

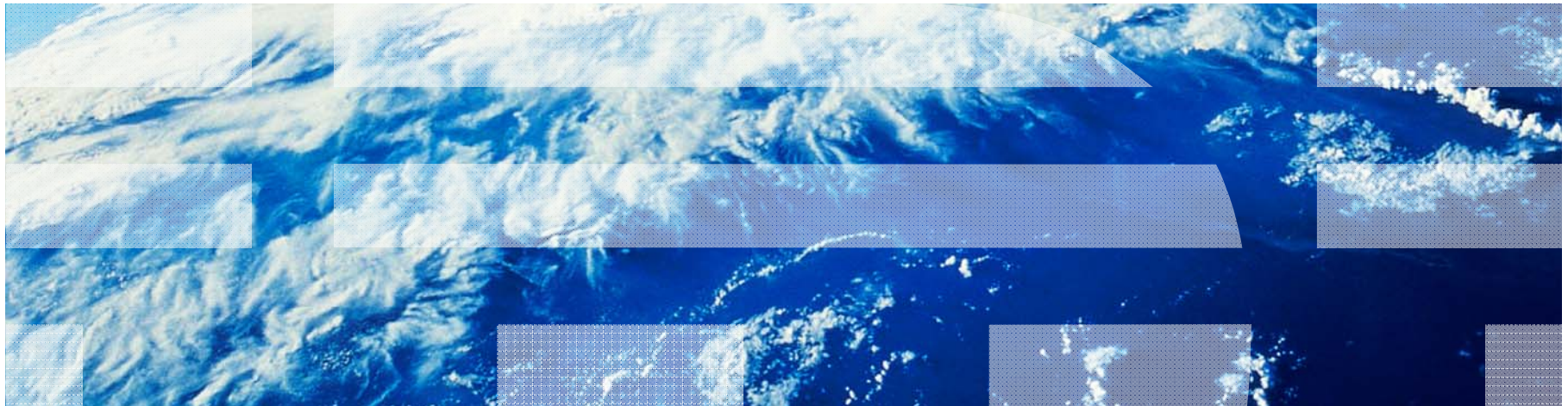
HillGang

October 2011



# z/VM 6.2 Overview

Richard Lewis  
rflewis@us.ibm.com



## Trademarks

**The following are trademarks of the International Business Machines Corporation in the United States, other countries, or both.**

z/VM® z10™ z/Architecture® zEnterprise™

Not all common law marks used by IBM are listed on this page. Failure of a mark to appear does not mean that IBM does not use the mark nor does it mean that the product is not actively marketed or is not significant within its relevant market.

Those trademarks followed by ® are registered trademarks of IBM in the United States; all others are trademarks or common law marks of IBM in the United States.

For a complete list of IBM Trademarks, see [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml):

**The following are trademarks or registered trademarks of other companies.**

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.

Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.

IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency, which is now part of the Office of Government Commerce.

\* All other products may be trademarks or registered trademarks of their respective companies.

### Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

## Disclaimer

The information contained in this document has not been submitted to any formal IBM test and is distributed on an "AS IS" basis without any warranty either express or implied. The use of this information or the implementation of any of these techniques is a customer responsibility and depends on the customer's ability to evaluate and integrate them into the operational environment. While each item may have been reviewed by IBM for accuracy in a specific situation, there is no guarantee that the same or similar results will be obtained elsewhere. Customers attempting to adapt these techniques to their own environments do so at their own risk.

In this document, any references made to an IBM licensed program are not intended to state or imply that only IBM's licensed program may be used; any functionally equivalent program may be used instead.

Any performance data contained in this document was determined in a controlled environment and, therefore, the results which may be obtained in other operating environments may vary significantly. Users of this document should verify the applicable data for their specific environments.

All statements regarding IBM's plans, directions, and intent are subject to change or withdrawal without notice, and represent goals and objectives only. This is not a commitment to deliver the functions described herein

## Acknowledgements

- The slides in this presentation come from several presentations authored by John Franciscovich, and Emily Hugenbruch. John and Emily are both members of z/VM Development in Endicott, NY

## Topics

- Introduction - z/VM Single System Image (SSI) Clusters
- Major Attributes of a z/VM SSI Cluster
- Planning and Creating a z/VM SSI Cluster

# *Introduction*

## Multi-system Virtualization with z/VM Single System Image (SSI)

- VMSSI Feature of z/VM 6.2
  
- Up to 4 z/VM instances (members) in a single system image (SSI) cluster
  - Same or different CECs
  
- Provides a set of shared resources for the z/VM systems and their hosted virtual machines
  - Managed as a single resource pool
  
- **Live Guest Relocation** provides virtual server mobility
  - Move virtual servers (guests) non-disruptively from one from one member of the cluster to another

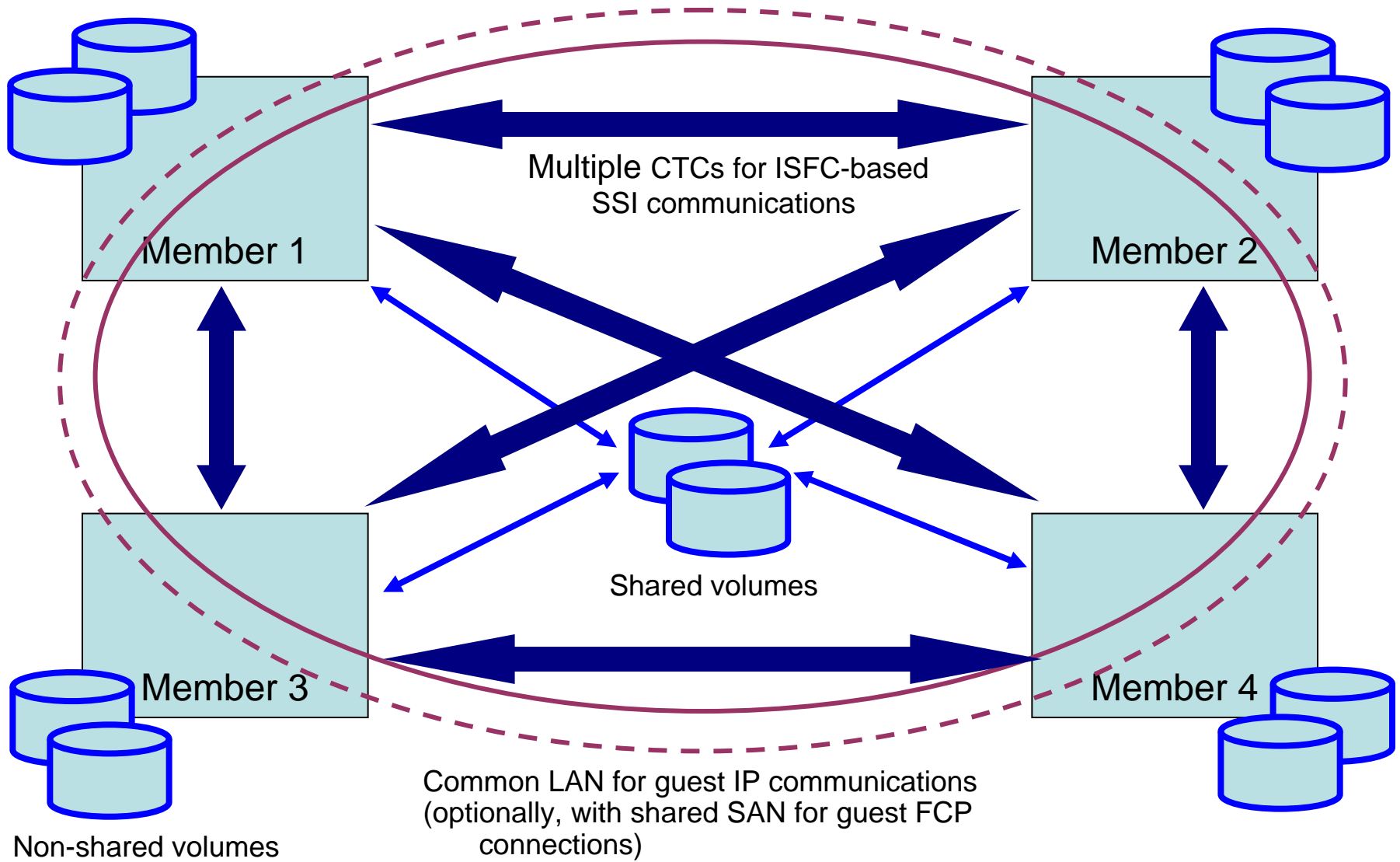
## z/VM Single System Image (SSI) Cluster

