



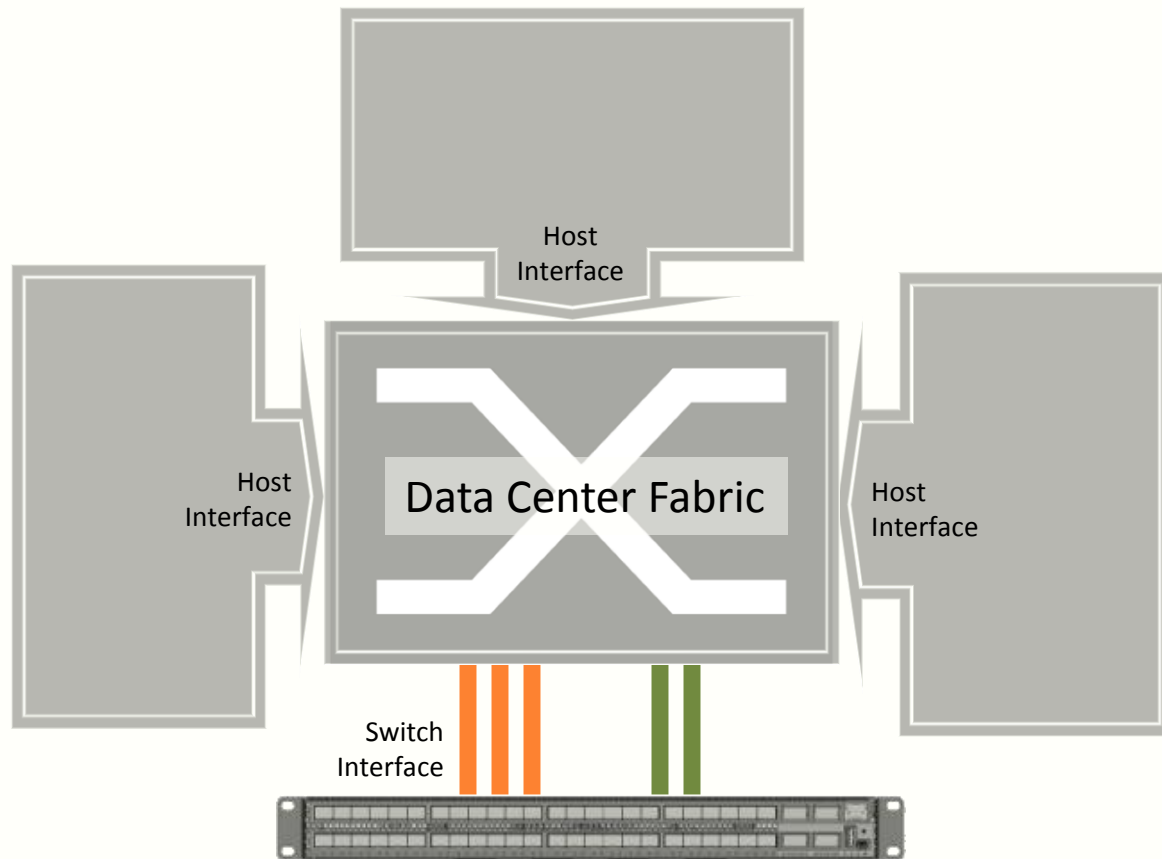
# VMready

*Implementation of a Profile Based Virtual Switching Framework*

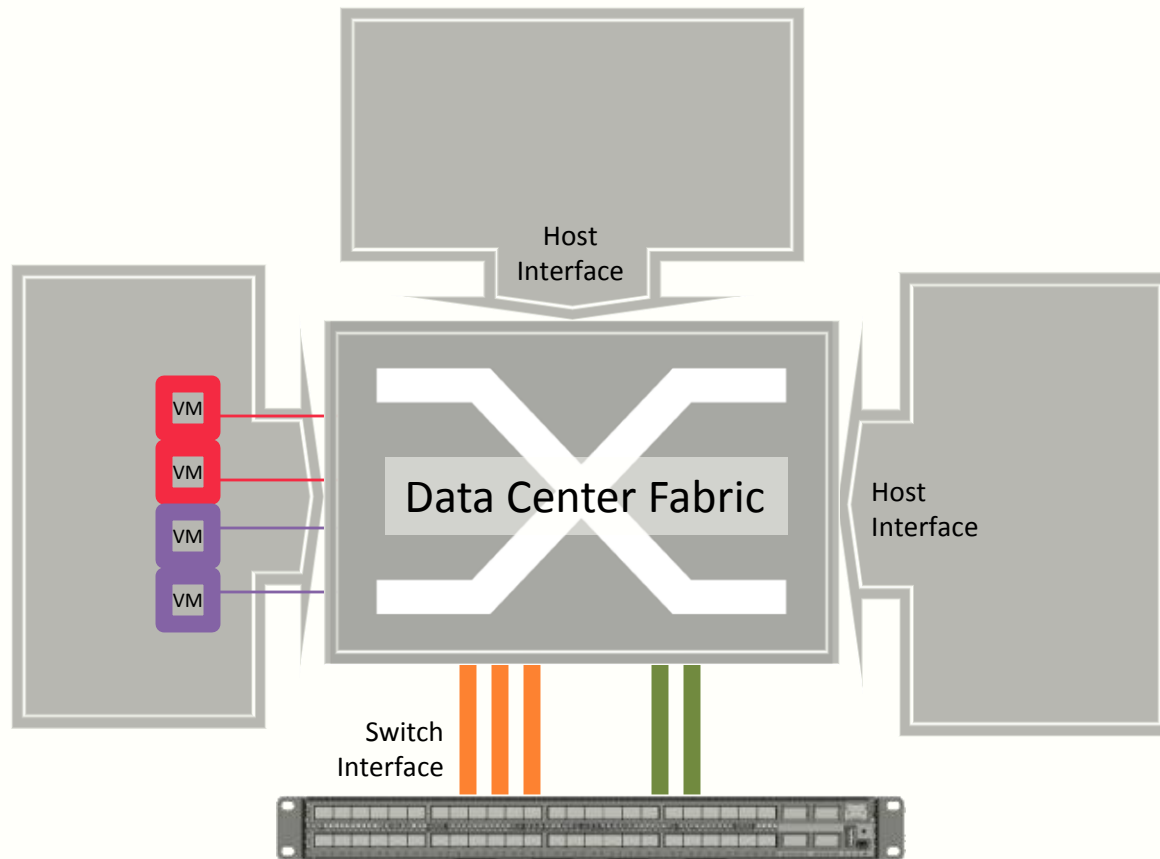
Vijoy Pandey and Jay Kidambi

DC-CAVES 2010

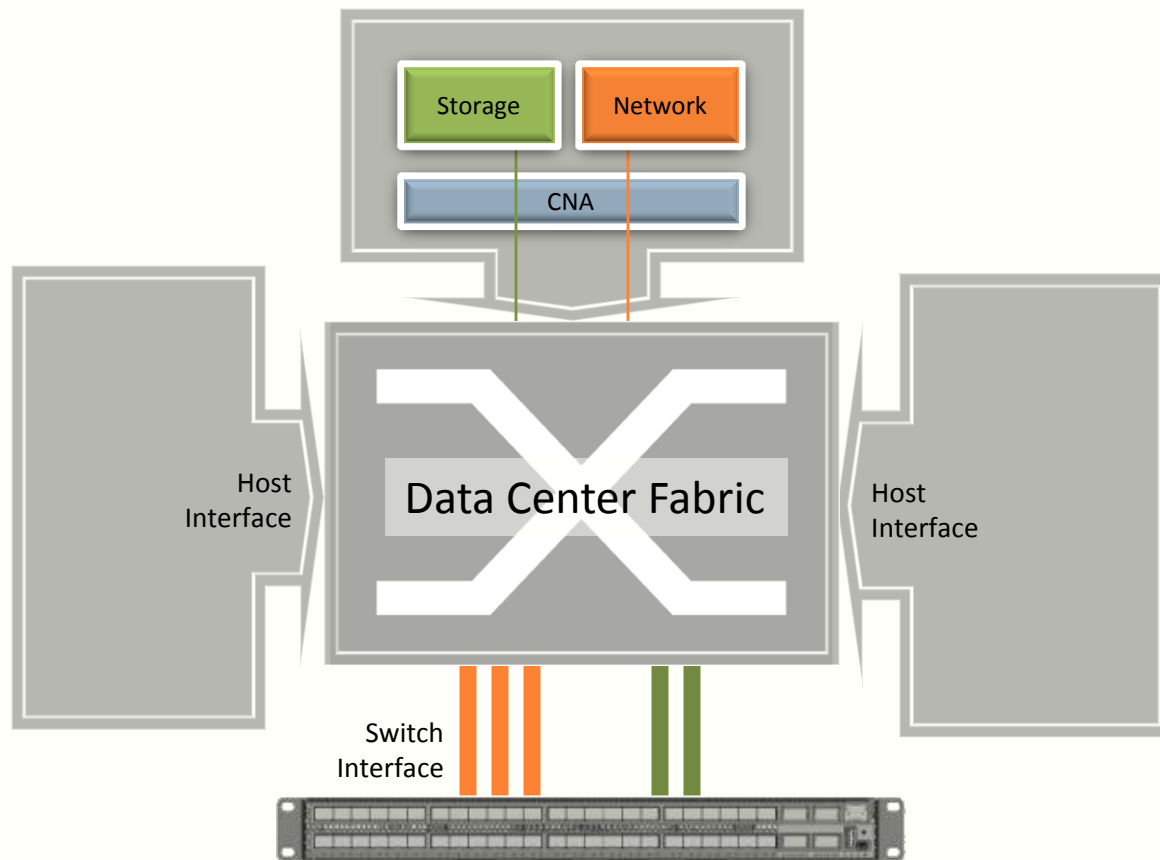
# The evolving switching primitive



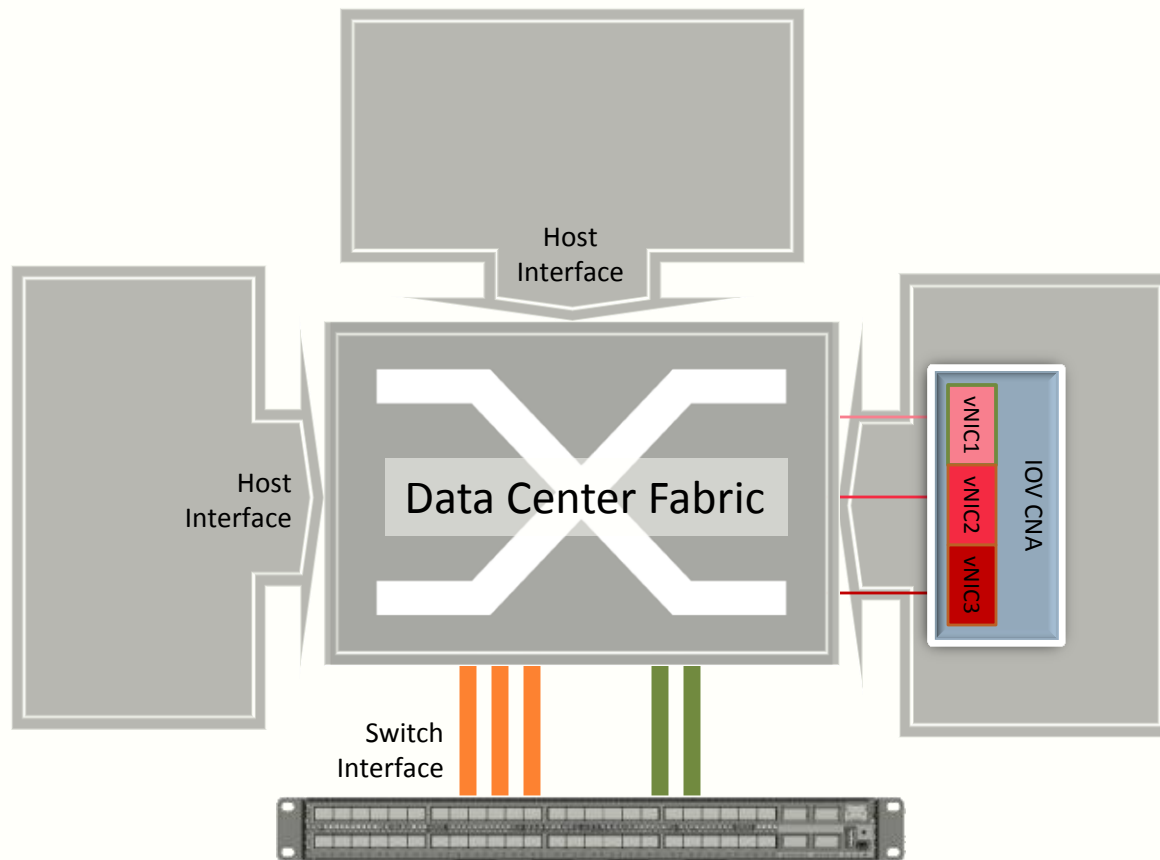
