



System z

z/VM Multi-VSwitch Link Aggregation (LAG)

Configuration and Operational Demonstration 101

Rick Tarcza

z/VM Networking Architecture and Design

© 2014 IBM Corporation

Today's Objectives

Discuss the Basics of a Multi-VSwitch LAG Configuration

Creating a Multi-VSwitch LAG Across Multiple z/VM Systems

Demonstrate the Capabilities of a Multi-VSwitch LAG Configuration

Multi-VSwitch LAG Development Core Team

- Mary Ellen Carollo
- Susan Farrell
- Tom Kovach
- Angelo Macchiano
- Bruce Mc Clellan
- Bruce Ratcliff
- Rick Tarcza
- Peggy Williams

Z/VM's Exclusive LAG Support

Virtual Switch

Physical Switch



No z/Hardware MIF Exploitation

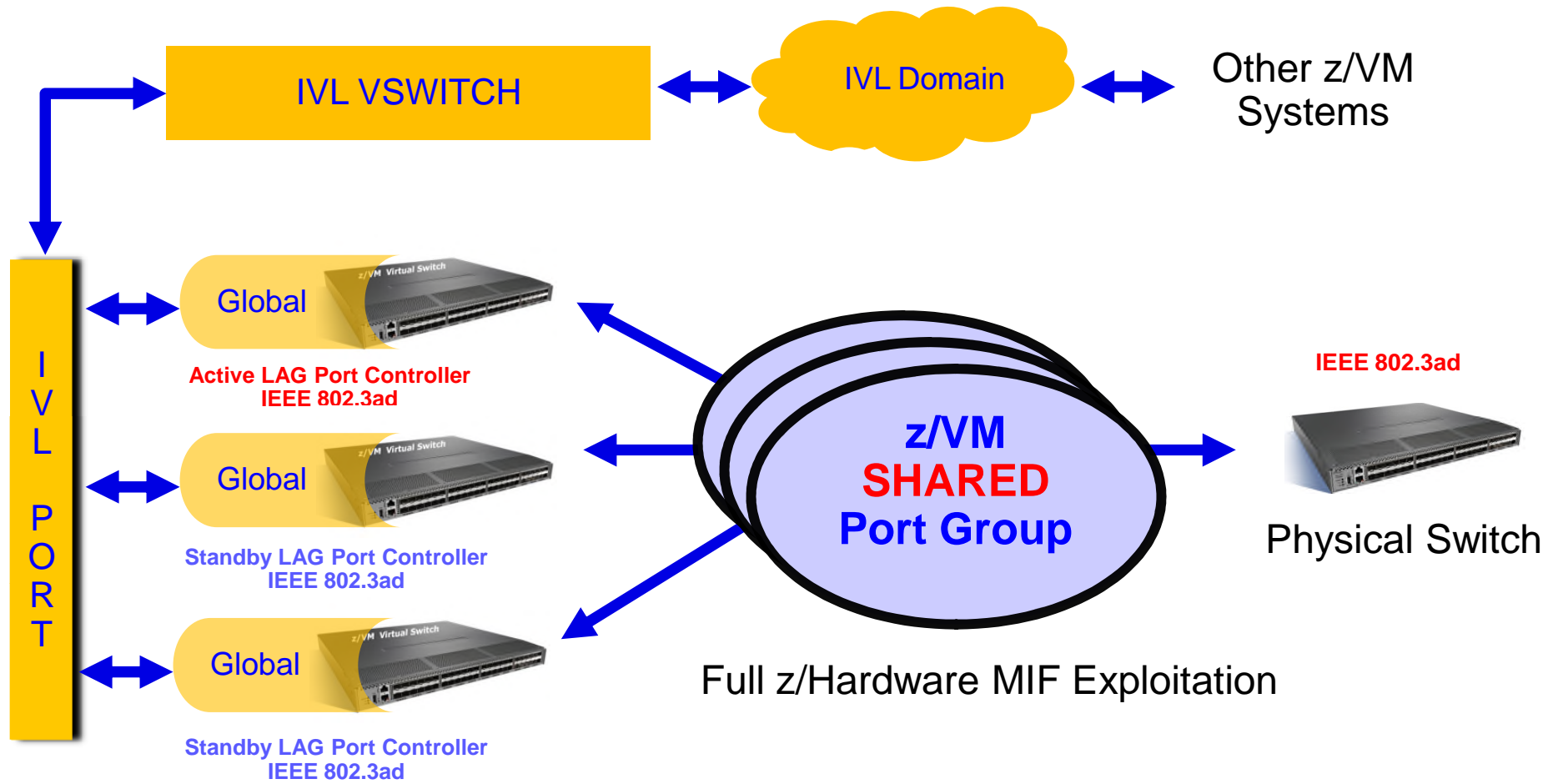
No Port Sharing of OSA Express Features
used by the Port Group