- Common resource pool accessible from all members
  - Shared disks for system and virtual server data
  - Common network access
  
- All members of an SSI cluster are part of the same ISFC collection
  - A system that is configured as a member of an SSI cluster joins the cluster during IPL
    - Verifies that its configuration is compatible with the cluster
    - Establishes communication with other members
  - Members leave the SSI cluster when they shut down
  
- CP validates and manages all resource and data sharing
  - Uses ISFC messages that flow across channel-to-channel connections between members
  - No virtual servers required
  
- **NOT** compatible with CSE (Cross System Extensions)
  - Cannot have SSI and CSE in same cluster
  - Disk sharing between an SSI cluster and a CSE cluster requires manual management of links
    - No automatic link protection or cache management

## Benefits of a z/VM SSI Cluster

- Facilitates horizontal growth of z/VM workloads
- Reduce effect of planned outages for z/VM and hardware maintenance
  - Less disruptive to virtual server workloads
- Simplifies system management of a multi-z/VM environment
  - Concurrent installation of multiple-system cluster
  - Single maintenance stream
- Enhances workload balancing

# z/VM SSI Cluster



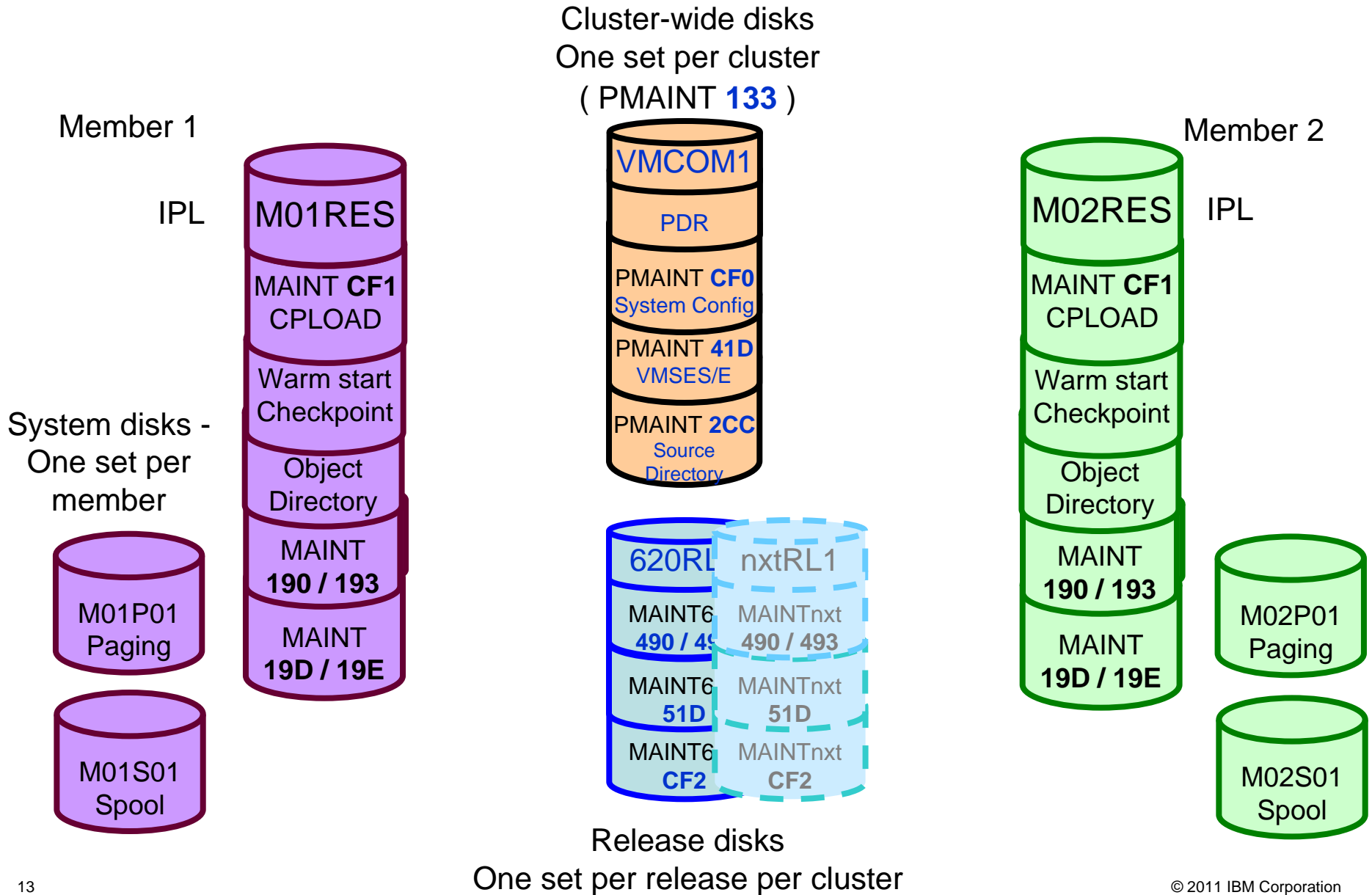
***Major Attributes of a  
z/VM SSI Cluster***

## Multisystem Installation

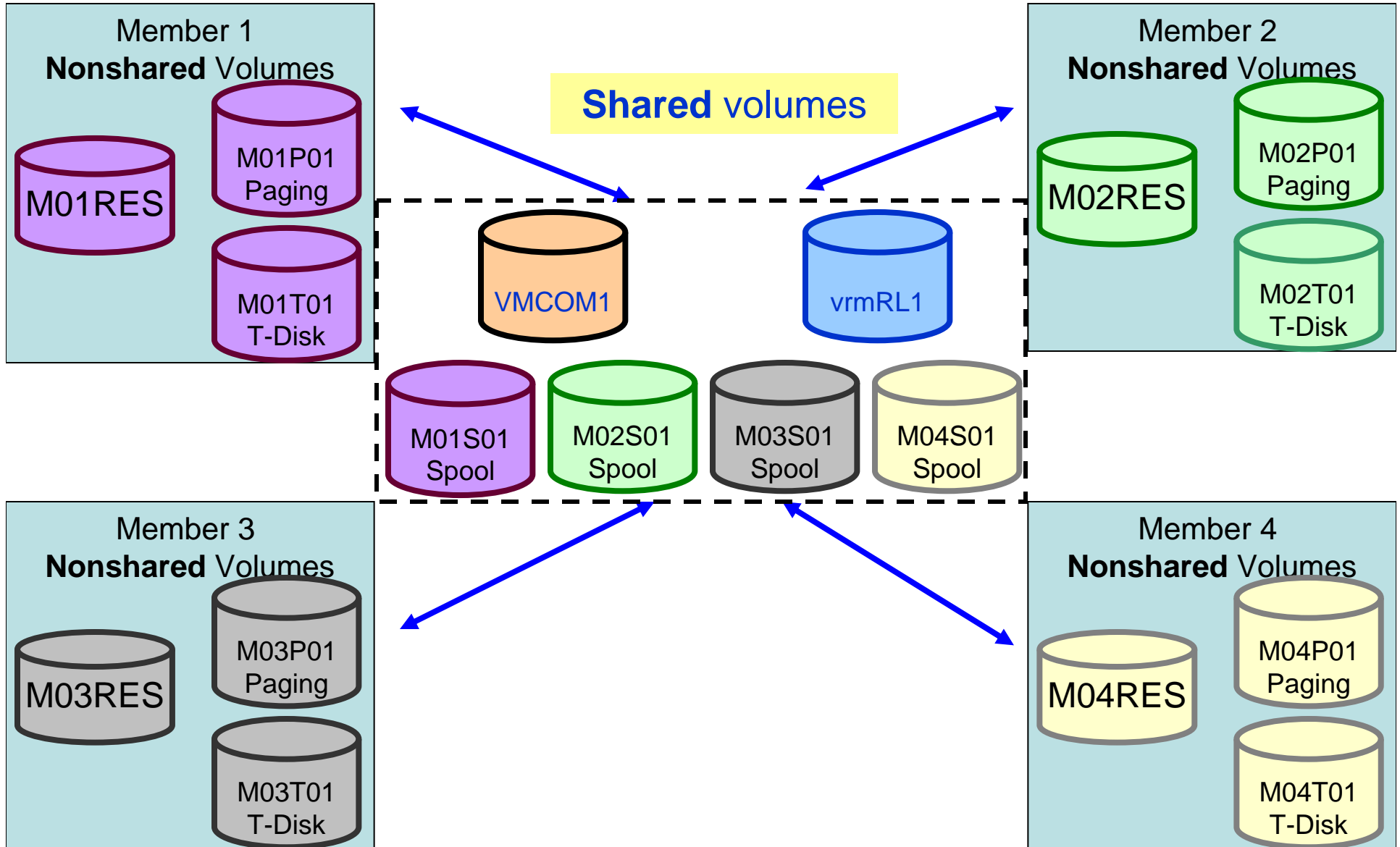
```
Select a System Type: Non-SSI or SSI (SSI requires the SSI feature)
  Non-SSI Install:      System Name _____
  X SSI Install:        Number of Members 4      SSI Cluster Name SAMPLE
```

