IBM Systems & Technology Group

CSE For High Availability and System Management

Jay Brenneman
rjbrenn@us.ibm.com

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.

DB2*
DB2 Connect
DB2 Universal Database
e-business logo*
e-business on demand
HiperSockets
IBM*
IBM eServer
IBM logo*
IMS
Resource Link
S/390*
Tivoli*
Tivoli Storage Manager
TotalStorage*
WebSphere*
z/OS*
z/VM*
zSeries*

* Registered trademarks of IBM Corporation

The following are trademarks or registered trademarks of other companies.

Java and all Java-related trademarks and logos are trademarks of Sun Microsystems, Inc., in the United States and other countries.
Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.
Microsoft, Windows and Windows NT are registered trademarks of Microsoft Corporation.
UNIX is a registered trademark of The Open Group in the United States and other countries.
SET and Secure Electronic Transaction are trademarks owned by SET Secure Electronic Transaction LLC.

* 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.

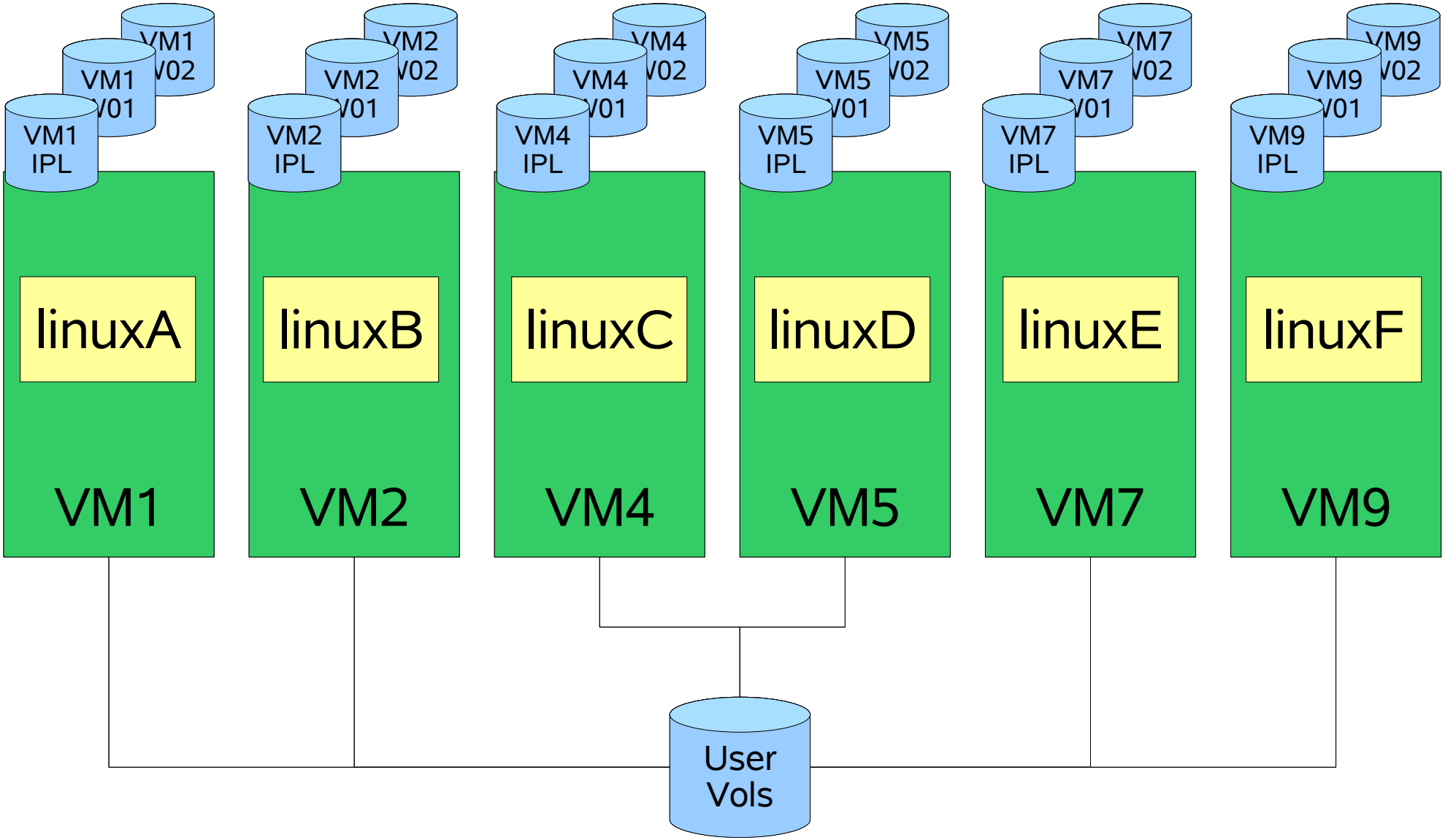
Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products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