# The evolving switching primitive



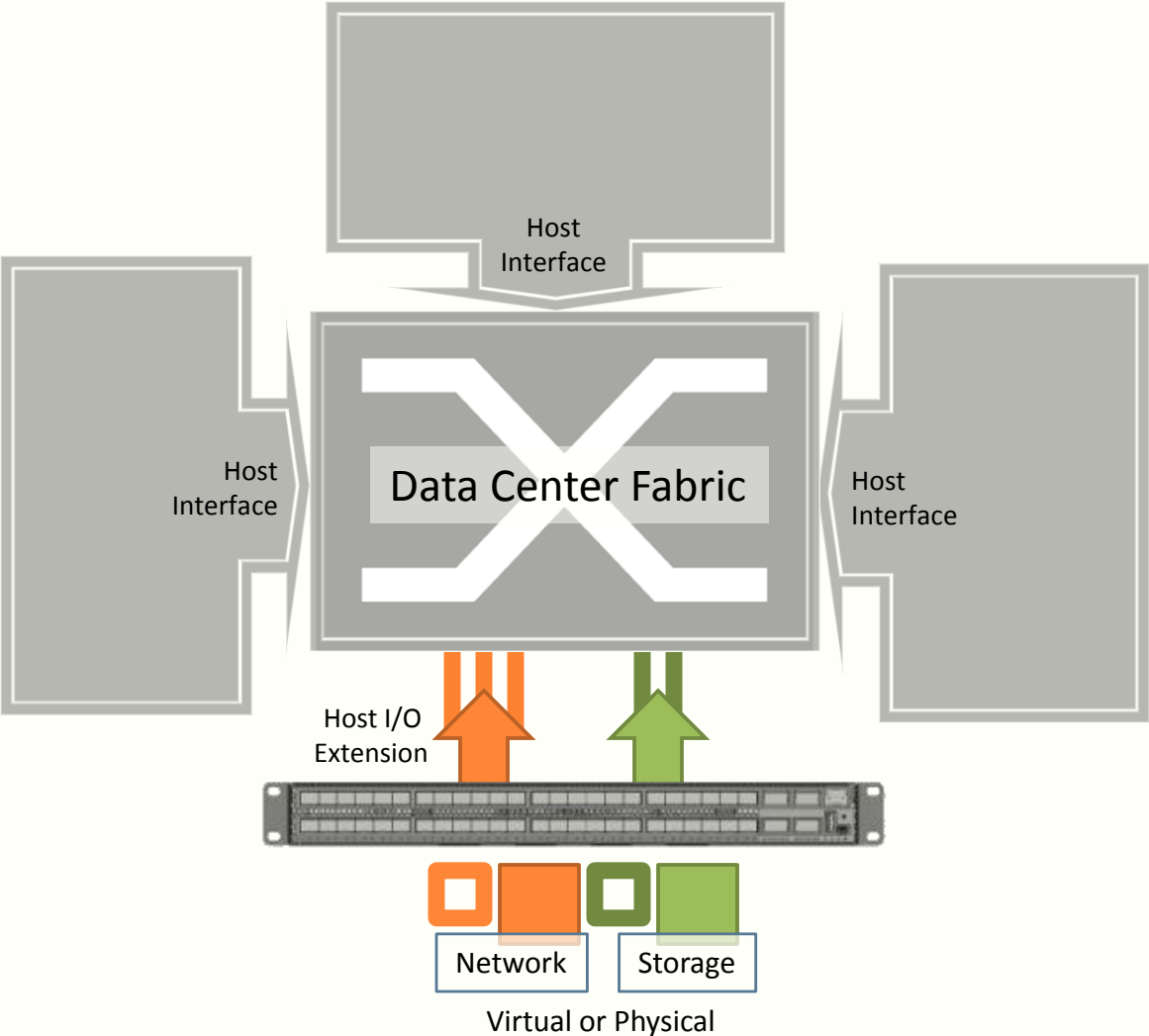
# The evolving switching primitive



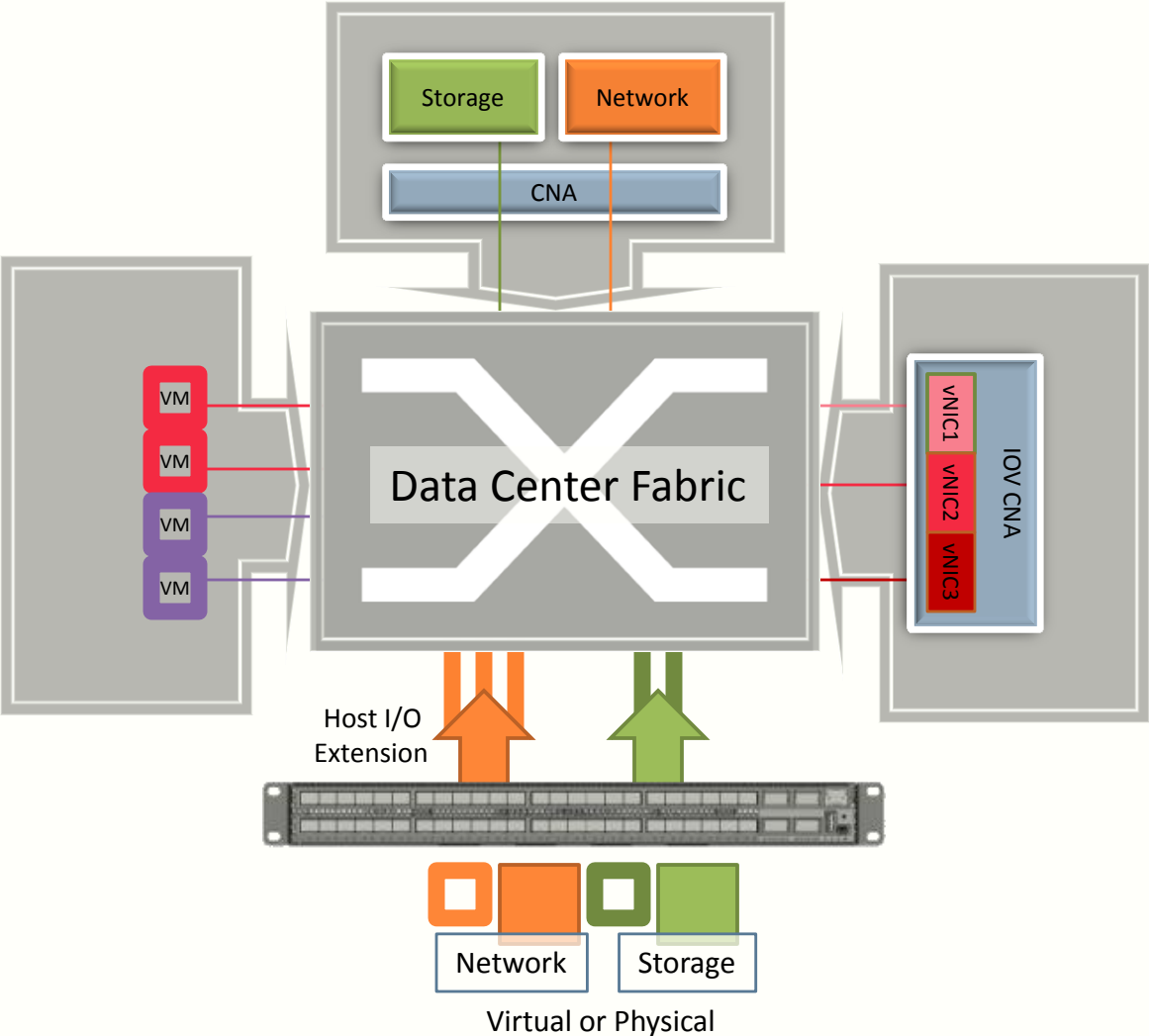
# The evolving switching primitive



# The evolving switching primitive

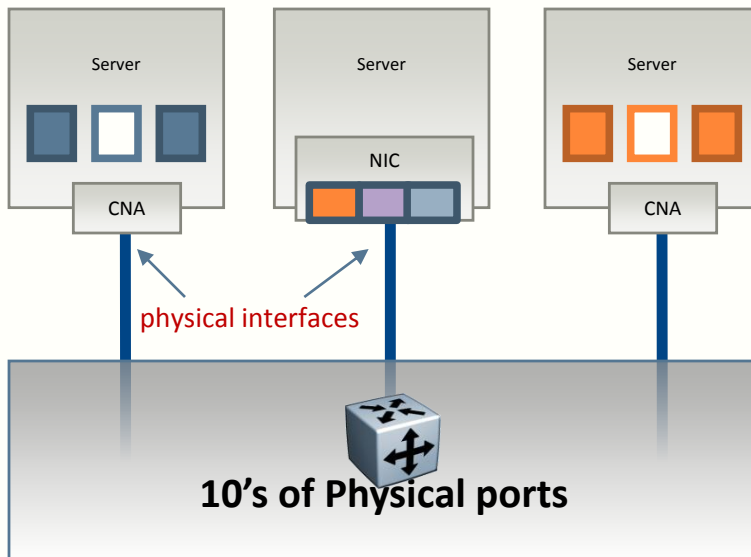


# The evolving switching primitive



# VMready : Virtualization Aware Networking

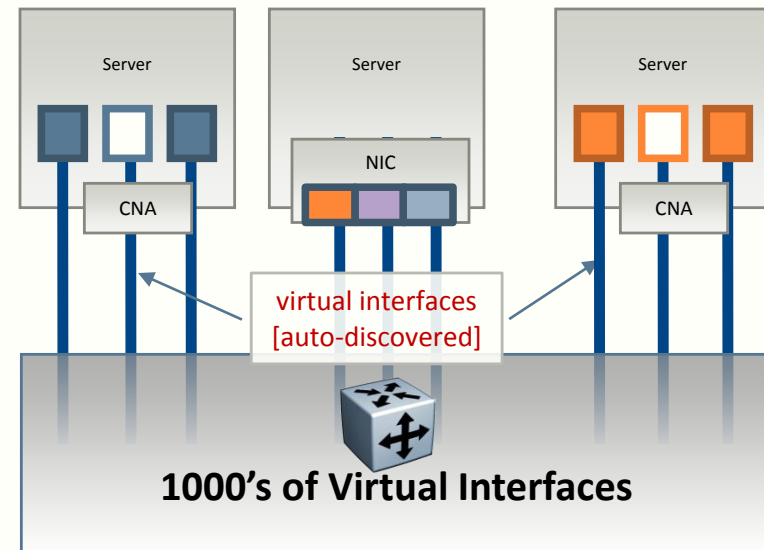
## Traditional Switches



### Physical interfaces are the old switching unit

- Not aware of server-side virtualization technologies
- Configuration per physical interface only
  - No flexibility for unique VM needs
- No flexibility to handle VM migrations