MIF: Multiple Image Facility

New Multi-VSwitch LAG Objects

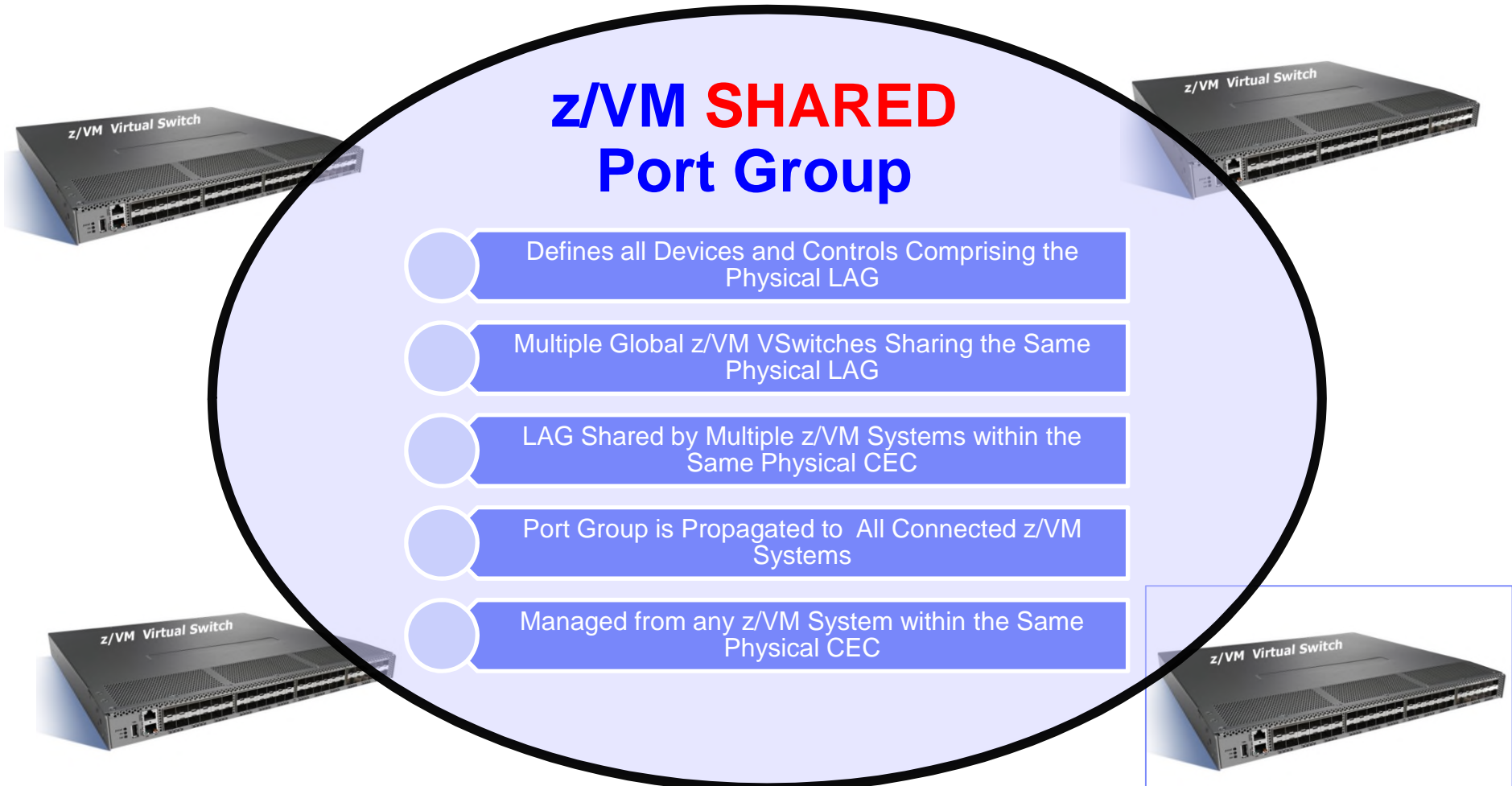
- **Shared Port Group** – Provides a single point of control for OSA Port management across multiple VSwitches sharing the same physical LAG Configuration.
- **Global VSwitch** – Provides the construct for a Virtual Switch to span multiple z/VM Systems by communicating with each other.
- **Inter-VSwitch Link (IVL)** - Provides management and data plane communications between Global VSwitches within the same or other z/VM Systems.

The Brave New World




Multiple Virtual Switches Sharing the Same Port Group

Multi-VSwitch LAG Configuration



Multi-VSwitch LAG Golden Rule

Each virtual switch sharing the same physical LAG must have identical connectivity.



A port added to the LAG must be added to all sharing VSwitches

A port removed from the LAG must be removed from all sharing VSwitches.

z/VM will insure Shared Port Group equilibrium.

Multi-VSwitch LAG Configuration Checklist

1

- **Defining the IVL Domain**

2

- System Connectivity to the IVL Network

3

- Verify System IVL Connectivity

4

- Configure a LAG on the Physical Switch

5

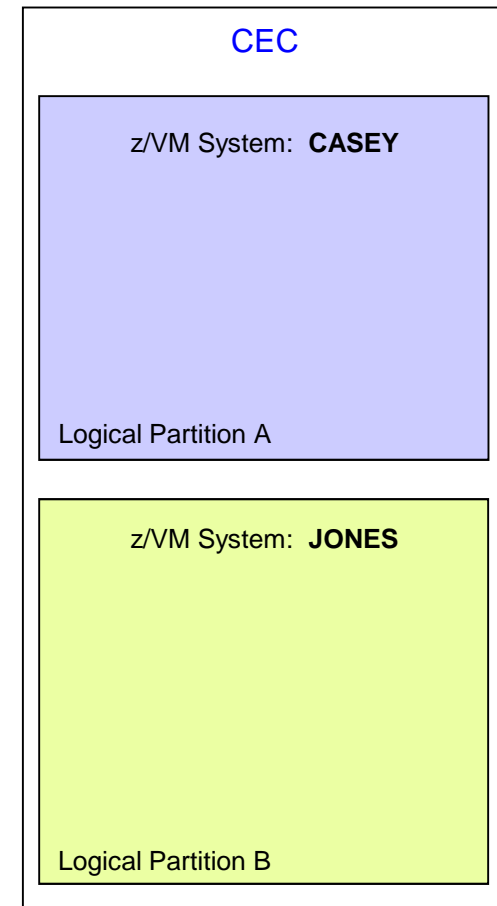
- Create a Shared Port Group

6

- Define a Global VSwitch

Step 1: Defining the IVL Domain

- The IVL is an Inter-VSwitch Link which provides the communication infrastructure to implement an IVL Domain.
- An IVL Domain is a group of up to 16 z/VM Systems connected by an Ethernet IVL LAN segment.
- All z/VM Systems sharing the same physical LAG must be members of the same IVL Domain.
- Up to 8 IVL Domains can be defined per VLAN
 - ▶ A z/VM System can be a member of only a single IVL Domain
 - ▶ VLANs are not required. The maximum number of domains that can be configured without VLANs is 8.
- The z/VM System is connected to the IVL network through a system administrator defined IVL VSwitch.
- The bandwidth required by the IVL consists of management and LAG data recovery communications.