- SSI cluster can be created with a single z/VM install
  - Cluster information is specified on installation panels
    - Member names
    - Volume information
    - Channel-to-channel connections for ISFC
  - Specified number of members are installed and configured as an SSI cluster
    - Shared system configuration file
    - Shared source directory
- Non-SSI single system installation also available
  - System resources defined in same way as for SSI
    - Facilitates later conversion to an SSI cluster

# DASD Volumes and Minidisks



# DASD Planning – Non-Shared and Shared System Volumes



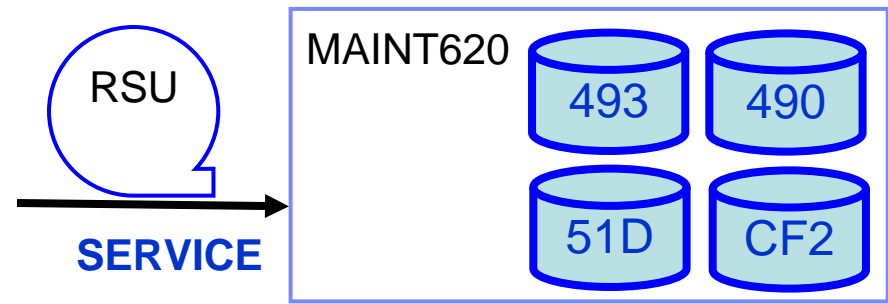
## Applying Service

### Single Maintenance Stream per release

1. Logon to MAINT620 on *either* member and run **SERVICE**

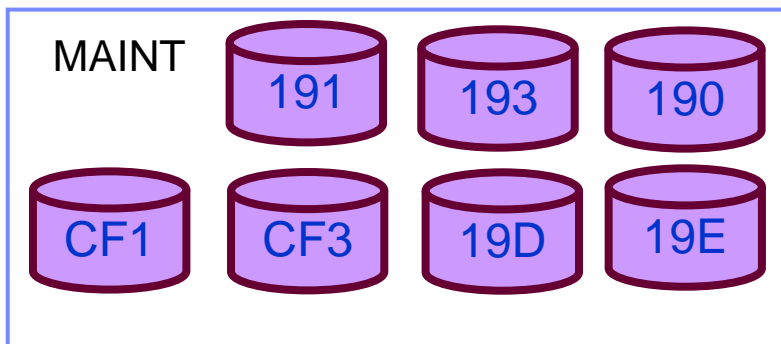
### Service applied privately to each member

2. Logon to MAINT620 on Member 1 and **PUT2PROD**
3. Logon to MAINT620 on Member 2 and **PUT2PROD**

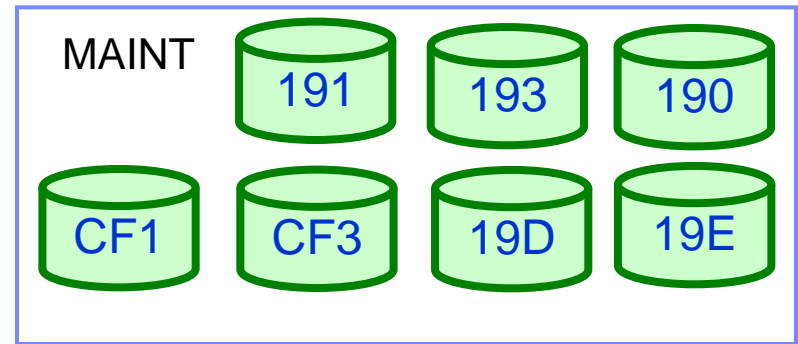


**PUT2PROD**

**PUT2PROD**



Member 1



Member 2

## Shared System Configuration File

- Resides on new shared parm disk
  - PMAINT CF0
  
- Can include member-specific configuration statements
  - Record qualifiers
    - New BEGIN/END blocks
  
- Define each member's system name
  - Enhanced SYSTEM\_IDENTIFIER statement
    - LPAR name can be matched to define system name

```
System_Identifier LPAR LP1 VMSYS01
```
  
    - System name can be set to the LPAR name

```
System_Identifier LPAR * &LPARNAME
```
  
- Define cluster configuration (cluster name and member names)

```
SSI CLUSTERA PDR_VOLUME VMCOM1,  
SLOT 1 VMSYS01,  
SLOT 2 VMSYS02,  
SLOT 3 VMSYS03,  
SLOT 4 VMSYS04
```

## Shared System Configuration File...

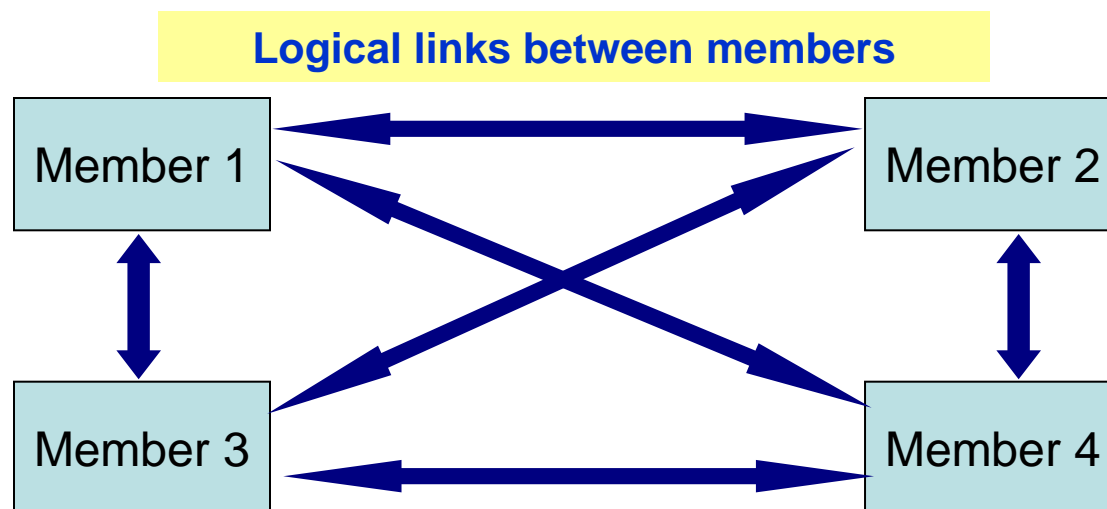
- Identify direct ISFC links between members
  - One set of statements for each member

```
VMSYS01: BEGIN
          ACTIVATE ISLINK 912A /* Member 1 TO Member 2 */
          ACTIVATE ISLINK 913A /* Member 1 TO Member 3 */
          ACTIVATE ISLINK 914A /* Member 1 TO Member 4 */
VMSYS01: END
```

- Define CP Owned volumes
  - Shared
    - SSI common volume
    - Spool
  - Private
    - Sysres
    - Paging
    - Tdisk