## VMready Switches

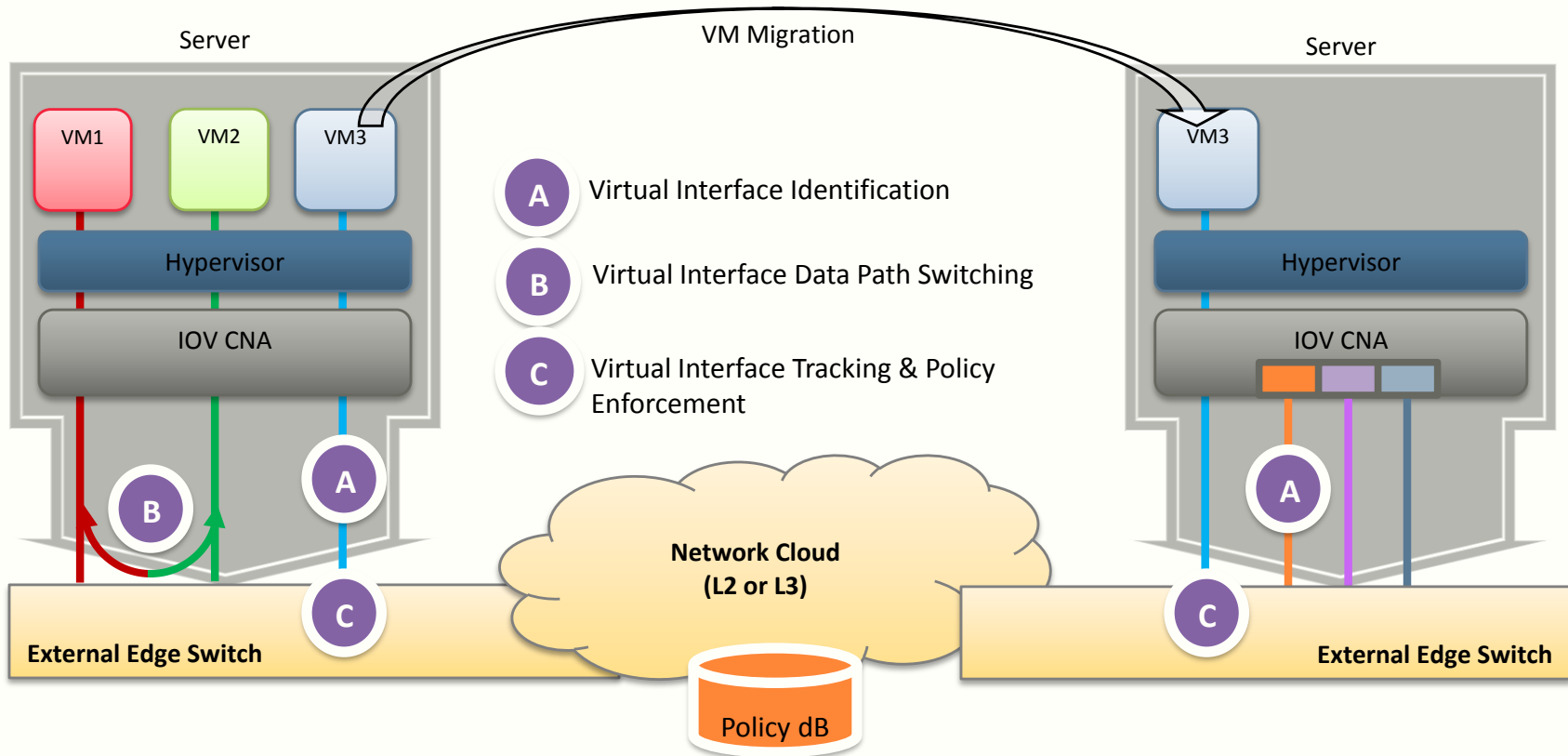


### Virtual interfaces are the new switching unit

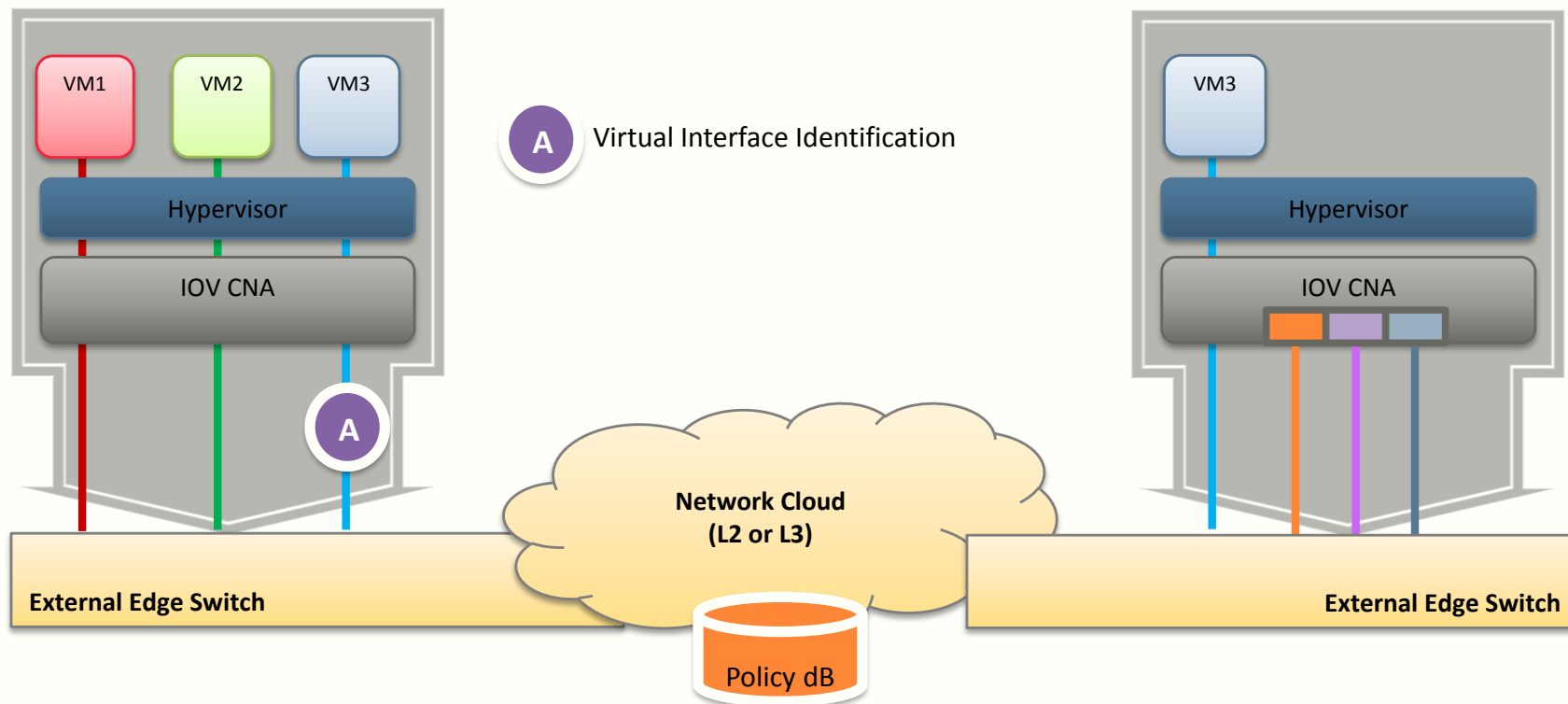
- Fully aware of server-side virtualization technologies
- Configuration per virtual interface
- Automated Network mobility
  - Configuration follows Virtual Machines in real-time during migration



# Big Picture : What is needed



# A: Identification



- VMready uses MAC based identification of VMs today
  - MAC assignment pools for the different vendors
- VMready uses Q-in-Q for NIC-partition (vNIC) identification

BNT 10-port 10Gb Ethernet Switch

- System
- Switch Ports
- Port-Based Port Mirroring
- Layer 2
- RMON Menu
- Layer 3
- QoS
- Access Control
- CEE
- FCOE
- Virtualization
  - VNIC
  - Virtual Machine
    - Global
    - VM Group General
    - VM Group Ports
    - Advanced Pre-Provisioning
    - VM Profile

Configure Via a Data Center ID:

Enter the UUID/MAC of a Virtual Entity:  VM ?

Configure via this ID

Configure Via a Host:

Host name / IP Address	UUID
<a href="#">cloudburst-esxi-newmgmt.cloudburst.net</a>	f7527cd2-b295-34c0-8e7f-9f27974ce48d
<a href="#">cloudburst-esxi2.cloudburst.net</a>	d6645b3c-4422-11de-8abf-00215e964158
<a href="#">cloudburst-esxi1.cloudburst.net</a>	dbbdcade-443e-11de-b425-00215e9655fc
<a href="#">192.168.70.80</a>	6ff9e28d-35c2-3143-b6d3-e9fa5b6aef6f
<a href="#">192.168.70.82</a>	df90d624-2bfa-3515-a45f-3f4aa7d60e9e
<a href="#">cloudburst-esxi3.cloudburst.net</a>	74693564-459f-11de-8d76-00215e9105c4
<a href="#">192.168.70.204</a>	00:50:56:7a:00:11:9b
<a href="#">192.168.70.81</a>	87:00:00:00:00:00
<a href="#">192.168.70.163</a>	5e:00:00:00:00:00

Configure Via a VM:

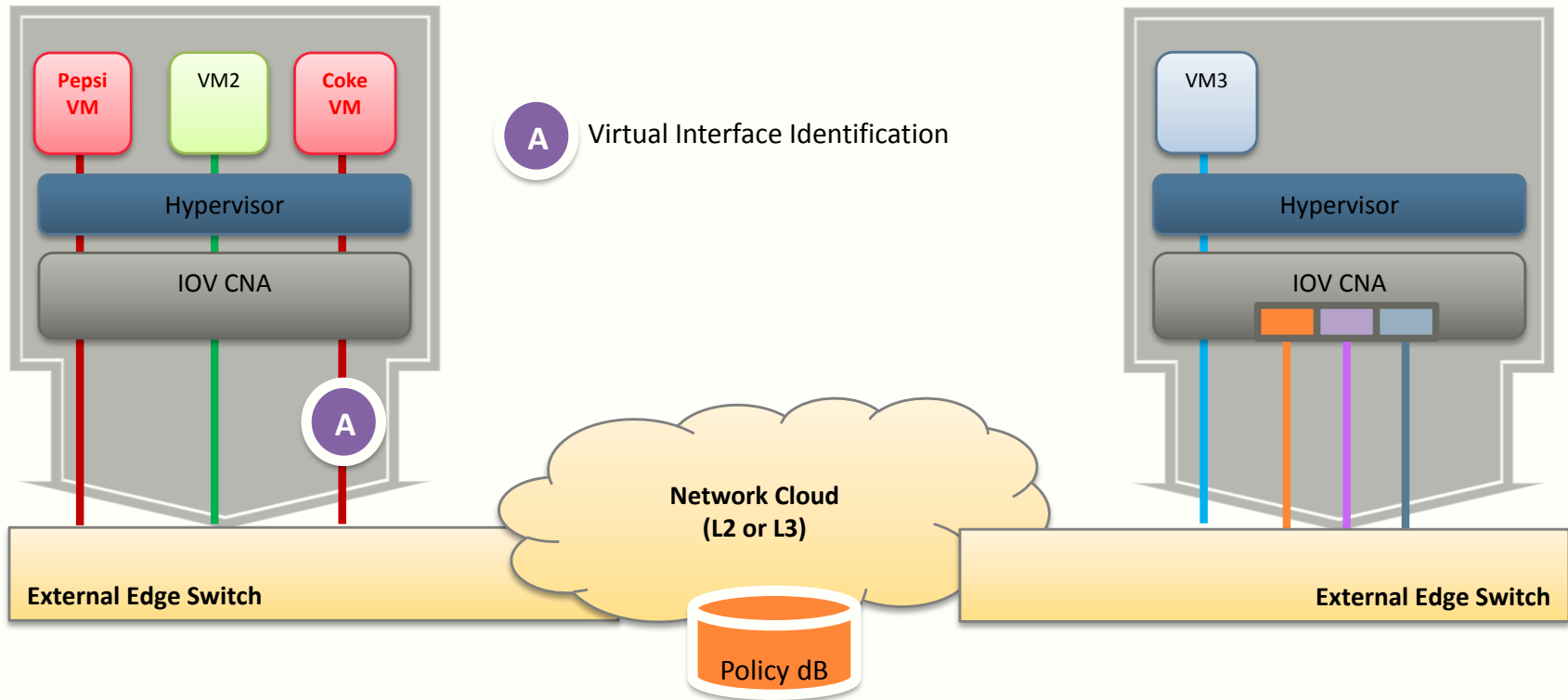
VM vCenter Name	UUID
<a href="#">vm4</a>	5002b8d5-0000-0000-0000-000000000000
<a href="#">RHEL5MGT</a>	5031a064-0000-0000-0000-000000000000

```

BCH1-10G-7A(config)#
BCH1-10G-7A(config)#show virt vm
IP Address          VMAC Address       Index Port      VM Group (Profile)
-----
10.160.5.1          00:50:56:00:11:9b  20  INT3
10.150.2.1          00:50:56:45:0c:e8  16  INT5
10.150.2.2          00:50:56:46:03:e9   4  INT3
10.150.2.4          00:50:56:4e:e0:ee   6  INT2
10.150.2.15         00:50:56:73:28:4e  21  INT12
10.150.3.1          00:50:56:74:a4:fe  17  INT5
10.150.2.17         00:50:56:78:c1:3c  14  INT10
10.150.3.4          00:50:56:79:45:75   5  INT2
10.150.3.2          00:50:56:7d:e3:e4  10  INT3
10.150.2.16         00:50:56:7d:e6:cc  13  INT11
203.1.1.12          00:50:56:82:0b:13   1  INT10  1
0.0.0.0             00:50:56:82:4a:dc   0  INT10  1
203.1.1.1           00:50:56:82:51:7b   3  INT10
203.1.1.13          00:50:56:82:5a:9f   2  INT10  2
203.1.1.3           00:50:56:82:5c:cb  18  INT12
203.1.1.2           00:50:56:82:5d:32  11  INT11
203.1.1.4           00:50:56:82:61:0f  19  INT12
10.160.0.102        00:50:56:bb:32:c9   8  INT2

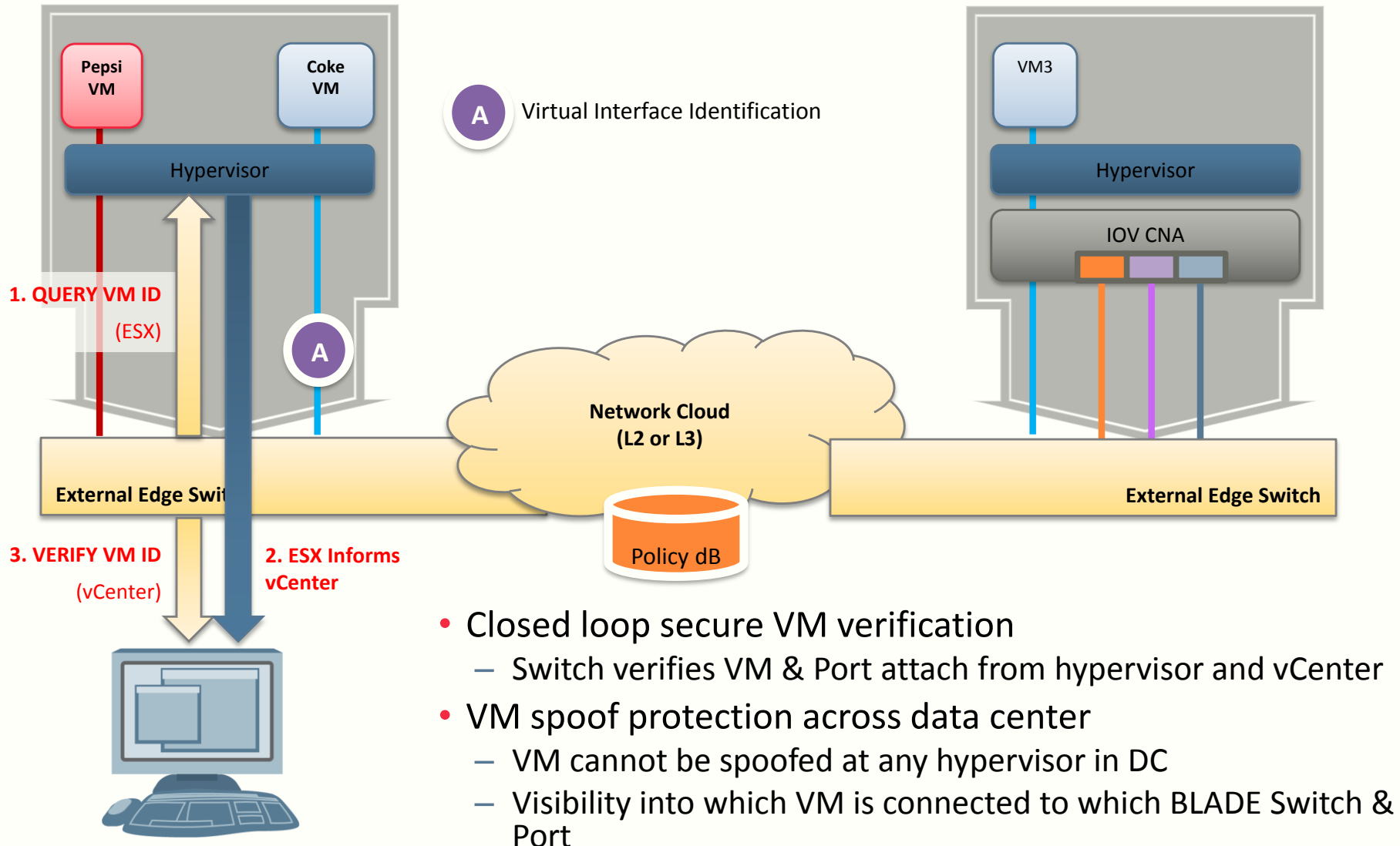
Number of entries: 18
* indicates VMware ESX Service Console Interface
# indicates VMware ESX/ESXi VMKernel or Management Interface
0.0.0.0 indicates IP address not yet available
BCH1-10G-7A(config)#
    
```

# A: Secure Identification



- MAC based identification of VMs is potentially **insecure**
  - Prone to spoofing attacks
- DCBX-based vNIC identification is **secure**
  - Agent running on both ends of the wire guarantees spoof-free operation

# A: Secure Identification for VMs



```

root@rslinux:~
File Edit View Terminal Help
*172.31.46.6    00:50:56:4d:80:56  2    INT2
 0.0.0.0      00:50:56:5b:4b:e9  3    INT2
+172.31.46.7    00:50:56:71:84:a7  1    INT2

Number of entries: 3
* indicates VMware ESX Service Console Interface
+ indicates VMware ESX/ESXi VMKernel or Management Interface
0.0.0.0 indicates IP address not yet available

>> Virtual Machine# /cfg/virt/vmg
Enter group number: (1-32) 1
-----
[VM group 1 Menu]
vlan      - Set the group's vlan (only for groups with no VM profile)
vmap      - Set VMAP for this group
tag       - Enable vlan tagging on all VM group ports
addvm     - Add a virtual entity to the group
remvm     - Remove a virtual entity from the group
validate  - Sets secure mode for all VMs in this group
addprof   - Add a VM profile to the group
remprof   - Delete any VM profile associated with the group
addport   - Add ports to the group
remport   - Remove ports from the group
addtrunk  - Add trunk to the group
remtrunk  - Remove trunk from the group
addkey    - Add LACP trunk to the group
remkey    - Remove LACP trunk from the group
stg       - Assign VM group vlan to a Spanning Tree Group
del       - Delete group
cur       - Display current group configuration

>> VM group 1# addvm 00:50:586:8b:48:a8

>> VM group 1# apply
.
-----
Apply complete; don't forget to "save" updated configuration.

>> VM group 1# 2000-08-09 15:06:03:Unauthenticated MAC [00:50:56:8b:48:a8] from port 1 switchId [172.31.]

Aug 9 6:58:41 172.31.213.1 NOTICE  vm: VM agent operation failed.
Disabling port 1 because verification failed on the port

Aug 9 6:58:41 172.31.213.1 NOTICE  server: link down on port INT1
1040:2000-08-09 15:07:44: Secure Validation OK for MAC [00:50:56:8b:48:a8] from port 2 switchId [172.31.]

Aug 9 7:00:22 172.31.213.1 NOTICE  vm: Virtual Machine with IP address 172.31.213.2 came online

Aug 9 7:02:37 172.31.213.1 NOTICE  mgmt: Port MGT1 ENABLED and MGT2 DISABLED because Management Modulee

```

# vCenter Visibility of BLADE Switch & Ports

172.31.46.6 VMware ESX, 4.0.0, 164009

- Getting Started
- Summary
- Virtual Machines
- Resource Allocation
- Performance
- Configuration
- Tasks & Events
- Alarms
- Permissions
- Maps
- Storage Views
- Hardware Status

## Hardware

- Processors
- Memory
- Storage
- ▶ Networking
  - Storage Adapters
  - Network Adapters
  - Advanced Settings

## Software

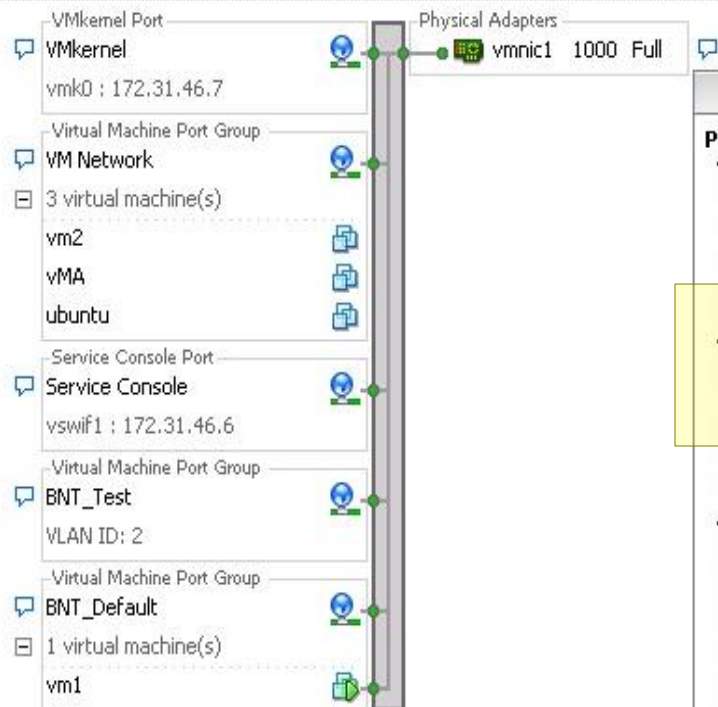
- Licensed Features
- Time Configuration
- DNS and Routing
- Power Management
- Virtual Machine Startup/Shutdown
- Virtual Machine Swapfile Location
- Security Profile
- System Resource Allocation
- Advanced Settings