Multi-VSwitch LAG Configuration Checklist

1

- Defining the IVL Domain

2

- **System Connectivity to the IVL Network**

3

- Verify System IVL Connectivity

4

- Configure a LAG on the Physical Switch

5

- Create a Shared Port Group

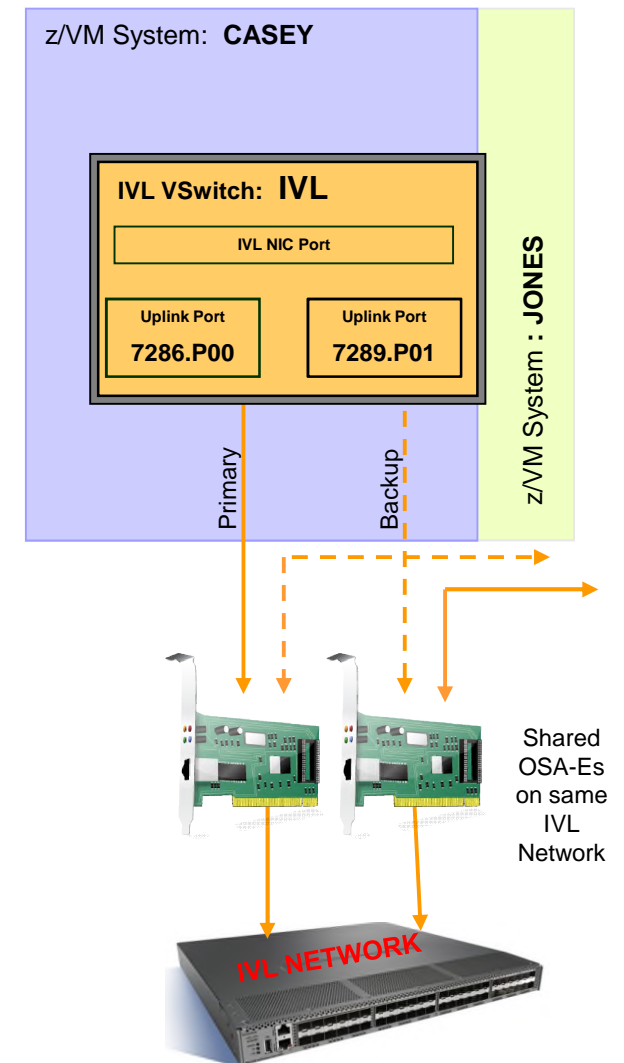
6

- Define a Global VSwitch

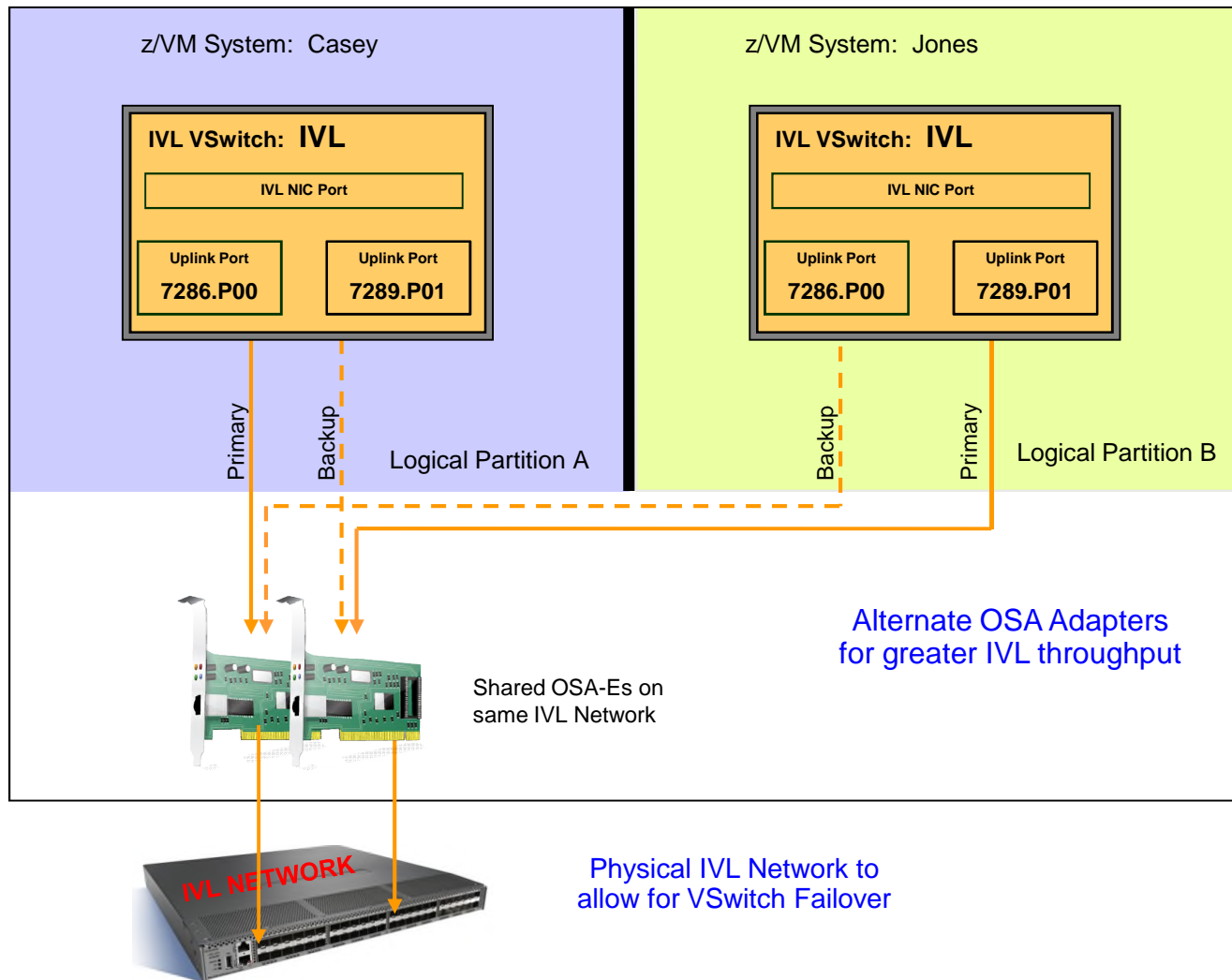
Step 2: Connectivity to the IVL Network

- **DEFINE VSWITCH IVL TYPE IVL DOMAIN B VLAN 8 NATIVE 1 UPLINK RDEV 7286.P00 7289.P01**
 - ▶ Only one IVL VSwitch allowed per z/VM System
 - ▶ By default Domain A is selected
 - ▶ Ethernet Layer 2 required for the IVL Network
 - ▶ Guest NIC Port not supported
 - ▶ VLAN support for IVL Network traffic isolation
 - ▶ Conventional UpLink Port with backups or an Exclusive LAG
 - ▶ Multi-VSwitch LAG UpLink Port is not supported
 - ▶ Recommend, but not required two OSA-Express (4 or 5) features to eliminate a single point of failure

- Repeat the same setup for z/VM System JONES, but alternate the RDEVs
 - ▶ **DEFINE VSWITCH IVL TYPE IVL DOMAIN B VLAN 8 NATIVE 1 UPLINK RDEV 7289.P01 7286.P00**
 - ▶ **SET VSWITCH IVL SWITCHOVER** may be used to alternate OSA UpLink Ports



IVL Network Configuration Domain B VLAN 8



Multi-VSwitch LAG Configuration Checklist

1

- Defining the IVL Domain

2

- System Connectivity to the IVL Network

3

- **Verify System IVL Connectivity**

4

- Configure a LAG on the Physical Switch

5

- Create a Shared Port Group

6

- Define a Global VSwitch

Step 3: Verify IVL Connectivity


Q VSWITCH IVL

```

VSWITCH SYSTEM IVL          Type: IVL      Connected: 0    Maxconn:
INFINITE
  PERSISTENT RESTRICTED     ETHERNET          Accounting: OFF
  PORTBASED LOCAL
  VLAN Aware Default VLAN: 0008   Default Porttype: Access  GVRP:
Enabled
      Native VLAN: 0001   VLAN Counters: OFF
  MAC address: 02-11-11-00-00-02   MAC Protection: Unspecified
  IPTimeout: 5      QueueStorage: 8
  Isolation Status: OFF      VEPA Status: OFF
Uplink Port:
State: Ready
PMTUD setting: EXTERNAL   PMTUD value: 8992   Trace Pages: 8
RDEV: 7286.P00 VDEV: 7286 Controller: DTCVSW2 ACTIVE
  Adapter ID: 28270007E386.0370
RDEV: 7289.P01 VDEV: 7289 Controller: DTCVSW1 BACKUP
  Adapter ID: 28270007E386.0370