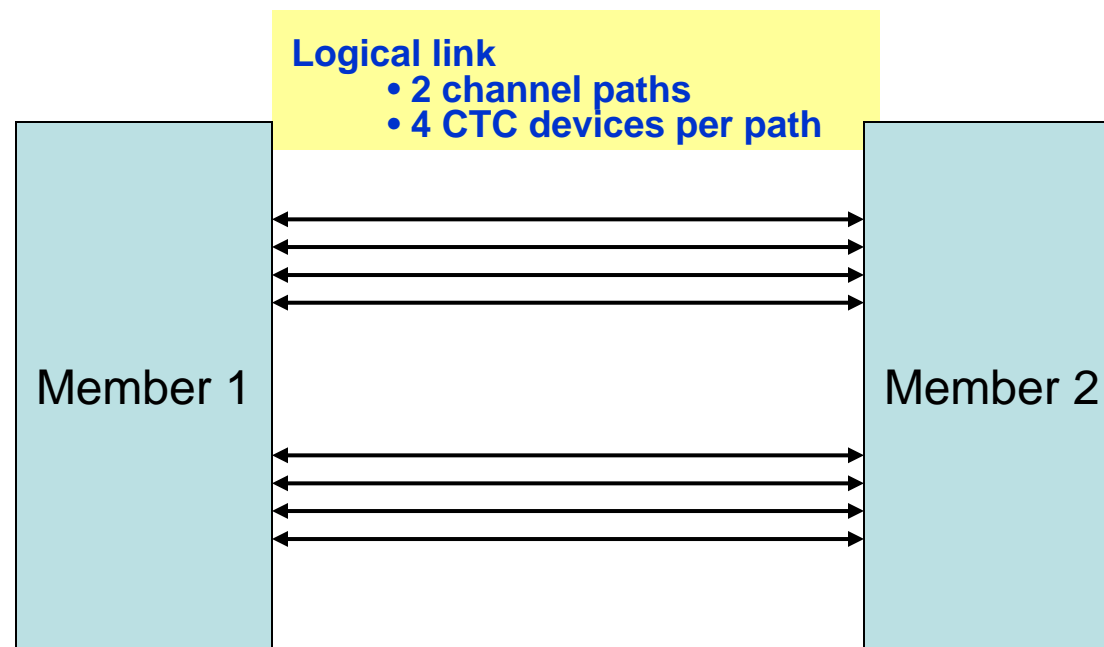
## CTC Connections

- Each member of an SSI cluster must have a direct ISFC connection to every other member (logical link)
- Logical links are composed of 1-16 CTC connections
  - FICON channel paths
  - May be switched or unswitched
- Use multiple CTCs distributed on multiple FICON channel paths between each pair of members
  - Avoids write collisions that affect link performance
  - Avoids severing logical link if one channel path is disconnected or damaged
- *Recommended practice:* Use same real device number for same CTC on each member



## CTC Connections – How Many Do I Need?

- 4 CTCs per FICON channel path provides most efficient ISFC data transfer
- For large guests, relocation time and quiesce time can be improved with more channel paths\*
  - Up to 4 channel paths, with 4 CTCs each
    - *Based on early performance measurements*
    - *Additional factors affect relocation and quiesce times*





## Shared System Configuration File – CP-Owned Volumes...

```
/*
/*
/* PAGE & TDISK VOLUMES */
/* To avoid interference with spool volumes and to */
/* automatically have all unused slots defined as */
/* "Reserved", begin with slot 255 and assign them in */
/* descending order. */
/*
*****/

VMSYS01: BEGIN
          CP_Owned Slot 254 M01T01
          CP_Owned Slot 255 M01P01
VMSYS01: END

VMSYS02: BEGIN
          CP_Owned Slot 254 M02T01
          CP_Owned Slot 255 M02P01
VMSYS02: END

VMSYS03: BEGIN
          CP_Owned Slot 254 M03T01
          CP_Owned Slot 255 M03P01
VMSYS03: END

VMSYS04: BEGIN
          CP_Owned Slot 254 M04T01
          CP_Owned Slot 255 M04P01
VMSYS04: END
```

## Persistent Data Record (PDR)

- Cross-system serialization point on disk
  - Must be a shared 3390 volume (VMCOM1)
  - Created and viewed with new FORMSSI utility
  
- Contains information about member status
  - Used for health-checking
  
- Heartbeat data
  - Ensures that a stalled or stopped member can be detected

## Ownership Checking – CP-Owned Volumes

- Each CP-owned volume in an SSI cluster will be marked with ownership information
  - Cluster name
  - System name of the owning member
  - The marking is created using CPFMTXA
  
- Ensures that one member does not allocate CP data on a volume owned by another member
  - Warm start, checkpoint, spool, paging, temporary disk, directory
  
- No need to worry about OWN and SHARED on CP\_OWNED definitions
  - Ignored on SSI members
  
- QUERY CPOWNED enhanced to display ownership information

## Ownership Checking – CP-Owned Volumes...

```
cpfmtxa
```

```
ENTER FORMAT, ALLOCATE, LABEL, OWNER OR QUIT:
```

```
owner
```

```
ENTER THE VDEV TO BE PROCESSED OR QUIT:
```

```
c4a8
```

```
ENTER THE VOLUME LABEL FOR DISK E000:
```

```
m01s01
```

```
ENTER THE OWNING SSI NAME (OR NOSSI) FOR DISK E000:
```

```
clustera
```

```
ENTER THE OWNING SYSTEM NAME (OR NOSYS) FOR DISK E000:
```

```
vmsys01
```

```
query cpowned
```

SLOT	VOL-ID	RDEV	TYPE	STATUS	SSIOWNER	SYSOWNER
				.		
				.		
				.		
10	M01S01	C4A8	OWN	ONLINE AND ATTACHED	CLUSTERA	VMSYS01

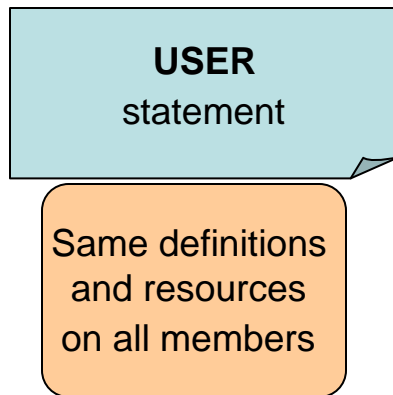
## Defining Virtual Machines – Shared Source Directory

- All user definitions in a single shared source directory
  
- Run DIRECTXA on each member
  
- No system affinity (SYSAFFIN)
  
- Identical object directories on each member
  
- Single security context
  - Each user has same access rights and privileges on each member

***Using a directory manager is strongly recommended!***

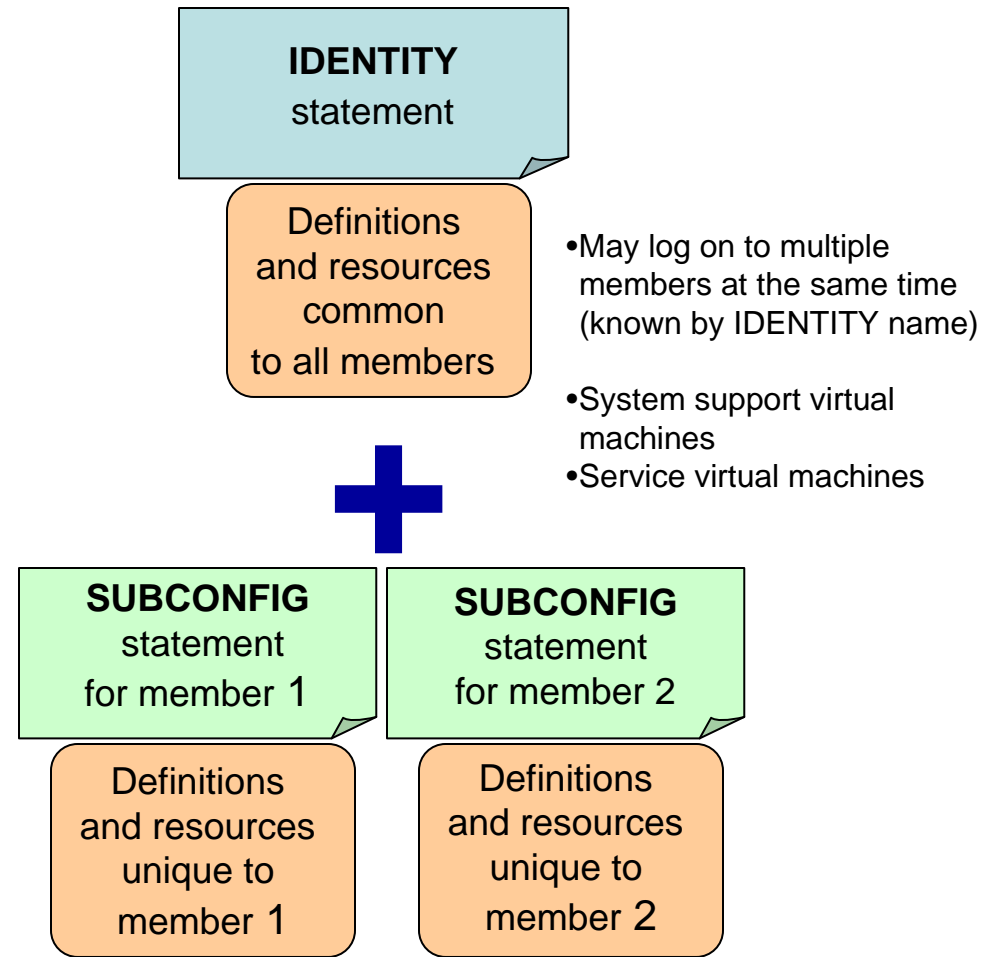
## Shared Source Directory – Virtual Machine Definition Types