Agenda – z/VM Cross System Extensions

- What does CSE do?
- What does CSE not do?
- How do I turn it on?
- How can I share the VM SysRes & Spool volumes?
- How does this get me closer to High Availability?

z/VM without CSE



CSE allows you to:

- Extend the mini disk access control semantics across VM nodes in Plex
  - Requires a common VM Source Directory across all nodes in Plex
- Share Spool files across VM nodes in Plex
  - Requires spxtape dump & restore to retrofit onto existing systems
- Extend the query and messaging tools across VM nodes in Plex

Ugly Bits: PVM and RSCS are not free...

- PVM is not included in the base cost of z/VM
  - Requires a Special Bid to get licensed on an IFL
  - needed for shared spooling
  - needed for cross system messaging, query
- RSCS is needed to support DirMaint communications if shared spool is not used
- DirMaint is required to support the single source directory
  - Or you have to figure out some other way to guarantee that the object directories are all identical, all the time

CSE does **NOT** allow you to:

- Share VM SysRes volumes for free
  - Each VM system must maintain its own object directory, warm start area, and checkpoint
- Get High Availability for free
  - Some infrastructure will have to be built or bought
- Share SFS pools across VM systems
  - Need TSAF, CS Collection, etc
- Virtual Reserve/Release across VM systems

Enabling CSE – The SYSTEM CONFIG file:

```
System_Identifier  2064 %01550 LTICVM1
System_Identifier  2094 %4299E LTICVM2
System_Identifier  2066 %10D05 LTICVM4
System_Identifier  2096 %168BE LTICVM5
System_Identifier  2064 %41550 LTICVM7
System_Identifier  2084 %4B52A LTICVM9

XLINK_System_Include Slot 1 LTICVM1
XLINK_System_Include Slot 2 LTICVM2
XLINK_System_Include Slot 3 LTICVM9
XLINK_System_Include Slot 4 LTICVM4
XLINK_System_Include Slot 5 LTICVM5
XLINK_System_Include Slot 6 LTICVM6
XLINK_System_Include Slot 7 LTICVM7
XLINK_System_Exclude LVL2VM

XLINK_Volume_Include VMP*

XSPOOL_SYSTEM Slot 1 LTICVM1 Share_Spool NO
XSPOOL_SYSTEM Slot 2 LTICVM2 Share_Spool NO
XSPOOL_SYSTEM Slot 3 LTICVM9 Share_Spool NO
XSPOOL_SYSTEM Slot 4 LTICVM4 Share_Spool NO
XSPOOL_SYSTEM Slot 5 LTICVM5 Share_Spool NO
XSPOOL_SYSTEM Slot 6 LTICVM6 Share_Spool NO
XSPOOL_SYSTEM Slot 7 LTICVM7 Share_Spool NO

XSPOOL_XLIST_OUTPUT  RSCSDNS VMSERVS VMSERVU PVM OPERATOR DTCVSW1 DTCVSW2
XSPOOL_XLIST_INPUT   RSCSDNS VMSERVS VMSERVU PVM OPERATOR DTCVSW1 DTCVSW2
```