```

New IVL Port Section
only on an
IVL VSwitch




IVL Port:

```

Adapter Owner: SYSTEM   NIC: FFFD.P00 Name: UNASSIGNED Type:
QDIO
  Porttype: Access
  RX Packets: 1040      Discarded: 27      Errors: 0
  TX Packets: 538      Discarded: 2       Errors: 0
  RX Bytes: 118222      TX Bytes: 61900
  Device: FFFD Unit: 000 Role: DATA   Port: 2100
  VLAN: 0008
  Options: Ethernet Broadcast
  Unicast MAC Addresses:
    02-11-11-00-00-00
  Multicast MAC Addresses:
    03-FF-FF-FF-FF-02
Ready; T=0.01/0.01 10:02:07

```

Unicast MAC Address
assigned to this system



Multicast MAC Address for
IVL Domain B



IVL Connectivity Verification

Q VMLAN

```

VMLAN maintenance level:
  Latest Service: VM65583
VMLAN MAC address assignment:
  System MAC Protection: OFF
  MACADDR Prefix: 021111 USER Prefix: 021111
  MACIDRANGE SYSTEM: 000001-FFFFFF
                   USER: 000000-000000
VMLAN Unified Resource Manager status:
  Hypervisor Access: NO      Status: UNAVAILABLE
  ID: NONE
  MAC Prefix: 000000
VMLAN default accounting status:
  SYSTEM Accounting: OFF    USER Accounting: OFF
VMLAN general activity:
  PERSISTENT Limit: INFINITE Current: 2
  TRANSIENT Limit: INFINITE Current: 0
Trace Pages: 8

```

```

IVL Domain: B  MAC address: 03-FF-FF-FF-FF-02  VLAN: 8
IVL Domain Heartbeat Timeout: 30
IVL Domain Capability: 8000000000000000
  Member: CASEY  MAC address: 02-11-11-00-00-00
  State: Active
  Port Group Status: Synchronized
  Global VSWITCH Status: Synchronized
  Heartbeat Count: <local>
  Member Capability: 8000000000000000 Maintenance Level: V632

  Member: JONES  MAC address: 02-22-22-00-00-00
  State: Active
  Port Group Status: Synchronized
  Global VSWITCH Status: Synchronized
  Heartbeat Count: 2
  Member Capability: 8000000000000000 Maintenance Level: V632
Ready; T=0.01/0.01 10:33:02

```

New Domain Section



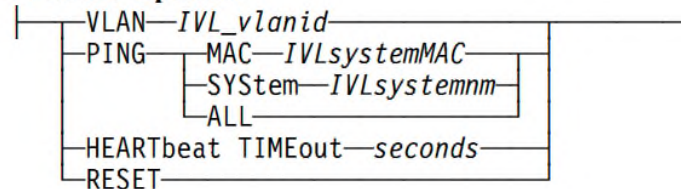
Optional



IVL Controls

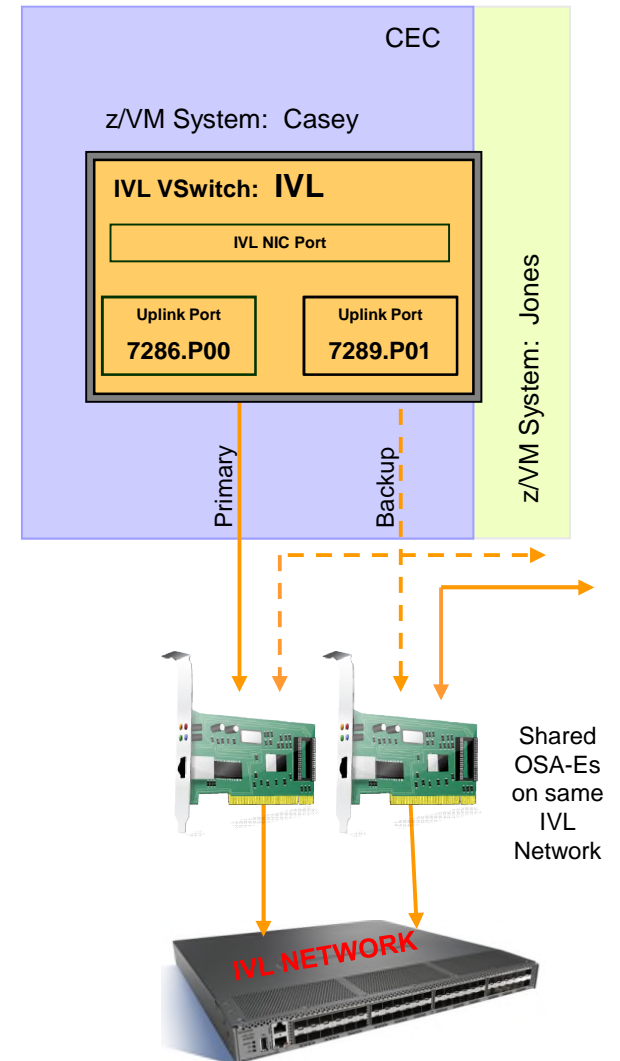
SET VSWITCH IVL IVLPORT -----| *IVLPort Options* |---

IVLPort Options:



- ▶ **VLAN** – Sets the VLAN ID associated with the IVL port (VLAN AWARE)
- ▶ **RESET** – Diagnostic which terminates and recreates IVL Port connection
- ▶ **PING** – Tests connectivity between z/VM Systems in the same IVL Domain
- ▶ **HEARTBEAT TIMEOUT** – Adjusts the frequency the local z/VM System confirms connectivity with z/VM Systems in the same IVL Domain

SET VSWITCH IVL IVLPORT PING ALL



Multi-VSwitch LAG Configuration Checklist

1

- Defining the IVL Domain

2

- System Connectivity to the IVL Network

3

- Verify System IVL Connectivity

4

- **Configure a LAG on the Physical Switch**

5

- Create a Shared Port Group

6

- Define a Global VSwitch

Step 4: Configure the Physical LAG

Each switch manufacture has its own unique method of configuring an LACP LAG Configuration. The following is an example of how a BNT G8052 Switch is configured:

1. After login to a switch telnet session with your *userid*, issue '**enable**' to enter admin mode
2. Issue '**conf t**' to enter configuration mode
3. Select the switch port to be configured into an LACP trunk group. Issue '**interface port x**' where x is the desired port
4. Select an *adminkey* to assign to the LACP trunk group. The key must be different from other trunk groups on the switch, but must be the same for all members of this trunk group. Issue '**lACP key <x>**' where x is a key from 1-65535.
5. Set LACP mode active: Issue: '**lACP mode active**'
6. Issue '**exit**' to leave the interface configuration mode
7. Repeat steps 3-6 for any additional switch ports you want to add to the LACP trunk group
8. When done with step 7, issue 'exit' to leave configuration mode

Note to save time: for step 3 above, if you are configuring multiple consecutive switch ports, you can issue '**interface port 4-7**' to configure ports 4-7 with a single set of LACP trunk group commands.