### Single Configuration Virtual Machine (traditional)



- May log on to any member
  - Only one member at a time
- General Workload
  - Guest Operating Systems
  - Service virtual machines requiring only one logon in the cluster

### Multiconfiguration Virtual Machine (new)



## New Directory Layout

- IBM-supplied directory will be significantly different than in previous releases
  - Both SSI and non-SSI installations
  - Directory for non-SSI installations will be in "SSI-ready" format
    - Facilitate future SSI deployment
  
- Many of the IBM-supplied userids will be defined as multiconfiguration virtual machines
  
- Determine if any of your guests should be defined as multiconfiguration virtual machines
  - Most will be single-configuration virtual machines
  - Userids defined on SYSTEM\_USERIDS statements will usually be multiconfiguration virtual machines
  
- Merge your user definitions into the IBM-supplied directory

## Multiconfiguration Virtual Machine Definition

```
IDENTITY MAINT      MAINTPAS      128M 1000M ABCDEFG
```

```
BUILD MAINT-1 USING SSIMEMB1
BUILD MAINT-2 USING SSIMEMB2
BUILD MAINT-3 USING SSIMEMB3
BUILD MAINT-4 USING SSIMEMB4
```

```
CONSOLE 009 3215 T
SPOOL 00C 2540 READER *
SPOOL 00D 2540 PUNCH A
SPOOL 00E 1403 A
LINK     USER1      2CC 2CC RR
LINK     USER1      551 551 RR
```

These statements apply to all instances of MAINT on all members

```
SUBCONFIG MAINT-1
MDISK 0191 3390 1000 20 MNTVL1 WR
MDISK CF1  3390 100  20 M01RES RR
* END OF MAINT-1
  b
```

These statements only apply to MAINT on member SSIMEMB1

```
SUBCONFIG MAINT-2
MDISK 0191 3390 1000 20 MNTVL2 WR
MDISK CF1  3390 100  20 M02RES RR
* END OF MAINT-2
```

These statements only apply to MAINT on member SSIMEMB2

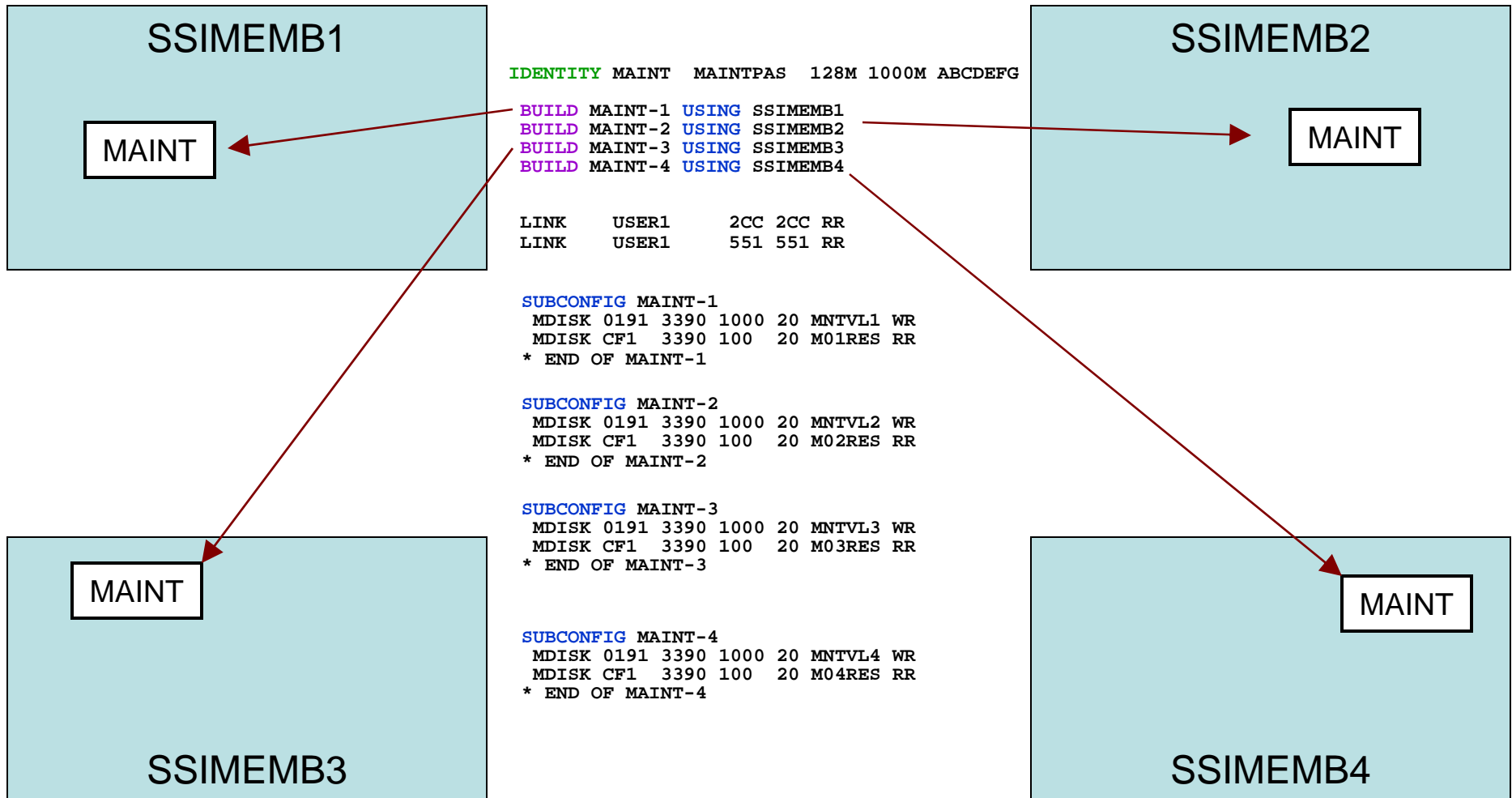
```
SUBCONFIG MAINT-3
MDISK 0191 3390 1000 20 MNTVL3 WR
MDISK CF1  3390 100  20 M03RES RR
* END OF MAINT-3
```

