP bus IE0 10.3.0.92
%CNXMAN, Have connection to system CHIC2
%CNXMAN, Have connection to system UK1
%CNXMAN, Discovered system FRANK1
%CNXMAN, Established connection to system FRANK1
%CNXMAN, Discovered system FRANK2
%CNXMAN, Established connection to system FRANK2
%CNXMAN, Sending VMScluster membership request to system UK1
%CNXMAN, Sending VMScluster membership request to system UK1
%CNXMAN, Now a VMScluster member -- system CHIC1
%STDV-I-STARTUP, OpenVMS startup begun at 18-FEB-2025 08:24:22.41
```

# RPC Failure Scenarios

Cluster status: ✓



# RPC Failure Scenarios



- What happens when 2 RPCs go down?
- Previous cluster: Frankfurt – Chicago
- When their RPC goes down: Chicago - London
- Chicago nodes have connection to each other + UK
  - 3/5
  - 2/5

```
Copyright 2024 VMS Software, Inc.

%DECnet-I-LOADED, network base image loaded, version = 05.92.07
%VMScluster-I-LOADIPICCFG, loading IP cluster configuration files
%VMScluster-S-LOADEDIPICCFG, Successfully loaded IP cluster configuration files

%SMP-I-CPUTRN, CPU #1 has joined the active set.
%SMP-I-CPUTRN, CPU #2 has joined the active set.
%VMScluster-I-LOADSECDB, loading the cluster security database
%EIA0, Link up: 1000 mbit, 02-00-17-0C-26-A6
%MSCPLOAD-I-CONFIGSCAN, enabled automatic disk serving
%SYSINIT-I- waiting to form or join an OpenVMS Cluster
%PEA0, Configuration data for IP cluster found
%PEA0, Successfully allocated IP vector
%PEA0, Cluster communication enabled on IP bus, IE0, 10.5.0.44
%PEA0, Successfully initialized with TCP/IP Services
%PEA0, Unicast list for IP bus IE0, added remote node address 10.4.0.92
%PEA0, Unicast list for IP bus IE0, added remote node address 10.3.0.92
%PEA0, Unicast list for IP bus IE0, added remote node address 20.0.0.210