Shared RACF database

- Change the RACF database from minis to dedicated volumes that support Hardware Reserve/Release
  - DDR from the minis to full volumes on VM1
  - Change the directory entry and recycle RACF on VM1
  - Add VM2 missing entries to the database
  - Change the directory entry and recycle RACF on VM2
- That's pretty much it – do the same for the rest of the VM systems

RSCS configuration

- RSCS is used for DirMaint message passing if you're not using Shared Spool

```
'CP ATTACH DB45 * DB4 '          /* Link to LTICVM2 */
'CP ATTACH C315 * C31 '          /* Link to LTICVM4 */
'CP ATTACH C215 * C21 '          /* Link to LTICVM5 */
'CP ATTACH D055 * D05 '          /* Link to LTICVM7 */
'CP ATTACH DA45 * DA4 '          /* Link to LTICVM9 */

LINKDEFINE LTICVM2 AST TYPE NJE LINE DB4
LINKDEFINE LTICVM4 AST TYPE NJE LINE C31
LINKDEFINE LTICVM5 AST TYPE NJE LINE C21
LINKDEFINE LTICVM7 AST TYPE NJE LINE D05
LINKDEFINE LTICVM9 AST TYPE NJE LINE DA4

'RSCS START LTICVM2'
'RSCS START LTICVM4'
'RSCS START LTICVM5'
'RSCS START LTICVM7'
'RSCS START LTICVM9'
```

PVM configuration

- Enables system to system messaging for
  - Indicate
  - Q Names
  - SMSG
- Enables Shared Spool

```
'CP ATT DB44 * DB4'   /* LTICVM2 */
'CP ATT C314 * C31'   /* LTICVM4 */
'CP ATT C214 * C21'   /* LTICVM5 */
'CP ATT D054 * D05'   /* LTICVM7 */
'CP ATT DA44 * DA4'   /* LTICVM9 */

LOCAL LTICVM1
LINK DB4 LTICVM2 CTCA
LINK C31 LTICVM4 CTCA
LINK C21 LTICVM5 CTCA
LINK D05 LTICVM7 CTCA
LINK DA4 LTICVM9 CTCA

START LINE DB4
START LINE C31
START LINE C21
START LINE D05
START LINE DA4
START CSECOM LTICVM2
START CSECOM LTICVM4
START CSECOM LTICVM5
START CSECOM LTICVM7
START CSECOM LTICVM9
```

Format a volume to support XLINK

- Attach the volume to your id
- XLINK FORMAT <vaddr> <volid>
  - XLINK command is split: part built in and part module
    - Module lives on maint's 193
  - Defaults to adding CSE tracks in CYL 0
    - 1 Cyl for mod 3, 9 cyl for mod 9 up to mod 54
- IPL to pick up SYSTEM CONFIG changes if you haven't already
- Make sure “q n” shows you everyone logged on everywhere
  - indicates CSE messaging is up
- Attach the volume to SYSTEM on all VMs
- XLINK CHECK <volid>
  - Volume <volid> is controlled by CSE LINK.

Dirmaint Configuration Overview

- Dirmaint will run on one node in the plex
- DirmSats will run on all other nodes in the plex
- Directory changes are made everywhere.
  - This can take some stern user re-education
- Dirmaint must be equal to or higher than the Dirmsats in code level

Directory Stuff

- Merging the directory is the hardest part of implementing CSE
- On VM1:
  - Add the system affinity information for all VM systems to the DIRECTORY control statement
  - Add SYSAFFIN statements to all guests which will have differences between systems
  - Enable dirmaint & make sure the lock disk (15D) is defined on XLINK controlled DASD

```
DIRECTORY 0123 3390 VM1IPL *01550-2064 LTICVM1
DIRECTORY 0123 3390 VM4IPL *10D05-2066 LTICVM4
DIRECTORY 0123 3390 VM7IPL *41550-2064 LTICVM7
DIRECTORY 0123 3390 VM2IPL *4299E-2094 LTICVM2
DIRECTORY 5502 3390 VM9CDS *4B52A-2084 LTICVM9
DIRECTORY 5623 3390 VM5CDS *168BE-2096 LTICVM5
```