Multi-VSwitch LAG Configuration Checklist

1

- Defining the IVL Domain

2

- System Connectivity to the IVL Network

3

- Verify System IVL Connectivity

4

- Configure a LAG on the Physical Switch

5

- **Create a Shared Port Group**

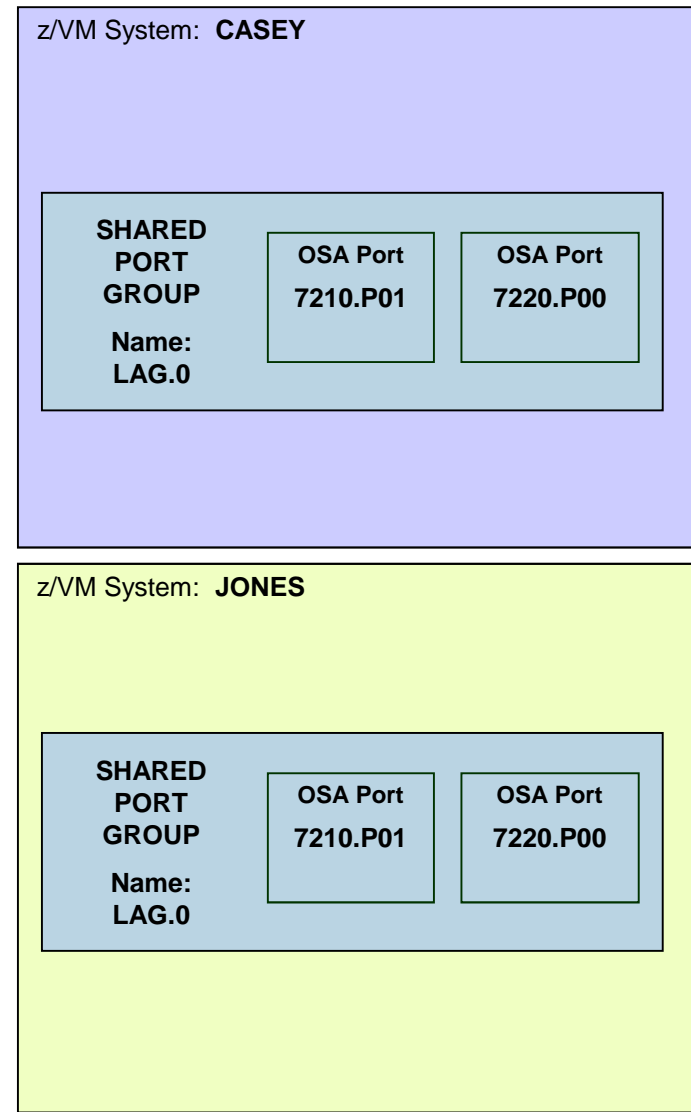
6

- Define a Global VSwitch

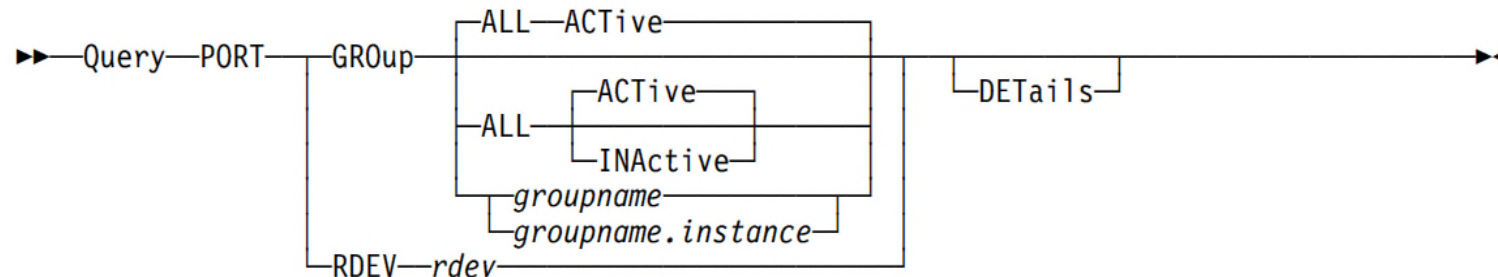
Step 5: Creation of a Shared Port Group

- **SET PORT GROUP LAG LACP ACTIVE SHARED**
 - ▶ To manage a Multi-VSwitch LAG Configuration, the SET PORT GROUP command must be created with a global scope by specifying: **LACP ACTIVE SHARED**.
 - ▶ Specify a single device on each OSA-Express feature that will be used to identify the adapter and reserve all devices for use by the port group..
 - ▶ The z/VM System will select the actual device numbers to be used on the target adapter.
 - ▶ Port Group name is increased to 10 characters.
 - 8 Character Name specified by the system administrator
 - 2 Character Instance ID assigned by z/VM
 - Default: *name.0*
 - Instance Number not used on SET Command
 - ▶ The z/VM System will automatically propagate Shared Port Group **LAG** to all active IVL Members in IVL Domain B.

- **SET PORT GROUP LAG JOIN 7210.P01 7220.P00**
 - ▶ Identify the OSA-Express features comprising the physical LAG.
 - ▶ The z/VM System will automatically add the OSA-Express features to the physical LAG.



Port Group Verification



- ▶ **ALL** – Return all active port groups defined in the system
- ▶ **ACTIVE** – Return only those port groups associated with a virtual switch
- ▶ **INACTIVE** – Return only those port groups NOT associated with a virtual switch
- ▶ **GROUP *groupname*** – Return only the specified port group
- ▶ **GROUP *groupname.instance*** – Return only the specified port group instance
- ▶ **RDEV** – Return only information for the specified real device
- ▶ **DETAILS** – Return additional information

Port Group Verification

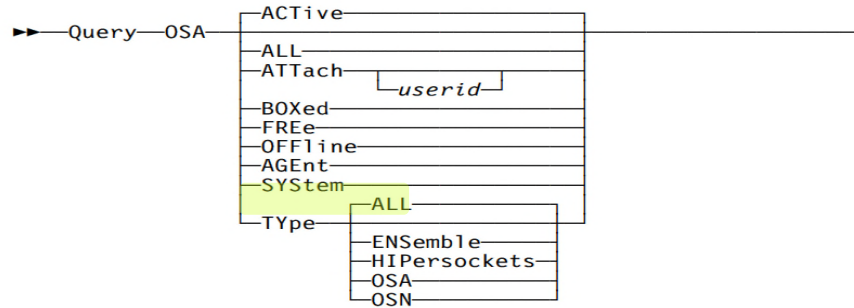
Q PORT GROUP LAG DET

```
Group: LAG.0          Inactive LACP Mode: Active   Shared
  VSWITCH <none>      Interval: 300   ifIndex:
2112
  RDEV: 7210.P01      Adapter ID: 28270007E386.0250
  RDEV: 7220.P00      Adapter ID: 28270007E386.0251
Member: CASEY
  Scope: Synchronized
  LAG Synchronization token: CE07E6D8FD203468
  Mode: Inactive
Member: JONES
  Scope: Synchronized
  LAG Synchronization token: CE07E6D8FD203468
  Mode: Inactive
Ready; T=0.01/0.01 10:03:50
```