%PEA0, Unicast list for IP bus IE0, added remote node address 10.1.0.61
%PEA0, Hello message sent on IP bus IE0
%PEA0, Cluster communication successfully initialized on IP bus IE0 10.5.0.44
%CNXMAN, Have connection to system CHIC1
%CNXMAN, Have connection to system CHIC2
```


Cluster will reform when Frankfurt leaves



# Recreating these tests

- Go to OCI web console
- Navigate to Dynamic Routing Gateways -> DRG\_name -> Remote Peering Connections -> RPC\_name

Networking > Customer connectivity > Dynamic routing gateways > myChicDRG1 > Remote peering connections > chicFrankRPC



AVAILABLE

chicFrankRPC

[Establish connection](#) [Edit](#) [Terminate](#)

**Remote peering connection Information**

**Compartment:** vsi (root)/internal/us/test/Jonathan  
**DRG OCID:** ...u6axobzfmq [Show](#) [Copy](#)  
**Peer status:** ● Peered  
**Peer region:** eu-frankfurt-1  
**Peer connection OCID:** ...p6ifs4vpsq [Show](#) [Copy](#)

**OCID:** ...72u2kechjq [Show](#) [Copy](#)  
**Created:** Mon, Feb 17, 2025, 09:27:33 UTC  
**Cross-tenancy:** No  
**Peer tenancy OCID:** ...p5e3vkelgq [Show](#) [Copy](#)

**Resources**

[Remote peering connection attachments \(1\)](#)

**Remote peering connection attachments**

Attachment name	Lifecycle state	Dynamic routing gateways	DRG route table	Created
<a href="#">DRG Attachment for RPC: chicFrankRPC</a>	● Attached	<a href="#">myChicDRG1</a>	<a href="#">Autogenerated Drg Route Table for RPC, VC, and IPSec attachments</a>	Mon, Feb 17, 2025, 09:27:39 UTC

Showing 1 item < 1 of 1 >

Terminate the RPC

# Node Failure Scenarios

- How many individual nodes can go down?
- Which nodes are allowed to go down?
- It does not matter which region the node(s) reside in
- Recreate? Just shut down or crash your nodes!

View of Cluster from system ID 1026 node: FRANK2