These statements only apply to MAINT on member SSIMEMB3

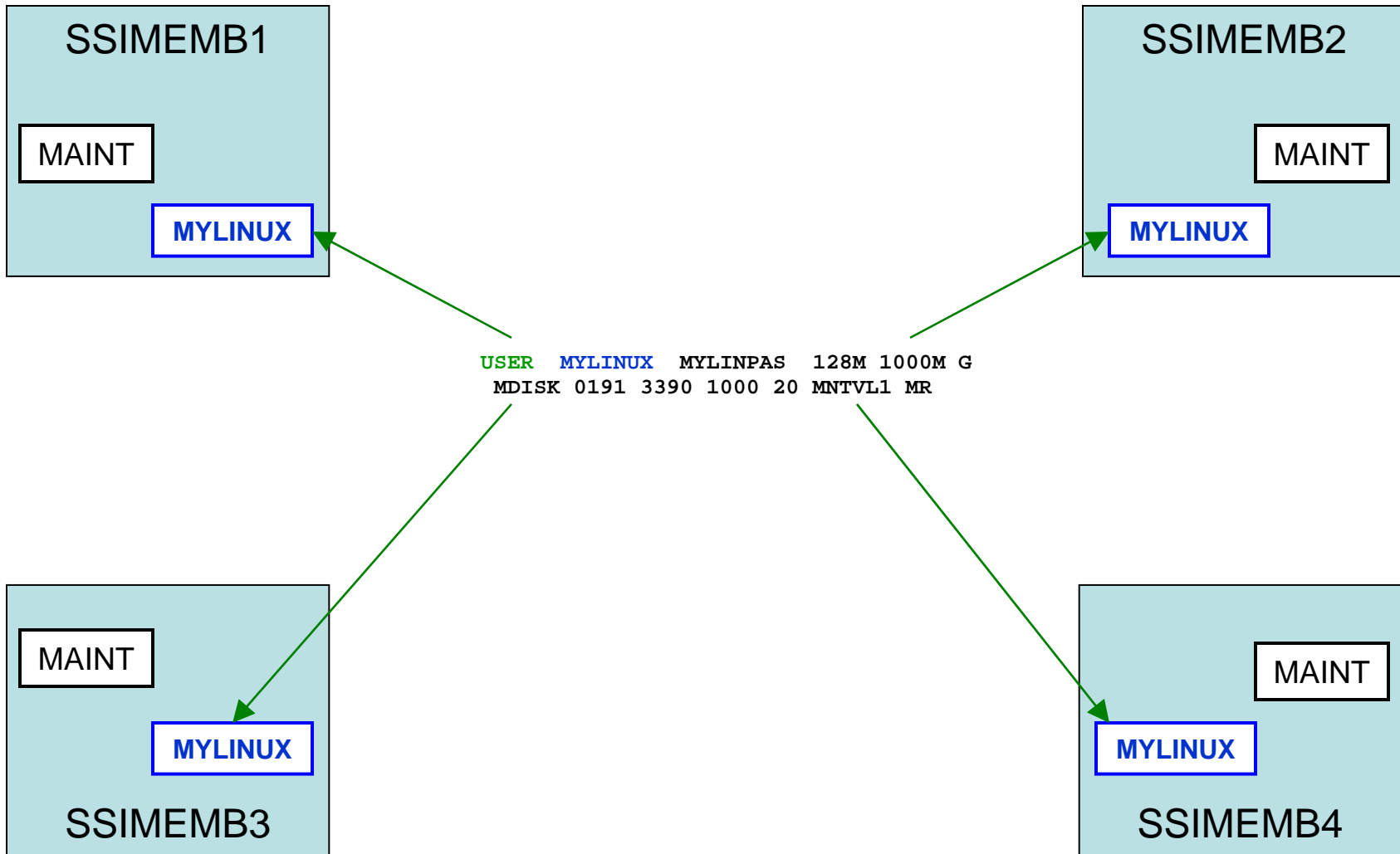
```
SUBCONFIG MAINT-4
MDISK 0191 3390 1000 20 MNTVL4 WR
MDISK CF1  3390 100  20 M04RES RR
* END OF MAINT-4
```

These statements only apply to MAINT on member SSIMEMB4

# Multiconfiguration Virtual Machines



# Single Configuration Virtual Machines



## New MAINT Userids

<b>MAINT</b>	<b>PMAINT</b>	<b>MAINT620</b>
Multi Configuration Virtual Machine	Single Configuration Virtual Machine	Single Configuration Virtual Machine
Owns CF1, CF3 parm disks, 190, 193, 19D, 19E, 401, 402, 990 CMS disks	Owns CF0 parm disk, 2CC, 550, 551 disks	Owns the service disks (e.g., 490, 493, 49D) and the CF2 parm disk
Use for work on a particular member, such as attaching devices, or relocating guests	Use for updating the system config, or for SSI-wide work, e.g., defining relocation domains	Use for applying 6.2.0 service. The CF2 parm disk contains 6.2.0 CPLOAD modules.

## Cross-System Spool

- Spool files are managed cooperatively and shared among all members of an SSI cluster
- Single-configuration virtual machines (most users) have a single logical view of all of their spool files
  - Access, manipulate, and transfer all files from any member where they are logged on
    - Regardless of which member they were created on
- Multiconfiguration virtual machines do not participate in cross-system spool
  - Each instance only has access to files created on the member where it is logged on
- All spool volumes in the SSI cluster are shared (R/W) by all members
  - Each member creates files on only the volumes that it owns
  - Each member can access and update files on all volumes

SLOT	VOL-ID	RDEV	TYPE	STATUS	SSIOWNER	SYSOWNER
10	M01S01	C4A8	OWN	ONLINE AND ATTACHED	CLUSTERA	VMSYS01
11	M02S01	C4B8	SHARE	ONLINE AND ATTACHED	CLUSTERA	VMSYS02
12	M01S02	C4A9	OWN	ONLINE AND ATTACHED	CLUSTERA	VMSYS01
13	M02S02	C4B9	SHARE	ONLINE AND ATTACHED	CLUSTERA	VMSYS02
14	M01S03	C4AA	DUMP	ONLINE AND ATTACHED	CLUSTERA	VMSYS01
15	M02S03	C4BA	DUMP	ONLINE AND ATTACHED	CLUSTERA	VMSYS02
16	-----	----	-----	RESERVED	-----	-----

## Cross-System SCIF

- Cross-System SCIF (Single Console Image Facility)
  - Allows one virtual machine (secondary user) to monitor and control one or more disconnected virtual machines (primary users)
    - CONSOLE statement in directory
    - SET SECUSER command
    - SET OBSERVER command
  - Secondary and primary users can be logged on different members of an SSI cluster
- Some restrictions for multiconfiguration virtual machines (MCVM):

Primary User or Observee	SECUSER or Observer	If Local	If Remote
<b>SCVM</b>	<b>SCVM</b>	<b>Yes</b>	<b>Yes</b>
<b>SCVM</b>	<b>MCVM</b>	<b>Yes</b>	<b>Yes</b>
<b>MCVM</b>	<b>SCVM</b>	<b>Yes</b>	<b>No</b>
<b>MCVM</b>	<b>MCVM</b>	<b>Yes</b>	<b>No</b>

## Cross-System CP Commands

- Virtual machines on other members can be the target of some CP commands
  - Single-configuration virtual machines are usually found wherever they are logged on
  - Multiconfiguration virtual machines require explicit targeting
  
- **AT *sysname*** operand for the following commands
  - MESSAGE (MSG)
  - MSGNOH
  - SEND
  - SMSG
  - WARNING

**MSG userid AT *sysname***

  - CMS TELL and SENDFILE commands require RSCS in order to communicate with multiconfiguration virtual machines on other members
  
- **AT** command can be used to issue most privileged commands on another active member