A SYSAFFIN'ed Directory entry

```
USER TCPIP TCPIP 64M 128M ABCG
   INCLUDE TCPCMSU
   IUCV ALLOW
   IUCV ANY PRIORITY
   IUCV *CCS PRIORITY MSGLIMIT 255
   IUCV *VSWITCH MSGLIMIT 65535
   OPTION QUICKDSP SVMSTAT MAXCONN 1024 DIAG98 APPLMON
   SHARE RELATIVE 3000
   LINK TCPMAINT 0591 0591 RR
   LINK TCPMAINT 0592 0592 RR
   LINK TCPMAINT 0198 0198 RR
SYSAFFIN LTICVM1 LTICVM4
   LINK 5VMTCP10 0491 0491 RR
   LINK 5VMTCP10 0492 0492 RR
SYSAFFIN LTICVM2 LTICVM5 LTICVM7 LTICVM9
   LINK 5VMTCP20 0491 0491 RR
   LINK 5VMTCP20 0492 0492 RR
SYSAFFIN LTICVM1 LTICVM4
   MDISK 0191 3390 2953 5 +VMRES MR RTCPIP WTCPIP MTCPIP
SYSAFFIN LTICVM2 LTICVM7
   MDISK 0191 3390 3125 5 +VMRES MR RTCPIP WTCPIP MTCPIP
SYSAFFIN LTICVM5
   MDISK 9191 3390 3125 5 +VMRES RR RTCPIP WTCPIP MTCPIP
   MDISK 0191 3390 0072 5 VM5CDS MR RTCPIP WTCPIP MTCPIP
SYSAFFIN LTICVM9
   MDISK 9191 3390 3125 5 +VMRES RR RTCPIP WTCPIP MTCPIP
   MDISK 0191 3390 0072 5 VM9CDS MR RTCPIP WTCPIP MTCPIP
```

More Directory Stuff

- Once VM1 has SYSAFFIN statements for all systems which will be part of the plex:
  - Using dirmsat as a template, create a new dirmsat user to run on each of the other VM systems: dirmsat2, dirmsat4, dirmsat5, dirmsat7, dirmsat9
  - Create a DVHPROFA DIRMSATx on Dirmaint's C disk for each new dirmsat user
  - Add RACF privileges for the dirmsats if needed
  - Run DIRM USER WITHPASS to consolidate the dirmaint files into a monolithic directory file
  - send USER WITHPASS to each of the other VM systems