New IVL Member Section

z/VM Automatic OSA Device Allocation



- ▶ All device numbers (subchannels) used within a SHARED Port Group are allocated by z/VM and not necessarily the ones specified on a SET PORT GROUP JOIN
- ▶ When a device is added to a SHARED Port Group, all devices on the same PCHID will be RESERVED for system use
- ▶ A SYSTEM device can't be attached or dedicated to a virtual machine
- ▶ z/VM will allocate a device from the SYSTEM list when creating a SHARED Port Group or an additional instance of a group
- ▶ All devices on the same adapter (PCHID) must be removed from all instances of a SHARED Port Group before they can be attached to a virtual machine

Q OSA SYSTEM

```

OSA 7210 SYSTEM , OSA 7211 SYSTEM , OSA 7212 SYSTEM , OSA 7220 SYSTEM
OSA 7221 SYSTEM , OSA 7222 SYSTEM , OSA 7223 SYSTEM , OSA 7224 SYSTEM
Ready; T=0.01/0.01 11:23:15
  
```


Multi-VSwitch LAG Configuration Checklist

1

- Defining the IVL Domain

2

- System Connectivity to the IVL Network

3

- Verify System IVL Connectivity

4

- Configure a LAG on the Physical Switch

5

- Create a Shared Port Group

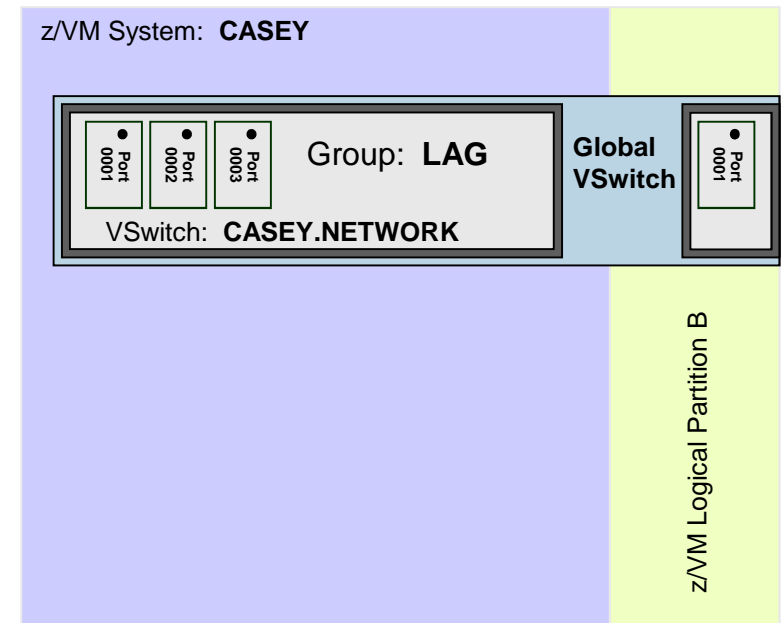
6

- Define a Global VSwitch

Step 6: Define a Global VSwitch

▪ DEFINE VSWITCH NETWORK GLOBAL ETHERNET GROUP LAG

- ▶ A Global VSwitch is a virtual switch which can span multiple z/VM System through the IVL Network.
 - A Global VSwitch connects to the IVL VSwitch's special IVL Port for Global VSwitch to VSwitch Communications.
 - Permits the Global VSwitch to manage and share the same physical LAG, share LACP and load balancing information.
- ▶ A Global VSwitch with the same 8 byte name needs to be defined in each z/VM System that is part of the Global VSwitch.
- ▶ A **Global ID** (*System Name.VSwitch*) is generated by the z/VM System.
- ▶ A Global VSwitch is created when its first Global VSwitch member is *defined* and is deleted only when the last Global VSwitch member is *detached*.
- ▶ Multiple Global VSwitches can be defined per z/VM System.
- ▶ Only a Global VSwitch can share a LAG configuration
- ▶ An instance of a Shared Port Group is created when it is configured to a virtual switch (LAG.0).



Repeat the same setup for VSwitch NETWORK on z/VM System JONES

Local VSwitch CASEY.NETWORK Port Group Connectivity

Q PORT GROUP

```
Group: LAG.0      Active   LACP Mode: Active   Shared
VSWITCH CASEY.NETWORK   Interval: 300   ifIndex: 2112
RDEV: 7210.P01 VDEV: 7210 Controller: DTCVSW1  ACTIVE
      Adapter ID: 28270007E386.0250
Uplink Port Connection:
MAC address: 02-11-11-00-00-03
RX Packets: 0          Discarded: 29          Errors: 0
TX Packets: 32         Discarded: 0           Errors: 0
RX Bytes: 0            TX Bytes: 4992
Device: 7210 Unit: 000 Role: DATA      Port: 2049
Partner Switch Capabilities: No_Reflective_Relay
LAG Port Controller: Active
RDEV: 7220.P00 VDEV: 7220 Controller: DTCVSW2  ACTIVE
      Adapter ID: 28270007E386.0251
Uplink Port Connection:
MAC address: 02-11-11-00-00-04
RX Packets: 0          Discarded: 700         Errors: 0
TX Packets: 27         Discarded: 0           Errors: 0
RX Bytes: 0            TX Bytes: 4212
Device: 7220 Unit: 000 Role: DATA      Port: 2050
Partner Switch Capabilities: No_Reflective_Relay
LAG Port Controller: Standby
Ready; T=0.01/0.01 10:37:10
```