SYSTEMS		MEMBERS
NODE	SOFTWARE	STATUS
FRANK2	VMS V9.2-3	MEMBER
FRANK1	VMS V9.2-3	BRK_NON
UK1	VMS V9.2-3	MEMBER
CHIC1	VMS V9.2-3	MEMBER
CHIC2	VMS V9.2-3	BRK_NON

Example: nodes in different regions going down



# Things to consider



Boot times



Inter-region latency



DRG creation can be scripted!

Region	N	MAD	MEL	MRS	MTY	MTZ	NRT	ORD	PHX	QRO
AMS	6	28.13	258.67	23.78	123.77	69.84	233.43	92.37	134.03	136.37
ARN	11	52.03	282.47	43.78	147.37	96.10	257.25	116.13	157.66	159.39
AUH	6	88.88	160.34	74.60	202.39	151.47	146.73	177.42	210.91	213.98
BOG	8	181.01	280.79	170.24	82.99	230.80	224.14	106.95	104.46	73.35
BOM	12	109.75	139.57	94.83	231.31	172.55	126.25	206.40	242.51	242.84
CDG	4	21.18	251.81	10.43	118.59	70.33	238.51	93.88	127.12	130.21
CWL	9	30.70	261.05	22.11	120.37	75.81	231.36	89.75	131.41	132.93
DXB	6	92.45	158.47	78.13	205.93	154.99	144.96	180.96	214.42	218.25
FRA	12	31.44	250.96	17.36	129.11	63.78	238.65	97.58	139.10	141.73
GRU	2	205.12	323.85	194.63	157.58	253.65	266.68	125.31	166.51	170.06

# Summary

- ▶ Cross-region VMS clusters on OCI achieved through DRGs
- ▶ Scalability potential!
- ▶ Cluster behaviour on the cloud

# Part 2: Hybrid VMS Clusters

Jonathan Bergdahl & Martin Schneider

Date: 14/5 2025

# Agenda Part 2

**01**

Introduction

**02**

System Topology

**03**

OCI Networking setup

**04**

Wireguard Configuration

**05**

Cluster and TCPIP configuration

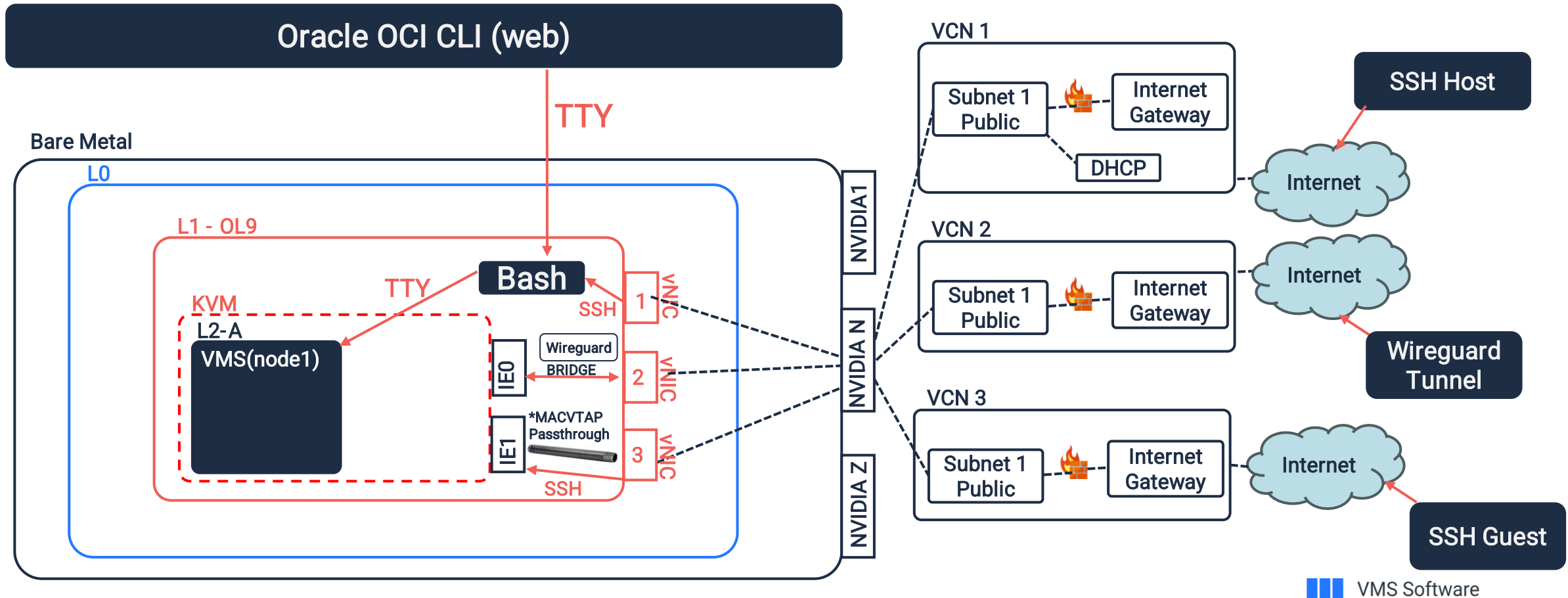
**06**

Challenges and Pitfalls

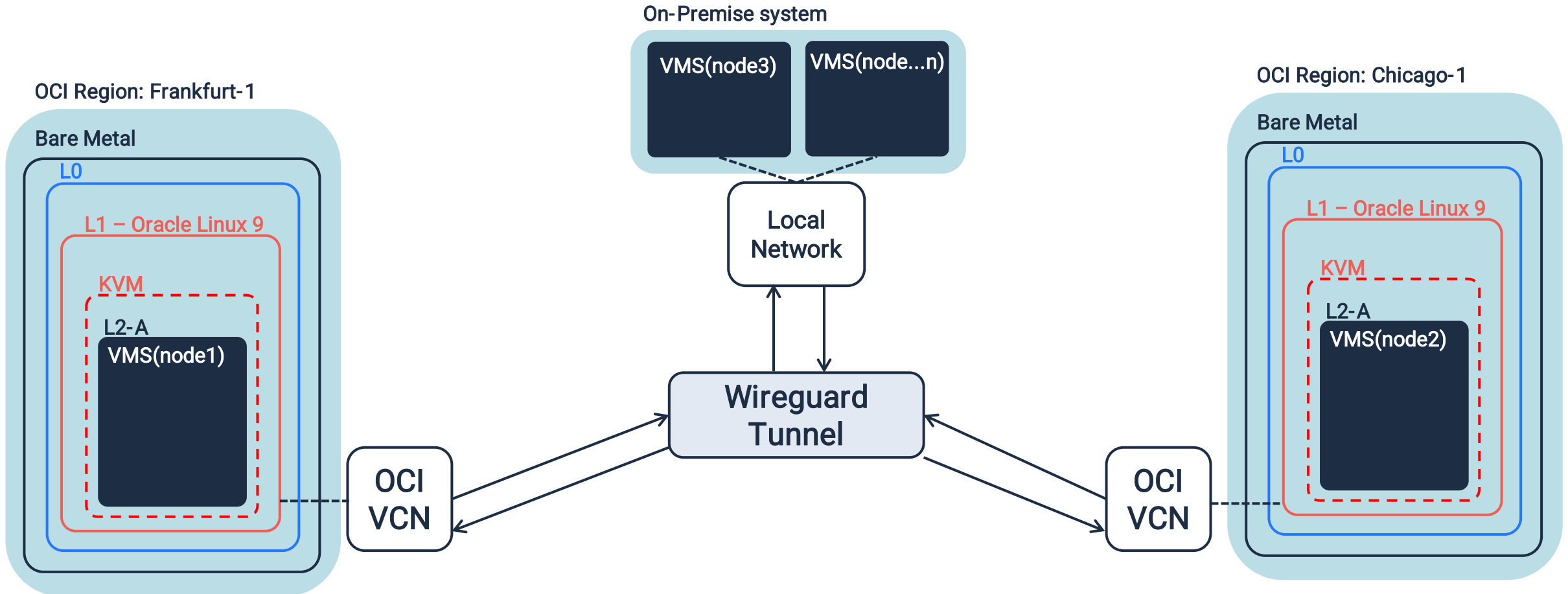
**07**

Q&A

# System Topology



# Hybrid Cluster Topology



# Configuring a Bridge Interface in the Cloud

- Secondary VNIC
- OCI networking under the hood
- Primary VNIC routes must always have a lower metric
- Can be done with a script!