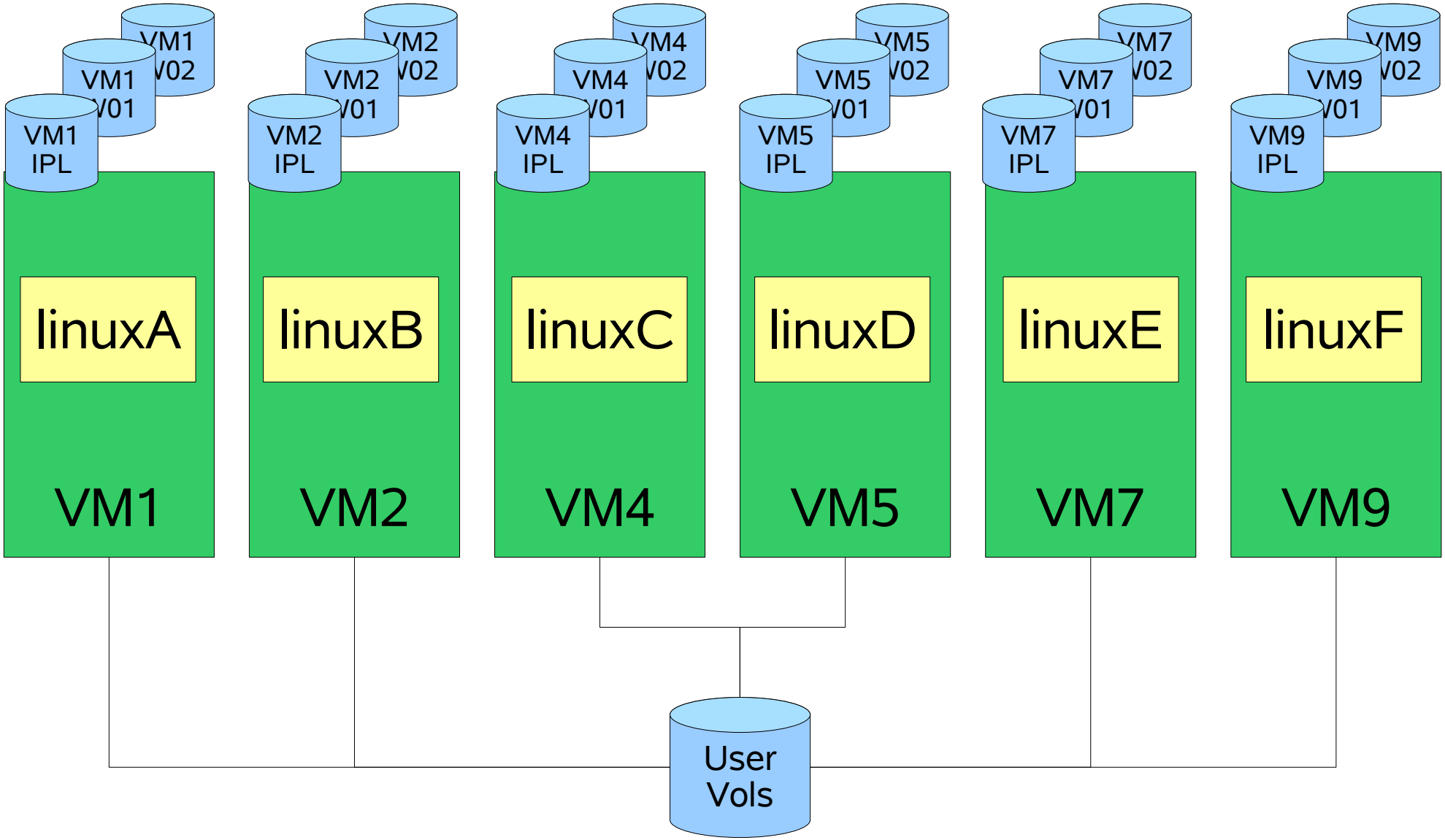
Yet More Directory Stuff

- On the other VM systems:
  - Rebuild the directory with directxa using the USER WITHPASS file from VM1
  - Xautolog this system's dirmsat
  - Add the RACF privileges you forgot about when you created it
- Add FROM= TO= and SATELLITE_SERVER= statements to Dirmaint Config