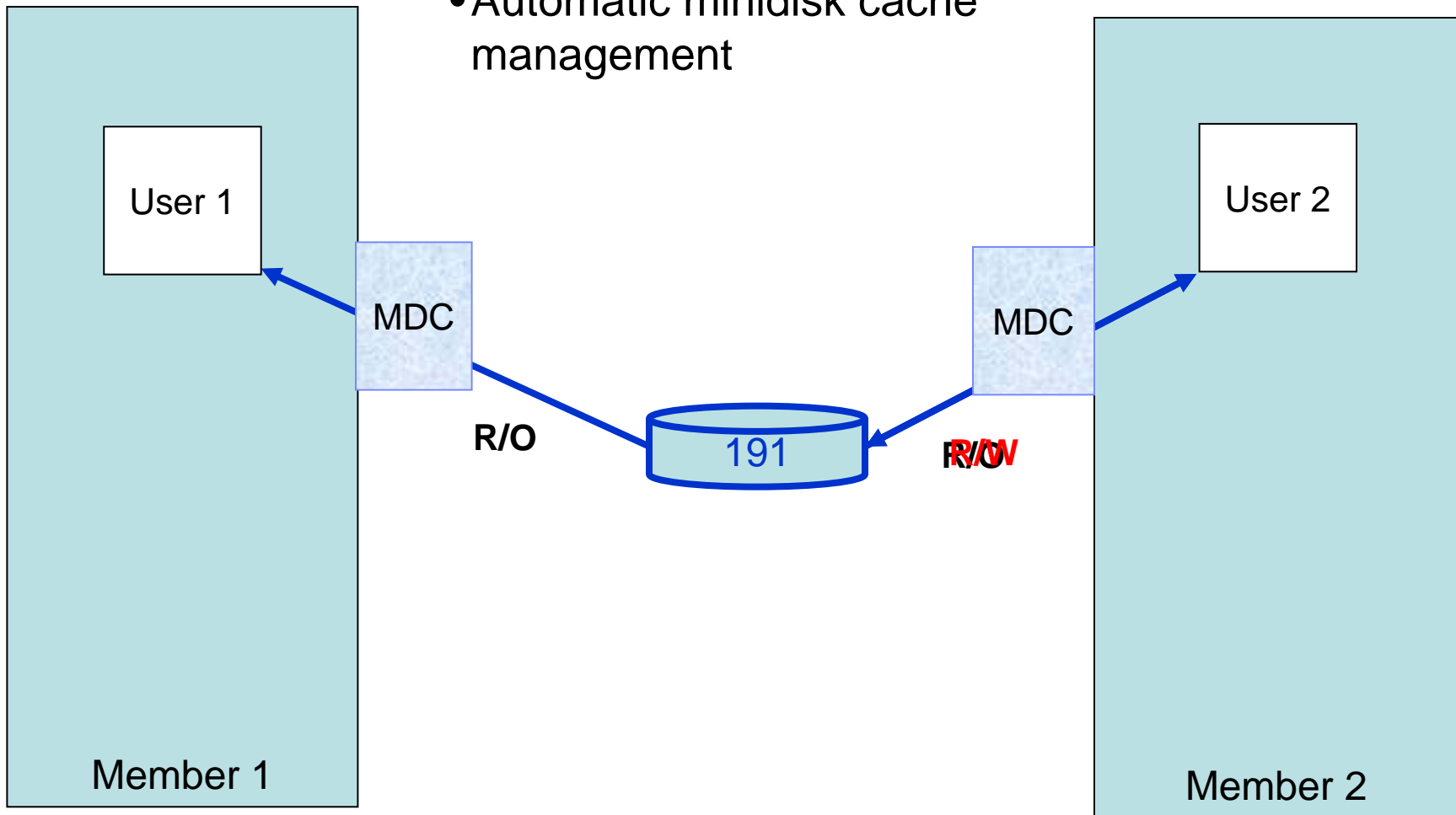
**AT *sysname* CMD *cmdname***

## Cross-System Minidisk Management

- Minidisks can either be shared across all members or restricted to a single member
  - CP checks for conflicts throughout the cluster when a link is requested
  
- Virtual reserve/release for fullpack minidisks is supported across members
  - Only supported on one member at a time for non-fullpack minidisks
  
- Volumes can be shared with systems outside the SSI cluster
  - **SHARED YES** on RDEVICE statement or SET RDEVICE command
  - **Link conflicts must be managed manually**
  - Not eligible for minidisk cache
  - **Use with care**

## Cross-System Minidisk Management...

- Automatic minidisk cache management



## Real Device Management

- Unique identification of real devices within an SSI cluster
  - Ensures that all members are using the same physical devices where required
  
- CP generates an equivalency identifier (EQID) for each disk volume and tape drive
  - Physical device has same EQID on all members
  
- EQID for network adapters (CTC, FCP, OSA, Hipersockets) must be defined by system administrator
  - Connected to same network/fabric
  - Conveying same access rights
  
- EQIDs used to select equivalent device for live guest relocation and to assure data integrity

## Network Planning

- All members should have identical network connectivity
  - Connected to same physical LAN segments
  - Connected to same SAN fabric
  
- Assign equivalence identifiers (EQIDs) to all network devices
  - Devices assigned same EQID on each member must be
    - same type
    - have the same capabilities
    - have connectivity to the same destinations

## Network Planning – Virtual Switches

- Define virtual switches with same name on each member
- For relocating guests:
  - Source and destination virtual switch guest NIC and port configurations must be equivalent
    - Port type
    - Authorizations (access, VLAN, promiscuous mode)
  - Source and destination virtual switches must be equivalent
    - Name and type
    - VLAN settings
    - Operational UPLINK port with matching EQID
    - Device and port numbers need not match, but connectivity to the same LAN segment is required

## Network Planning – MAC Addresses

- MAC address assignments are coordinated across an SSI cluster
  - VMLAN statement
    - MACPREFIX must be set to different value for each member
      - ♦ Default is 02-xx-xx where xx-xx is "system number" of member (e.g., 02-00-01 for member 1)
    - USERPREFIX must be set for SSI members
      - ♦ Must be identical for all members
      - ♦ Must not be equal to any member's MACPREFIX value
      - ♦ Default is 02-00-00
    - MACIDRANGE is ignored in an SSI cluster
      - ♦ Because MAC assignment is coordinated among members
    - Example:

```
VMSYS01: VMLAN MACPREFIX 021111 USERPREFIX 02AAAA
VMSYS02: VMLAN MACPREFIX 022222 USERPREFIX 02AAAA
VMSYS03: VMLAN MACPREFIX 023333 USERPREFIX 02AAAA
VMSYS04: VMLAN MACPREFIX 024444 USERPREFIX 02AAAA
```

## Live Guest Relocation

- Relocate a running Linux virtual server (guest) from one member of an SSI cluster to another
  - Load balancing
  - Moving workload off a member requiring maintenance
  
- Relocating guests continue to run on source member until destination is ready
  - Briefly quiesced
  - Resumed on destination member
  
- New **VMRELOCATE** command will initiate and manage guest relocations
  - Relocation capacity determined by various factors (e.g. system load, ISFC bandwidth, etc.)
  
- A guest to be relocated must meet eligibility requirements, including:
  - It must be logged on but disconnected
  - Architecture and functional environment on destination must be comparable
  - Destination member must have capacity to accommodate the guest
  - Devices and resources needed by guest must be shared and available on destination
  
- Relocation domains can be used to define a set of members among which virtual machines can relocate freely

## Guest Configuration for Live Guest Relocation

- In order to be eligible to relocate, a guest must be:
  - Defined as a single configuration virtual machine
  - Running in an ESA or XA virtual machine in ESA/390 or z/Architecture mode
  - Logged on and disconnected
  - Running only type CP or type IFL virtual processors
  - IPLed from either a
    - Device
    - Named saved system (NSS)
  
- If a guest is using a DCSS or NSS:
  - Identical NSS or DCSS must be available on the destination member
  - It cannot have the following types of page ranges
    - SW (shared write)
    - SC (shared with CP)
    - SN (shared with no data)