View: Virtual Switch Distributed Virtual Switch

## Networking

Virtual Switch: vSwitch0

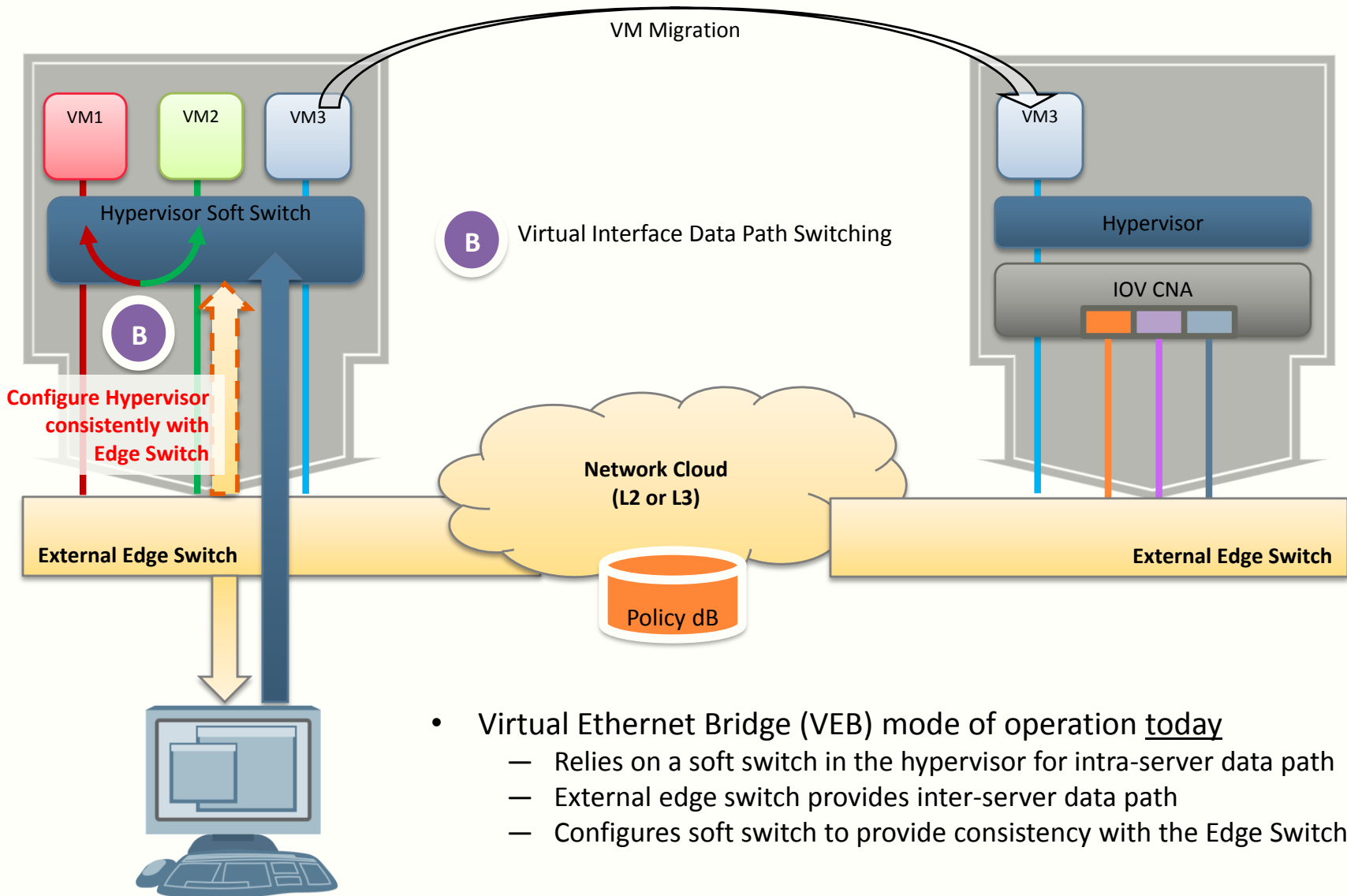
Remove... Properties...



### Properties

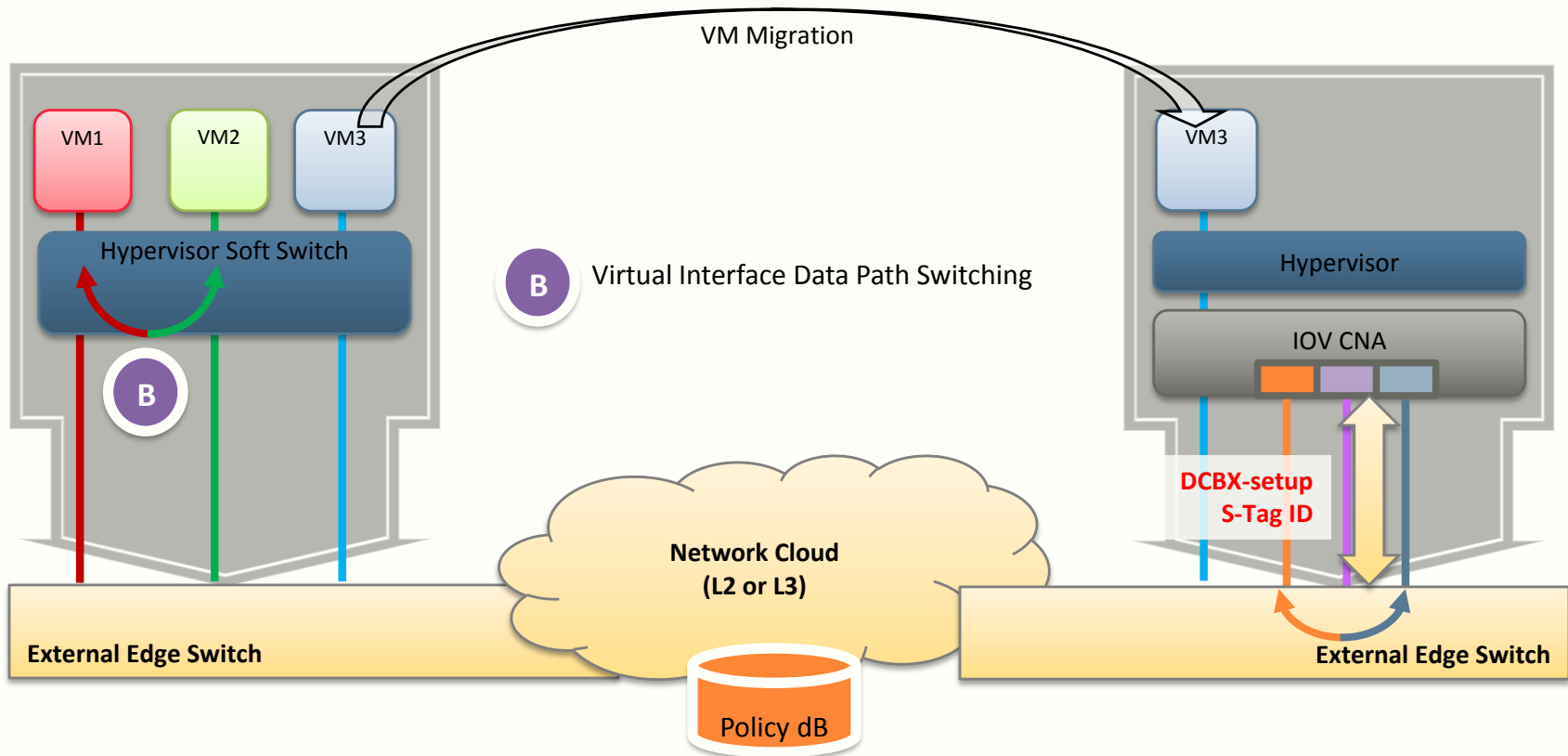
Version	1
Timeout	0
Time to live	83
Samples	723929
Device ID	BLADE
Address	172.31.213.1
Port ID	INT2
Software Version	Unreleased: FW_VIEW:
Hardware Platform	BNT 1/10Gb Uplink
IP Prefix	0.0.0.0
IP Prefix Length	0
VLAN	0
Full Duplex	false
MTU	0
System Name	
System Old	
Management Address	0.0.0.0
Location	