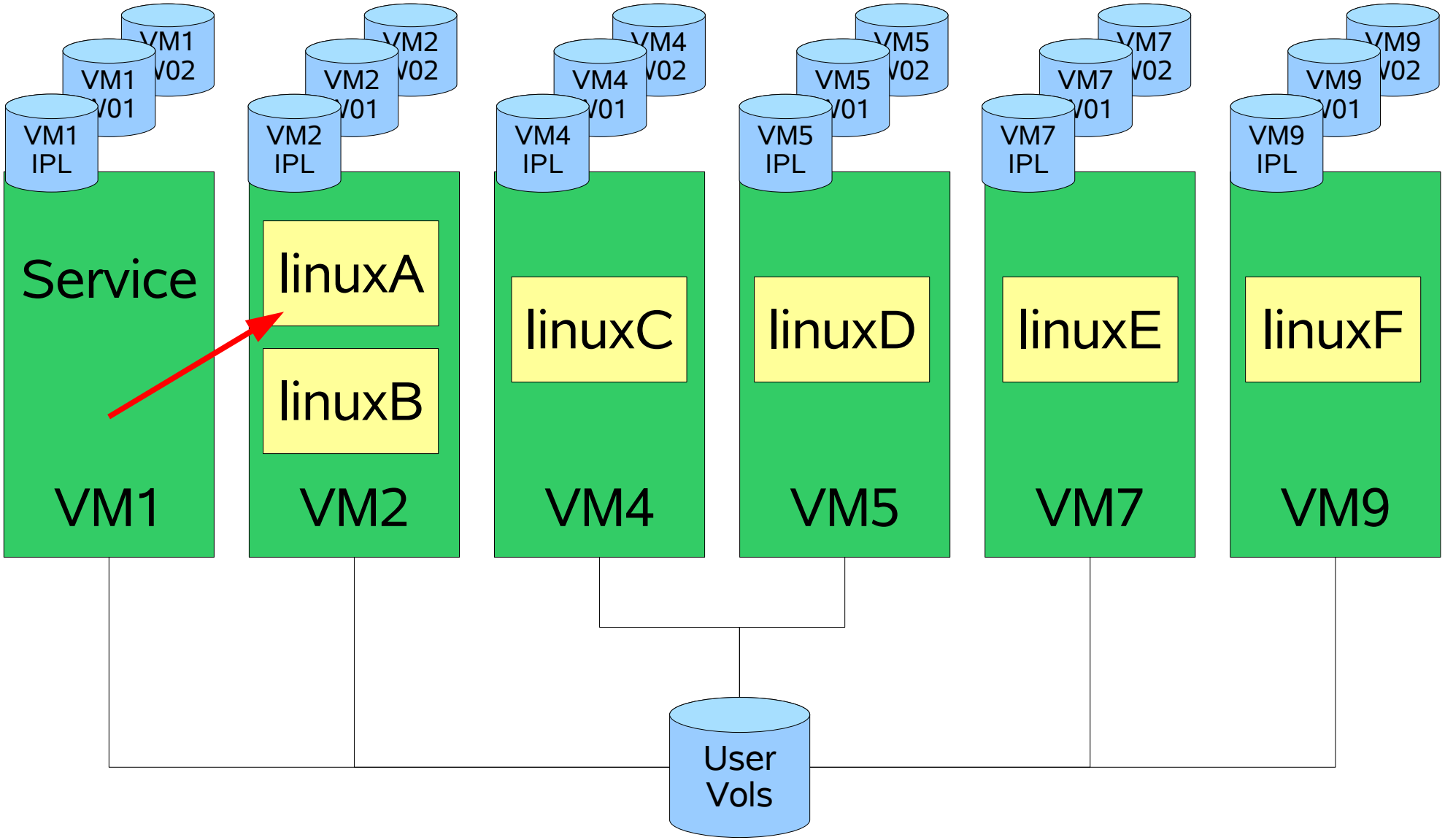
```
SATELLITE_SERVER= DIRMSAT9 LTICVM9
SATELLITE_SERVER= DIRMSAT5 LTICVM5
SATELLITE_SERVER= DIRMSAT4 LTICVM4
SATELLITE_SERVER= DIRMSAT2 LTICVM2
SATELLITE_SERVER= DIRMSAT1 LTICVM1

FROM= LTICVM1  TO= LTICVM2  S= RSCS T=LTICVM2
FROM= LTICVM2  TO= LTICVM1  S= RSCS T=LTICVM1
... And so on
```

- Force and restart dirmaint on VM1 and dirmsats everywhere else
- Enjoy the utopia of CSE enabled VM.