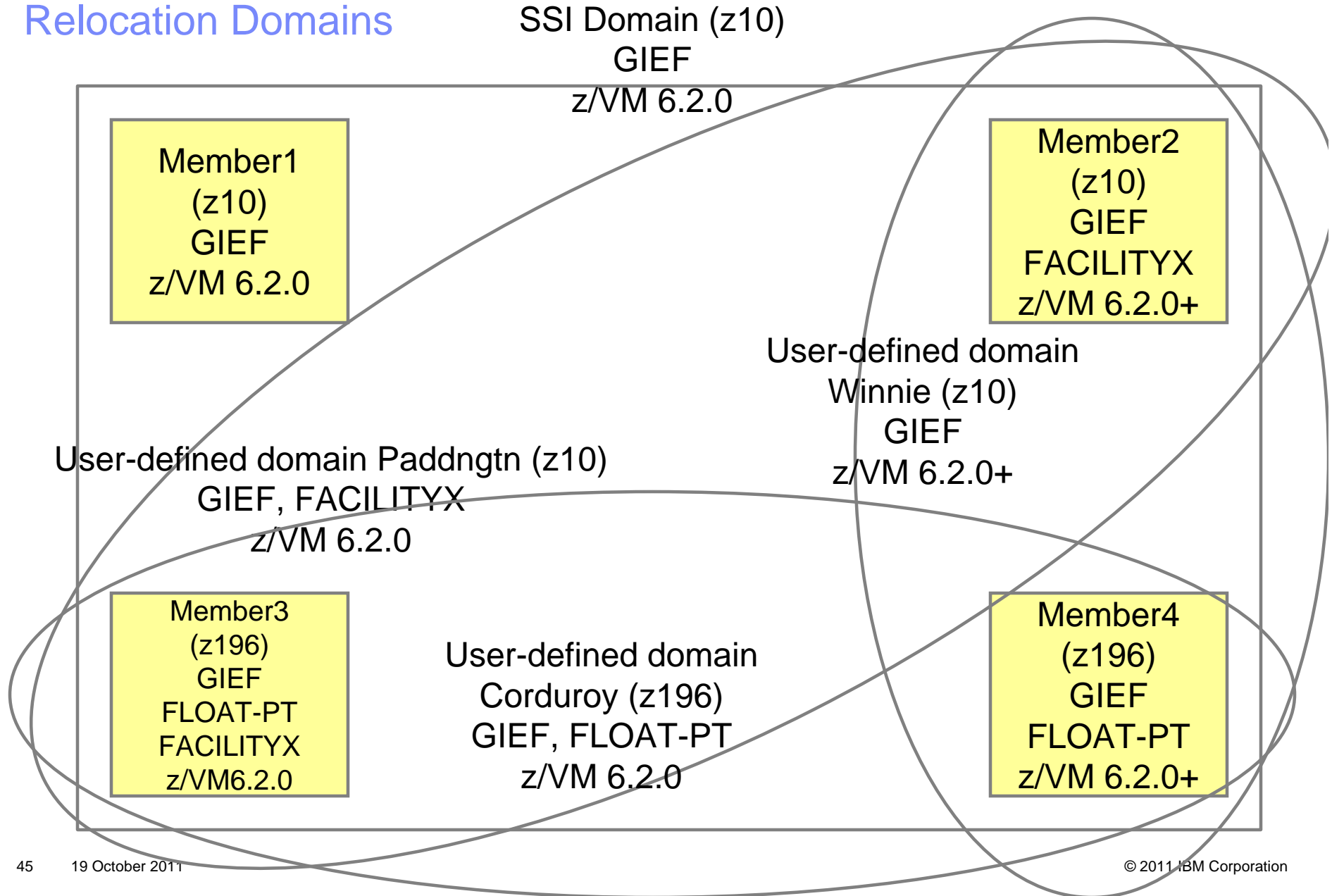
## Guest Configuration for Live Guest Relocation (cont.)

- A guest can relocate if it has any of the following:
  - Dedicated devices
    - Equivalent devices and access must be available on destination member
  - Private Vdisks
  - No open spool files other than console files
  - VSWITCHes
    - Equivalent VSWITCH and network connectivity must be available on destination
  
- A relocating guest can be using any of the following facilities:
  - Cryptographic adapter
    - Crypto cards for shared domains on source and destination must be same AP type
  - Virtual machine time bomb (Diag x'288')
  - IUCV connections to \*MSG and \*MSGALL CP system services
  - Application monitor record (APPLDATA) collection
    - If guest buffer is not in a shared DCSS
  - Single Console Image Facility
  - Collaborative Memory Management Assist (CMMA)

## Relocation Domains

- A relocation domain defines a set of members of an SSI cluster among which virtual machines can relocate freely
  
- Relocation domains can be defined for business or technical reasons
  
- Regardless of differences in the facilities of the individual members, a domain has a common architectural level
  - This is the maximal common subset of all the members' facilities
  
- Several default domains are automatically defined by CP
  - Single member domains for each member in the SSI
  - An SSI domain that will have the features and facilities common to all members
  
- Defining your own domains is useful in a 3+ member cluster
  - In a 1 or 2 member cluster, all possible domains are defined by default

# Relocation Domains



## Relocation Domains – How to define

- In the *SYSTEM CONFIG* on **PMAINT**'s 2CC disk
- Dynamically via a **DEFINE** command

```
88
89 RELOCATION_DOMAIN PADDNGTN MEMBER2 MEMBER3
90 RELOCATION_DOMAIN WINNIE MEMBER2 MEMBER4
91 RELOCATION_DOMAIN CORDUROY MEMBER3 MEMBER4
92
```

```
define relodomain paddngtn members member2 member3
Ready; T=0.01/0.01 15:12:08
define relodomain winnie members member2 member4
Ready; T=0.01/0.01 15:12:43
define relodomain corduroy members member3 member4
Ready; T=0.01/0.01 15:13:38
```

## Relocation Domains

- Virtual machines may be assigned to a domain in their directory entry
  - Default for single configuration virtual machines is the SSI domain
  - Default for multiconfiguration virtual machines is their singleton domain, which cannot be changed
  
- Virtual machines are assigned a virtual architecture level when they log on, according to what domain they are in
  
- They cannot use facilities or features not included in the domain even if the member they are on has access to those features
  - We call this “fencing”
  
- Examples of commands/instructions with “fenced” responses:
  - **Q CPUID** -the model number will always reflect the virtual architecture level, the processor number is set at logon and not affected by relocation or relocation domain changes
  - **Q CPLEVEL**
  - **STFLE**

## Relocation Domains

```
set vmrelocate * domain ssi
Running on member GDLRCTS2
Relocation enabled in Domain SSI
Ready;
q cpuid
CPUID = FF3B6D0520970000
Ready;
define relodomain winnie gdlrcts1 gdlrcts2
Ready;
set vmrelocate * domain winnie
Running on member GDLRCTS2
Relocation enabled in Domain WINNIE
Ready;
q cpuid
CPUID = FF3B6D0528170000
Ready;
```



***Planning and Creating a  
z/VM SSI Cluster***

## SSI Cluster Requirements

- Servers must be IBM System z10 or later (z/VM Version 6)
- Shared and non-shared DASD
  - 3390 volume required for the PDR
  - All volumes should be cabled to all members
    - Makes non-shared disks accessible to other members to fix configuration problems
- LPARs
  - 1-16 FICON CTC devices between LPARs
    - Provide direct ISFC links from each member to all other members
  - FICON channels to shared DASD
  - OSA access to the same LAN segments
  - FCP access to same storage area networks (SANs) with same storage access rights
- Shared system configuration file for all members
- Shared source directory containing user definitions for all members
- Capacity planning for each member of the SSI cluster
  - Ensure sufficient resources are available to contain shifting workload
    - Guests that will relocate
    - Guests that logon to different members

## SSI Cluster Restrictions

- Physical systems must be close enough to allow
  - FICON CTC connections
  - Shared DASD
  - Common network and disk fabric connections
  
- Installation to SCSI devices is not supported
  - Guests may use SCSI devices
  
- If using RACF, the database must reside on a fullpack 3390 volume
  
- Live Guest Relocation is only supported for Linux on System z guests

## SSI Cluster Setup – Suggested Practices

- Use the same real device numbers across LPARs to simplify cloning of z/VM systems
  - DASD volumes
  - Ranges for OSA and hipersockets subchannels connected to same network
  - Ranges for FCP subchannels connected to the same fabric
  
- Install no more than 2 members of an SSI cluster on the same server
- Maintain parallel volume layouts for each member (again, simplifies cloning)
- Allocate object directory (DRCT) extents only on the system residence volume for each member
- Do not place user data on the installation volumes
  - Simplifies release-to-release migration
  
- Keep member-specific data and SSI cluster data on separate volumes
  - Simplifies cloning and release-to-release migration
- Use a directory manager

## Summary

- An SSI cluster gives you
  - Workload balancing (move work to system resources)
  - Maintenance on your schedule (not the application owner)
  - Easier operation and management of multiple z/VM images
  
- Allow sufficient time to plan for an SSI cluster
  - Migration from current environment
  - Configuration
  - Sharing resources and data
  
- Plan for extra
  - CPU capacity
  - Memory
  - CTC connections

# Thanks!

*Contact Information:*

Richard Lewis  
IBM  
Washington Sys Ctr  
Gaithersburg, MD  
  
rflewis@us.ibm.com