# B: Data Path Switching (VMs)



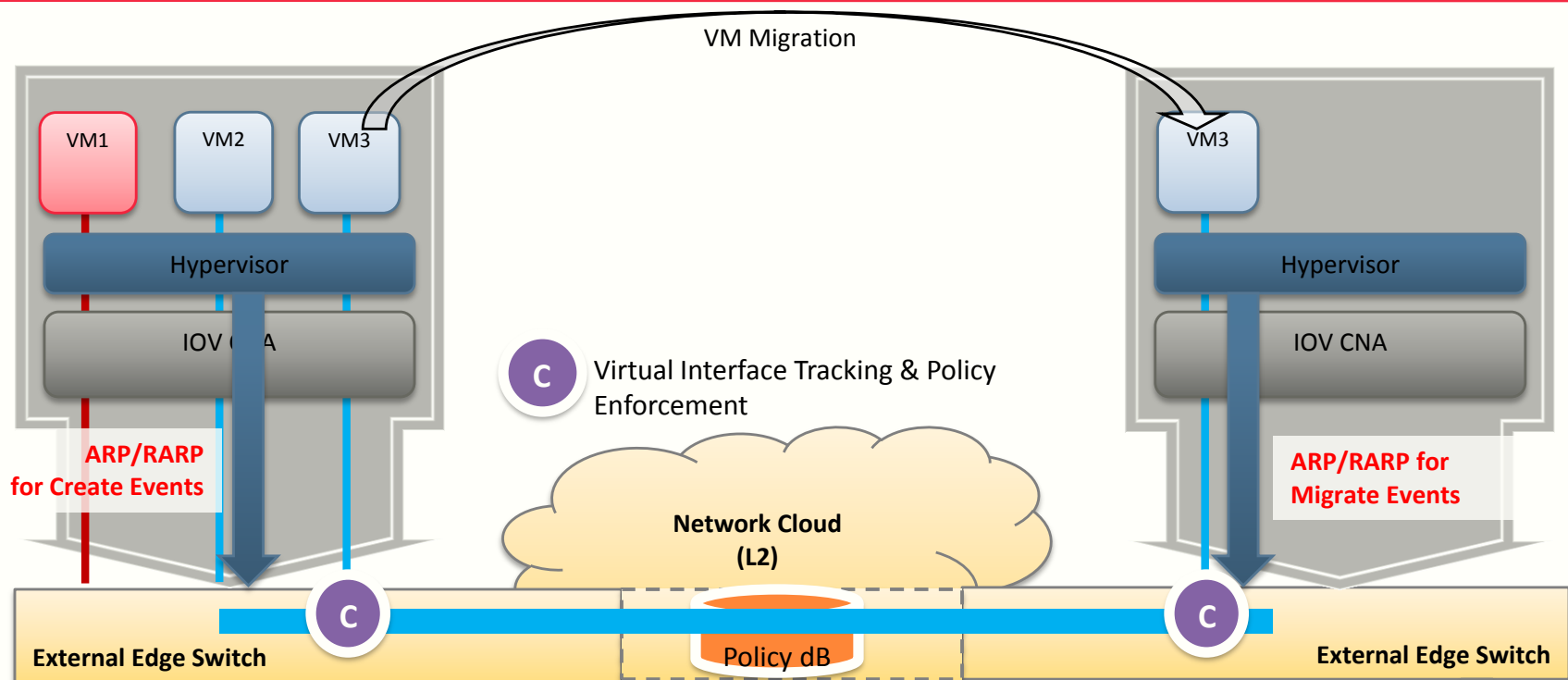


# B: Data Path Switching (vNICs)



- Isolation is key property
  - DCBX to exchange vNIC parameters, S-Tags for ID
  - Two models : switch mode and I/O extender mode of operation
  - S-Tags either
    - carried to Network Cloud (Host I/O extender mode) or
    - stripped on ingress (Switch mode)

# C : Stack Switch Tracking & Policy Enforcement



- Policy-based configuration on Edge Switch
  - Profiles created on switch or distributed switch
  - Virtual Interfaces (live or pending) assigned to Port Group
  - Profile attached to Port Group
- Migrate events cause Migrated VM ID to be attached to the same (BLUE) Port Group
  - Post-verification with VM Manager
- **Drawback**
  - Post-creation / post-move association of profile
  - Window small in the normal case since RARP sent before VM traffic

40. Oct 20 21:14:24 192.168.70.133 NOTICE system: 3m DAC inserted at port EXT5

**BNT 10-port 10Gb Ethernet Switch**

- System
- Switch Ports
- Port-Based Port Mirroring
- Layer 2
- RMON Menu
- Layer 3
- QoS
- Access Control
- CEE
- FCOE
- Virtualization
  - VNIC
  - Virtual Machine
    - Global
    - VM Group General
    - VM Group Ports
    - Advanced Pre-Provisioning
    - VM Profile

203.1.1.13 (InterVM13)
Group 2

+ ESX/ESXi VMkernel/Management

**INT11**

VM IP/MAC Address (Name)	VM Group
+10.150.2.16 (192.168.70.81)	None
203.1.1.2 (vm2)	None
+ ESX/ESXi VMkernel/Management	

**INT12**

VM IP/MAC Address (Name)	VM Group
+10.150.2.15 (192.168.70.82)	None
203.1.1.3 (vm3)	None
203.1.1.4 (vm4)	None
+ ESX/ESXi VMkernel/Management	

**INT13**

**INT14**

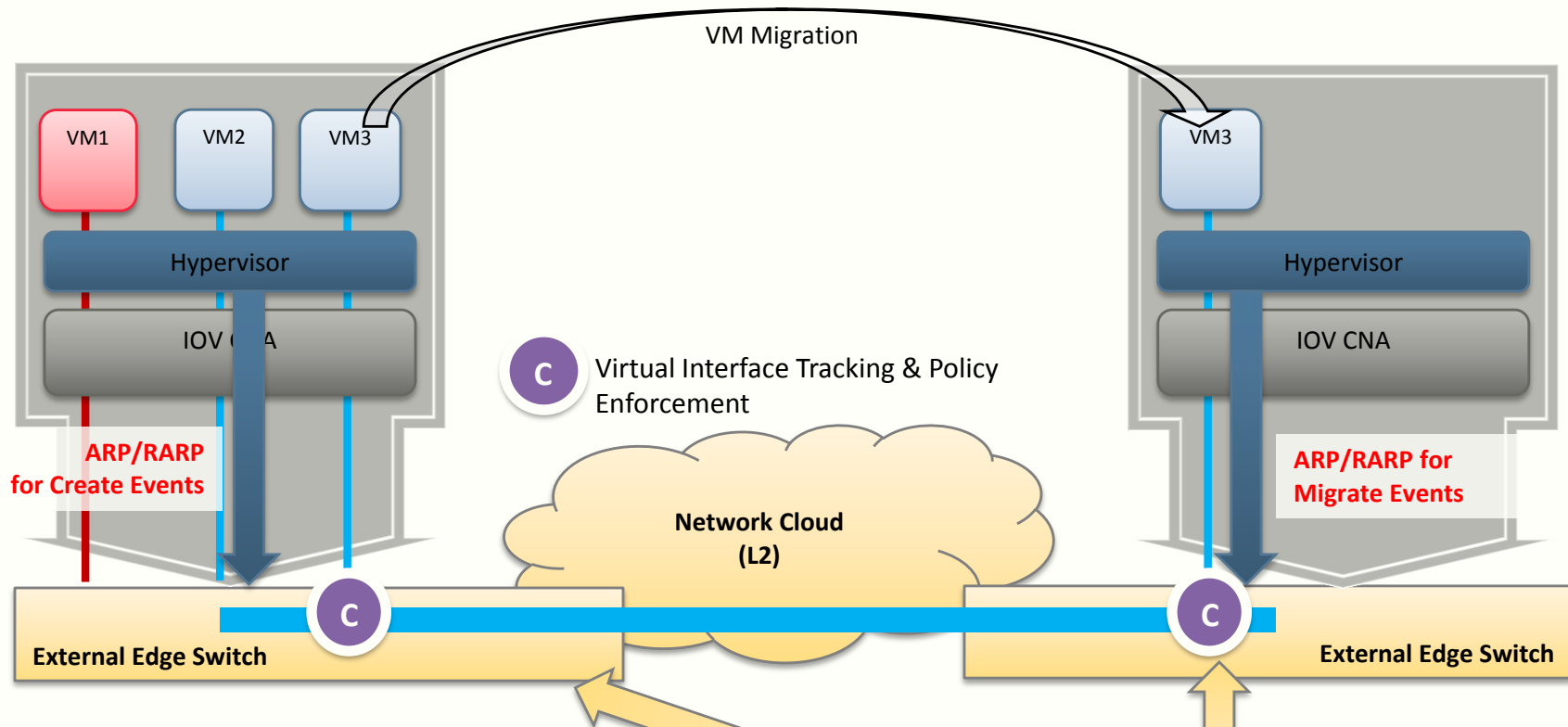
**VM Pre-Provisioning**

Done

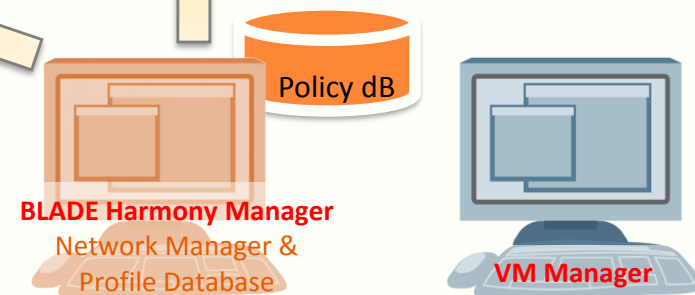
```

BCH1-10G-7A(config)#virt vmgroup 3 vm vm1
BCH1-10G-7A(config)#show vlan
VLAN          Name               Status MGT           Ports
-----
1             Default VLAN      ena    dis    INT1-INT14 EXT1-EXT11
100          VLAN 100          ena    dis    INT1-INT14 EXT5 EXT6
101          VM Group 2 (T)    ena    dis    INT10 EXT5 EXT6
220          VM Group 1 (T)    ena    dis    INT10 EXT5 EXT6
4093         VM group 3        ena    dis    INT10
4094         VM group 4        ena    dis    empty
4095         Mgmt VLAN         ena    ena    INT1-INT14 MGT1 MGT2
BCH1-10G-7A(config)#virt vmgroup 4 vm vm2
BCH1-10G-7A(config)#
    
```