```
[opc@myprimvnic ~]$ sudo brctl show
bridge name      bridge id        STP  enabled  interfaces
br0               8000.0200170599cb  no   enabled  enp2s0
```

# OCI Networking Challenges

- Under the hood port forwarding
- Firewall
- Route rules and metrics
- Unmanaged network interfaces

```
[opc@cluster-frank-1 ~]$ cat fix_net.sh
#!/bin/sh
#
#####
#           MAGIC SECOND vNIC SCRIPT           #
#####
#
# Quick and dirty script to get the secondary vNIC#
# working on OCI. Might take several tries to   #
# ping with both cards.                         #
# Keep in mind that ping can be unreliable, and #
# that you need to have ICMP allowed on your    #
# firewall.                                     #
#
# If it fails to unconfigure, make sure you have #
# all network connected virtual machines stopped #
# and have tools like WireGuard turned off.      #
#
#####
```



# Configuring Wireguard

- Various route rules
- Ensuring the correct VNIC is used
- Allowed IPs

Static route to endpoint(s)

```
[Interface]
Address = 192.168.2.10/24
SaveConfig = true
PostUp = iptables -A FORWARD -o wg0 -m conntrack --ctstate RELATED,ESTABLISHED -j ACCEPT
PostUp = ip route add 10.0.2.1/32 via 10.0.2.1 dev br0
PostUp = iptables -A FORWARD -i br0 -o wg0 -j ACCEPT
PostUp = iptables -A FORWARD -i wg0 -o br0 -m conntrack --ctstate RELATED,ESTABLISHED -j ACCEPT
PostUp = iptables -t nat -A POSTROUTING -o br0 -j MASQUERADE
PostUp = iptables -t nat -A POSTROUTING -o wg0 -j MASQUERADE
PostDown = ip route del 10.0.2.1/32 via 10.0.2.1 dev br0
PostDown = iptables -t nat -D POSTROUTING -o br0 -j MASQUERADE
PostDown = iptables -t nat -D POSTROUTING -o wg0 -j MASQUERADE
PostDown = iptables -D FORWARD -o wg0 -m conntrack --ctstate RELATED,ESTABLISHED -j ACCEPT
PostDown = iptables -D FORWARD -i br0 -o wg0 -j ACCEPT
PostDown = iptables -D FORWARD -i wg0 -o br0 -m conntrack --ctstate RELATED,ESTABLISHED -j ACCEPT
ListenPort = 30300
PrivateKey = [REDACTED]

[Peer]
PublicKey = [REDACTED]
AllowedIPs = 192.168.2.1/32, 10.1.0.61/32, 10.1.0.62/32
Endpoint = [REDACTED]:51820
```

ALL IPs of peer

```
[opc@myprimvnic ~]$ sudo wg show
interface: wg0
  public key: [REDACTED]
  private key: (hidden)
  listening port: 30300

peer: [REDACTED]
  endpoint: [REDACTED]:51820
  allowed ips: 192.168.2.1/32, 10.1.0.61/32, 10.1.0.62/32
  latest handshake: 1 minute, 48 seconds ago
  transfer: 103.38 MiB received, 103.67 MiB sent
```

# MacVTap

- Passthrough to the guest
- Disable Decnet Phase IV addressing!
  - @NET\$CONFIGURE ADVANCED
- Used for all non-cluster communication

## Virtual Network Interface

Network source:

Device name:

 In most configurations, macvtap does not work for host to guest network communication.

Device model:

MAC address: 02:00:17:12:7d:03