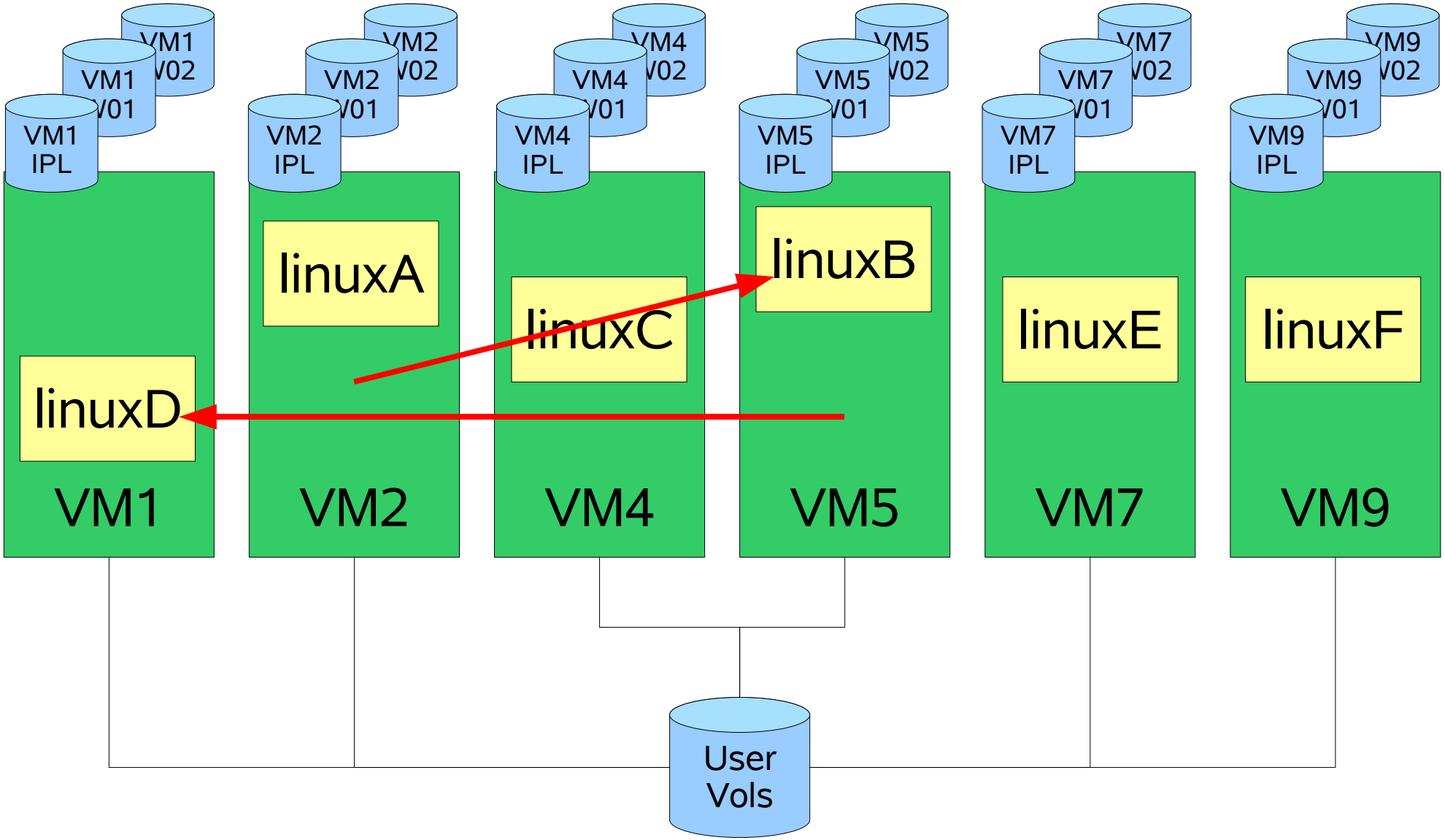
z/VM with CSE