New LAG Controller Status



Detail VSwitch CASEY.NETWORK Port Group Connectivity

Q PORT GROUP DET

```

Group: LAG.0      Active   LACP Mode: Active   Shared
VSWITCH CASEY.NETWORK Interval: 300   ifIndex: 2112
GROUP Information:
  PORT Information - Total Frames per Interval:
    Device Status Previous Current
    7210 Active 91 145
    7220 Active 383 424
  ROUTING Information - Frame Distribution per Interval:
    MAC Device Previous Current
    0 7210 0 0
    1 7220 0 0
    2 7210 0 0
    3 7220 0 0
    4 7210 0 0
    5 7220 0 0
    6 7210 0 0
    7 7220 0 0
RDEV: 7210.P01 VDEV: 7210 Controller: DTCVSW1 ACTIVE
Adapter ID: 28270007E386.0250
Uplink Port Connection:
MAC address: 02-11-11-00-00-03
RX Packets: 0 Discarded: 29 Errors: 0
TX Packets: 50 Discarded: 0 Errors: 0
RX Bytes: 0 TX Bytes: 7800
Device: 7210 Unit: 000 Role: DATA Port: 2049
Partner Switch Capabilities: No_Reflective_Relay
LAG Port Controller: Active
PROTOCOL Counters:
LACP RX: 288 Marker RX: 0
LACP TX: 46 Marker TX: 4 Timeouts: 0
ACTOR Information:
System ID: 32768,02-11-11-00-00-02 Oper Key: 2
Port Priority: 32768 Port: 2049 Group Key: 2
State: 3D - LACP_Active Slow AGG SYNC DIST COLL
PARTNER Information:
System ID: 32768,74-99-75-35-DF-00 Oper Key: 30
Port Priority: 32768 Port: 0004 Group Key: 30
State: 3D - LACP_Active Slow AGG SYNC DIST COLL

```

```

RDEV: 7220.P00 VDEV: 7220 Controller: DTCVSW2 ACTIVE
Adapter ID: 28270007E386.0251
Uplink Port Connection:
MAC address: 02-11-11-00-00-04
RX Packets: 0 Discarded: 1213 Errors: 0
TX Packets: 45 Discarded: 0 Errors: 0
RX Bytes: 0 TX Bytes: 7020
Device: 7220 Unit: 000 Role: DATA Port: 2050
Partner Switch Capabilities: No_Reflective_Relay
LAG Port Controller: Active
PROTOCOL Counters:
LACP RX: 256 Marker RX: 0
LACP TX: 45 Marker TX: 0 Timeouts: 0
ACTOR Information:
System ID: 32768,02-11-11-00-00-02 Oper Key: 2
Port Priority: 32768 Port: 2050 Group Key: 2
State: 3D - LACP_Active Slow AGG SYNC DIST COLL
PARTNER Information:
System ID: 32768,74-99-75-35-DF-00 Oper Key: 30
Port Priority: 32768 Port: 0005 Group Key: 30
State: 3D - LACP_Active Slow AGG SYNC DIST COLL

```

Member: CASEY

```

Scope: Synchronized
LAG Synchronization token: CE07E6D8FD203468
Mode: Connected

```

Member: JONES

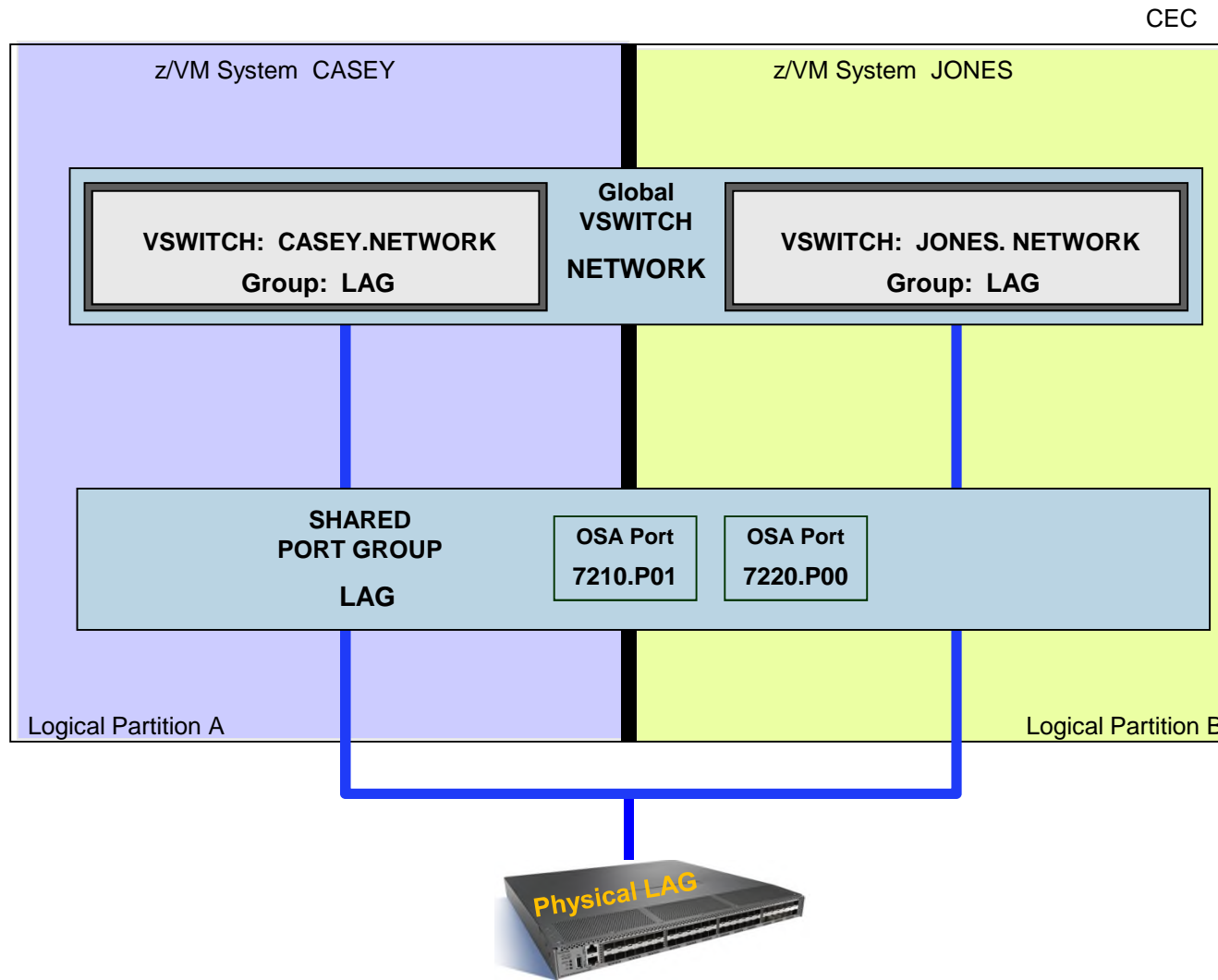
```

Scope: Synchronized
LAG Synchronization token: CE07E6D8FD203468
Mode: Connected