```
* Data Link name to use for EIB0 (EI/82558)?      [CSMACD-1] :  
* Routing Circuit Name for Data Link 'CSMACD-1'? [CSMACD-1] :  
* Enable Phase-IV Addressing on Routing Circuit 'CSMACD-1'? [NO] :
```

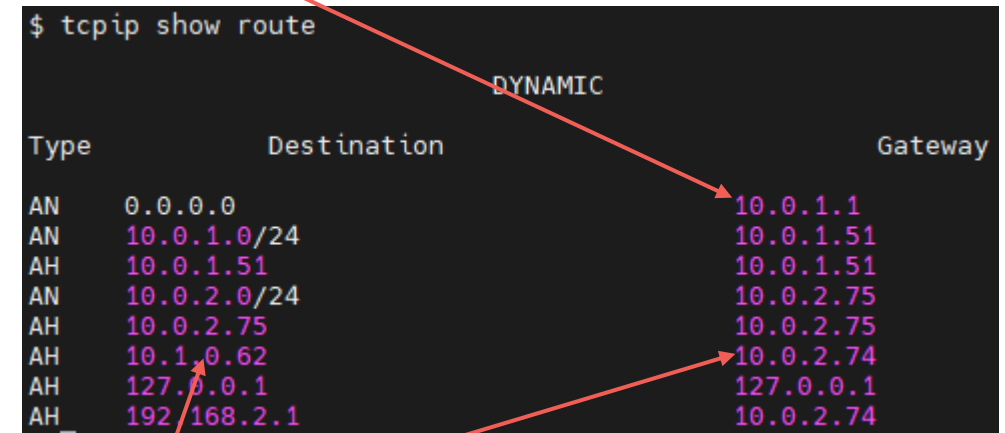
# Configuring TCPIP

- Configure 2 interfaces, macVTap + bridge
- Default route – IP address of bridge interface in the host
- Redefine default route + static routes
  - SYS\$MANAGER:SYSTARTUP\_VMS.COM
  - Have a better idea? Please let us know!

```
$ @SYS$STARTUP:TCPIP$STARTUP.COM
$ tcpip set route /default /gateway=10.0.1.1
$ tcpip set noroute /default /gateway=10.0.2.74
```

```
$ tcpip set route 192.168.122.102 /gateway=10.0.2.74 /permanent
```

Default route (used by macVTap interface)



The screenshot shows the output of the 'tcpip show route' command. A red arrow points from the text 'Default route (used by macVTap interface)' to the first row of the table, which shows the default route (0.0.0.0) with gateway 10.0.1.1. Another red arrow points from the text 'Static route to other nodes (used by bridge interface)' to the last row of the table, which shows a static route (192.168.2.1) with gateway 10.0.2.74.

Type	Destination	Gateway
AN	0.0.0.0	10.0.1.1
AN	10.0.1.0/24	10.0.1.51
AH	10.0.1.51	10.0.1.51
AN	10.0.2.0/24	10.0.2.75
AH	10.0.2.75	10.0.2.75
AH	10.1.0.62	10.0.2.74
AH	127.0.0.1	127.0.0.1
AH	192.168.2.1	10.0.2.74

Static route to other nodes (used by bridge interface)

# Cluster Configuration

- Cluster over IP (IPCI)
- Choose the bridged interface for cluster communication
- Unicast addresses

# Summary

- ▶ Cloud + On-Prem clusters, Possible!
- ▶ Networking on OCI is challenging, but not impossible
- ▶ On-prem Wireguard setup varies

# One last thing... OpenVMS native on OCI

▶ Buckets, custom images

▶ Netboot

▶ Custom iPXE

Shape: VM.Standard.E5.Flex  
Chipset: q35

# Get in touch with Professional Services!

- We are a mixed team, with lots of years of VMS experience and well versed in many contemporary IT tools
- We are here to help, get in touch!
- Save your time! Tools, scripting, automation
- Learn from our mistakes! Processes, best practices
- Leverage our attention to details: Performance Testing, Fine Tuning
- Delegate your responsibilities: Full system migration, updates and upgrades, plus 24/7 managed services
- And much more! Contact your account manager for details.



# Questions and Suggestions Please!



OCI



KVM



VMS