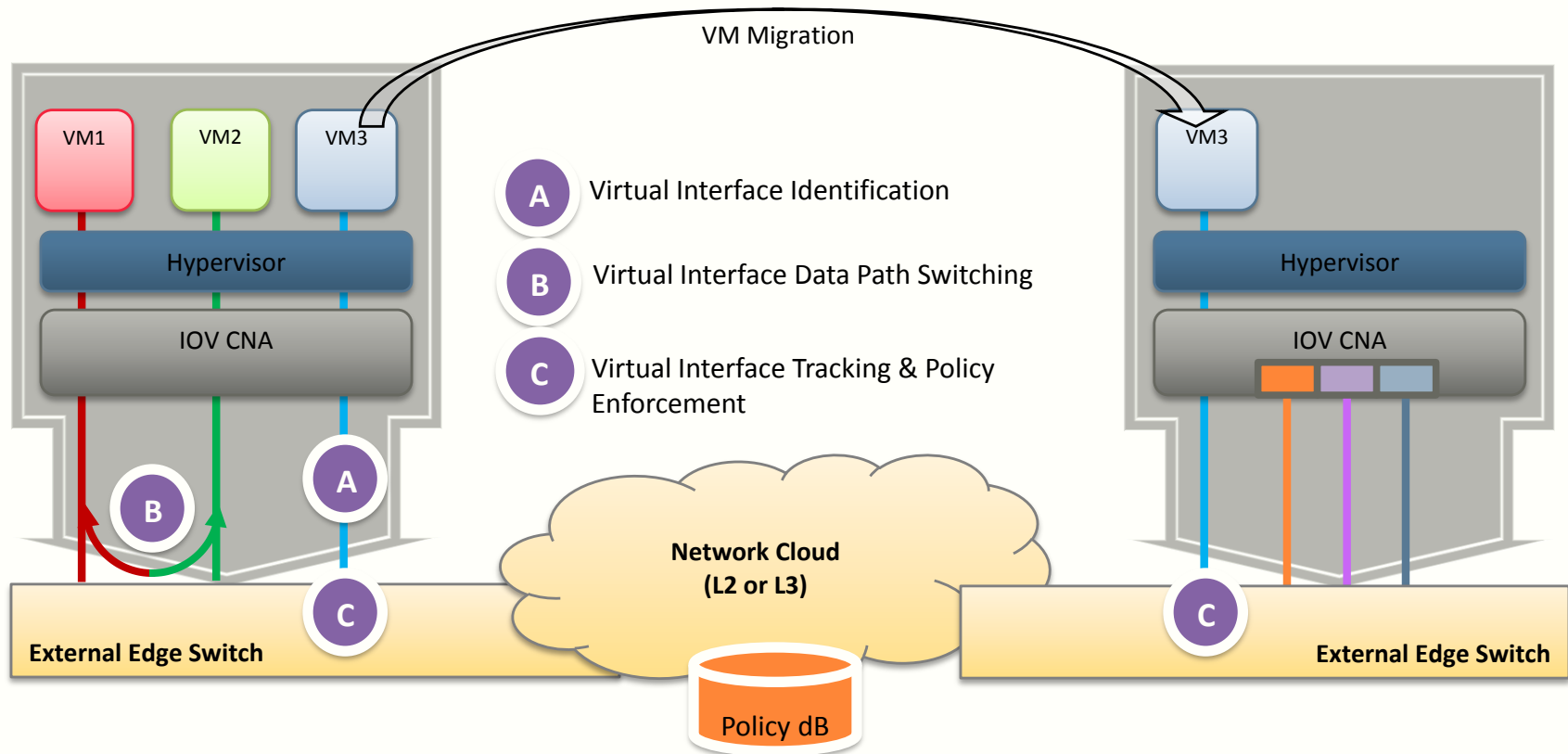
# C : DC-wide Tracking & Policy Enforcement



- Policy-based configuration on Policy DB Manager
  - Profiles created on BHM
  - On creation, switch queries BHM for Profile
  - On Migrate, destination switch queries BHM for Profile
  - BHM performs post-verification with VM Manager
- Drawback
  - Post-move / creation association of profile
  - Window small in the normal case since RARP sent before VM traffic

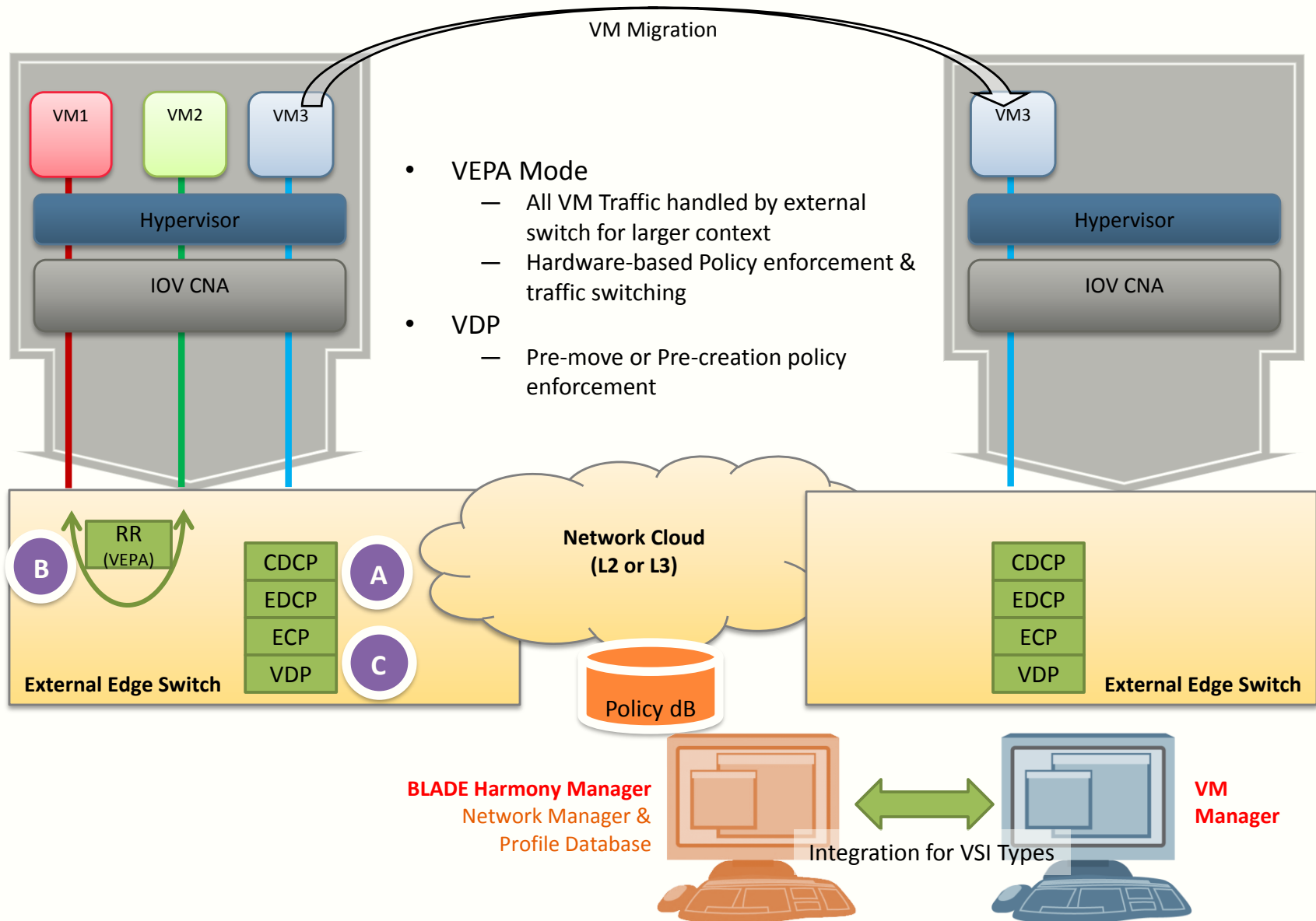


# Futures : ABC of Qbg



- A** (MAC, VLAN) or S-Tag [CDCP] (Channel Discovery & Control Protocol)
- B** VEB (Virtual Ethernet Bridge) or VEPA (Virtual Ethernet Port Aggregator)
- C** VDP (VSI Discovery & Control Protocol)

# VMready Changes for Qbg & Value Proposition



# Firmware Architecture