```

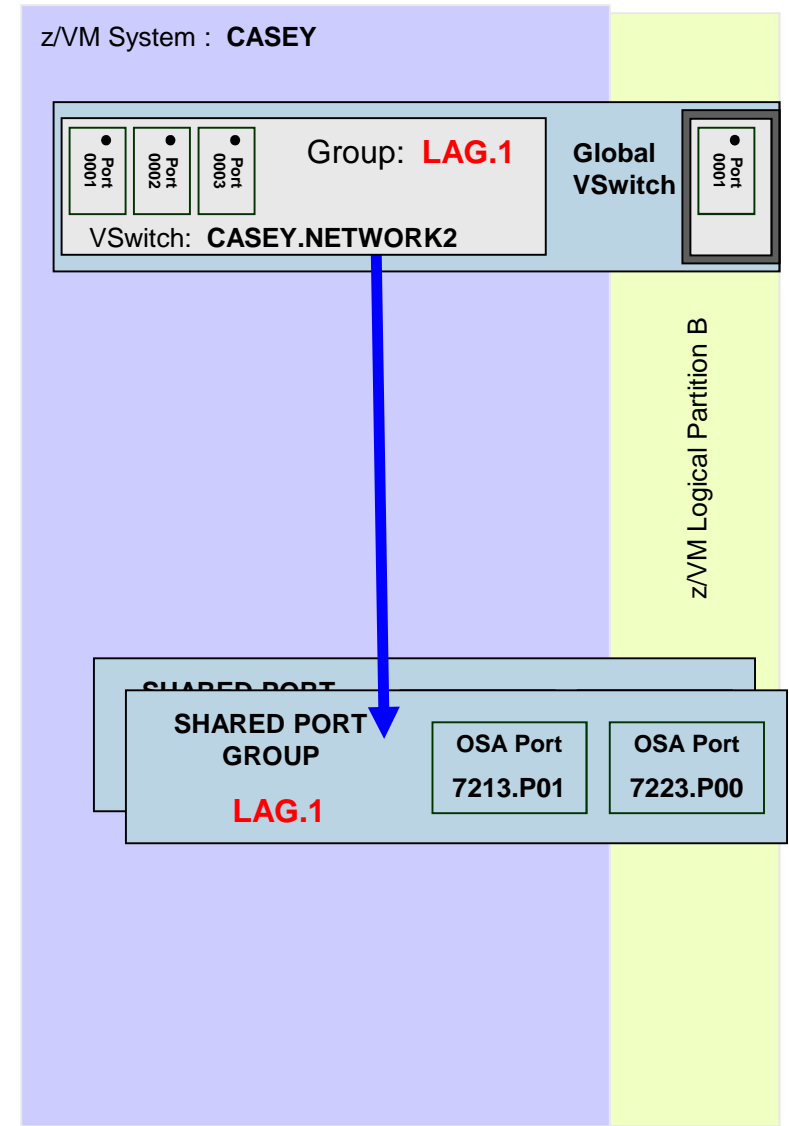
Ready; T=0.01/0.01 10:49:22

Multi-VSwitch LAG Configuration



Sharing A LAG within the Same z/VM System

- **DEFINE VSWITCH NETWORK2 GLOBAL ETHERNET GROUP LAG**
 - ▶ An instance of a Shared Port Group LAG is created (LAG.1) when it is configured to virtual switch NETWORK2.
 - ▶ z/VM will automatically allocate an OSA Triplet for each adapter within in the group from the available devices in the LPAR.
 - ▶ Up to four port group instances can be defined within an LPAR.
 - ▶ LAG.0 is the base instance of a Shared Port Group and is the only instance propagated to other IVL Members within the same domain.
 - ▶ Instance one to three (non-base) remain local to the system where it is being used.
 - ▶ The only difference between the base and its other instances are the device numbers allocated for each adapter within the LAG.



Q PORT GROUP

```

Group: LAG.0      Active   LACP Mode: Active   Shared
VSWITCH CASEY.NETWORK   Interval: 300   ifIndex: 2112
RDEV: 7210.P01 VDEV: 7210 Controller: DTCVSW1 ACTIVE
Adapter ID: 28270007E386.0250
Uplink Port Connection:
MAC address: 02-11-11-00-00-03
RX Packets: 0          Discarded: 29          Errors: 0
TX Packets: 2879       Discarded: 0           Errors: 0
RX Bytes: 0            TX Bytes: 449124
Device: 7210 Unit: 000 Role: DATA Port: 2049
Partner Switch Capabilities: No_Reflective_Relay
LAG Port Controller: Active
RDEV: 7220.P00 VDEV: 7220 Controller: DTCVSW2 ACTIVE
Adapter ID: 28270007E386.0251
Uplink Port Connection:
MAC address: 02-11-11-00-00-04
RX Packets: 0          Discarded: 83247       Errors: 0
TX Packets: 2874       Discarded: 0           Errors: 0
RX Bytes: 0            TX Bytes: 448344
Device: 7220 Unit: 000 Role: DATA Port: 2050
Partner Switch Capabilities: No_Reflective_Relay
LAG Port Controller: Active

```

```

Group: LAG.1      Active   LACP Mode: Active   Shared
VSWITCH CASEY.NETWORK2 Interval: 300   ifIndex: 2112
RDEV: 7213.P01 VDEV: 7213 Controller: DTCVSW1 ACTIVE
Adapter ID: 28270007E386.0250
Uplink Port Connection:
MAC address: 02-11-11-00-00-06
RX Packets: 0          Discarded: 0           Errors: 0
TX Packets: 4          Discarded: 0           Errors: 0
RX Bytes: 0            TX Bytes: 624
Device: 7213 Unit: 000 Role: DATA Port: 2049
Partner Switch Capabilities: No_Reflective_Relay
LAG Port Controller: Standby
RDEV: 7223.P00 VDEV: 7223 Controller: DTCVSW2 ACTIVE
Adapter ID: 28270007E386.0251
Uplink Port Connection:
MAC address: 02-11-11-00-00-07
RX Packets: 0          Discarded: 83158       Errors: 0
TX Packets: 0          Discarded: 0           Errors: 0
RX Bytes: 0            TX Bytes: 0
Device: 7223 Unit: 000 Role: DATA Port: 2050
Partner Switch Capabilities: No_Reflective_Relay
LAG Port Controller: Standby
Ready; T=0.01/0.01 08:31:38

```

Multi-VSwitch LAG Operations

Changing a Global VSwitch Member's Connectivity

SET VSWITCH NETWORK UPLINK [DISCONNECT|CONNECT]

- The operation of entirely removing or restoring a single Global VSwitch Member's connectivity from/to the physical LAG is provided through the existing SET VSWITCH UPLINK Command.
- The scope of the SET VSWITCH command remains local for Multi-VSwitch LAG support.

Removing or Adding a Physical Port from/to the LAG

SET PORT GROUP LAG [LEAVE|JOIN] 7220.P0

- ▶ Issued on either z/VM System CASEY or JONES
 - ▶ Executes on every Global VSwitch Member configured to use port group LAG
 - ▶ If any one of the VSwitches sharing port group LAG is unable to execute the command successfully, the specified operation will be backed out from all
-
- The operation of removing or adding a physical port (OSA) from/to a LAG is provided through the existing SET PORT GROUP Command.
 - Multi-VSwitch LAG support provides a global scope to the SET PORT GROUP command. This allows a Global VSwitch's Uplink Port to be managed with a single command issued from any IVL Domain Member.