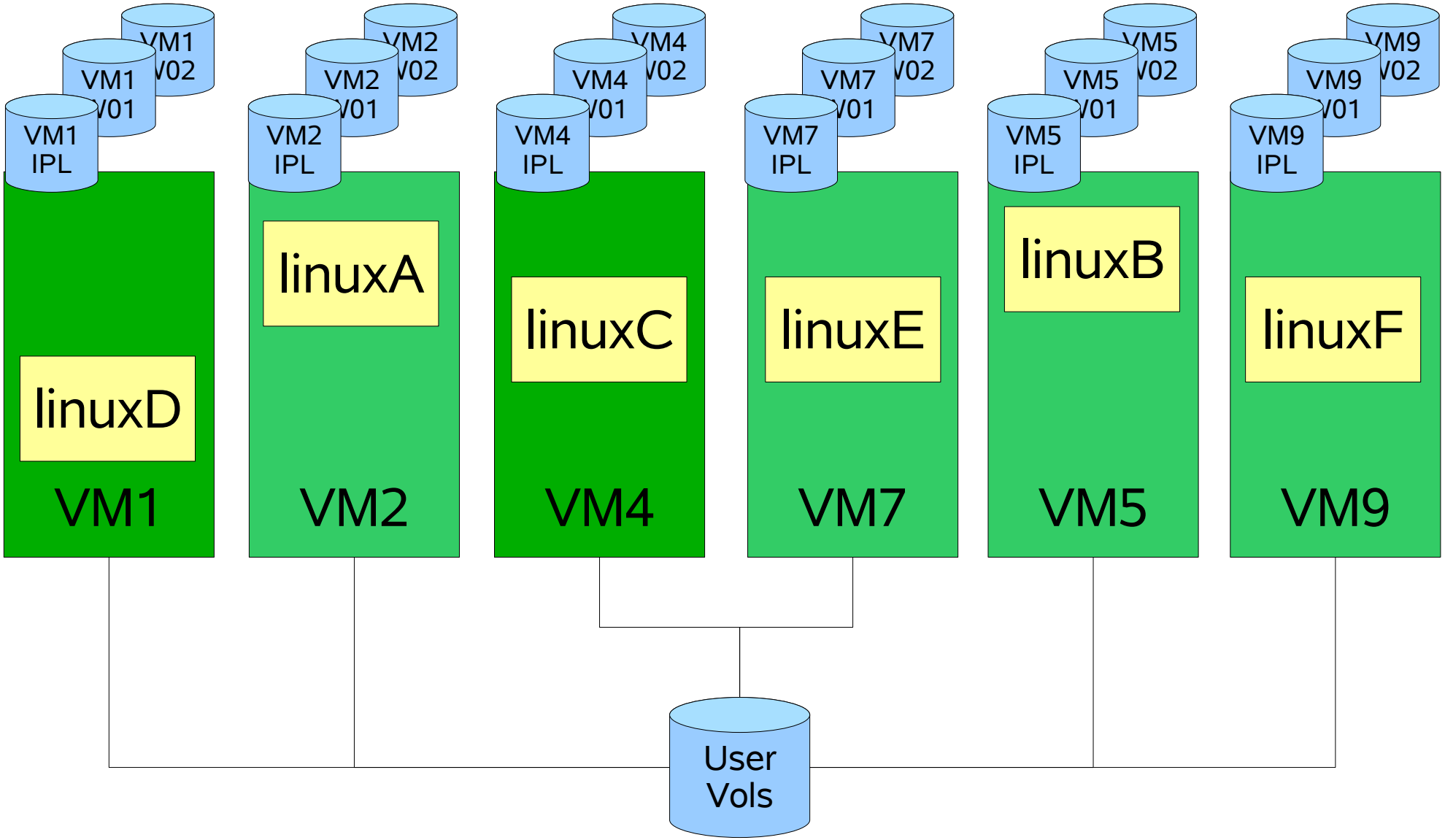
z/VM with CSE



z/VM with CSE

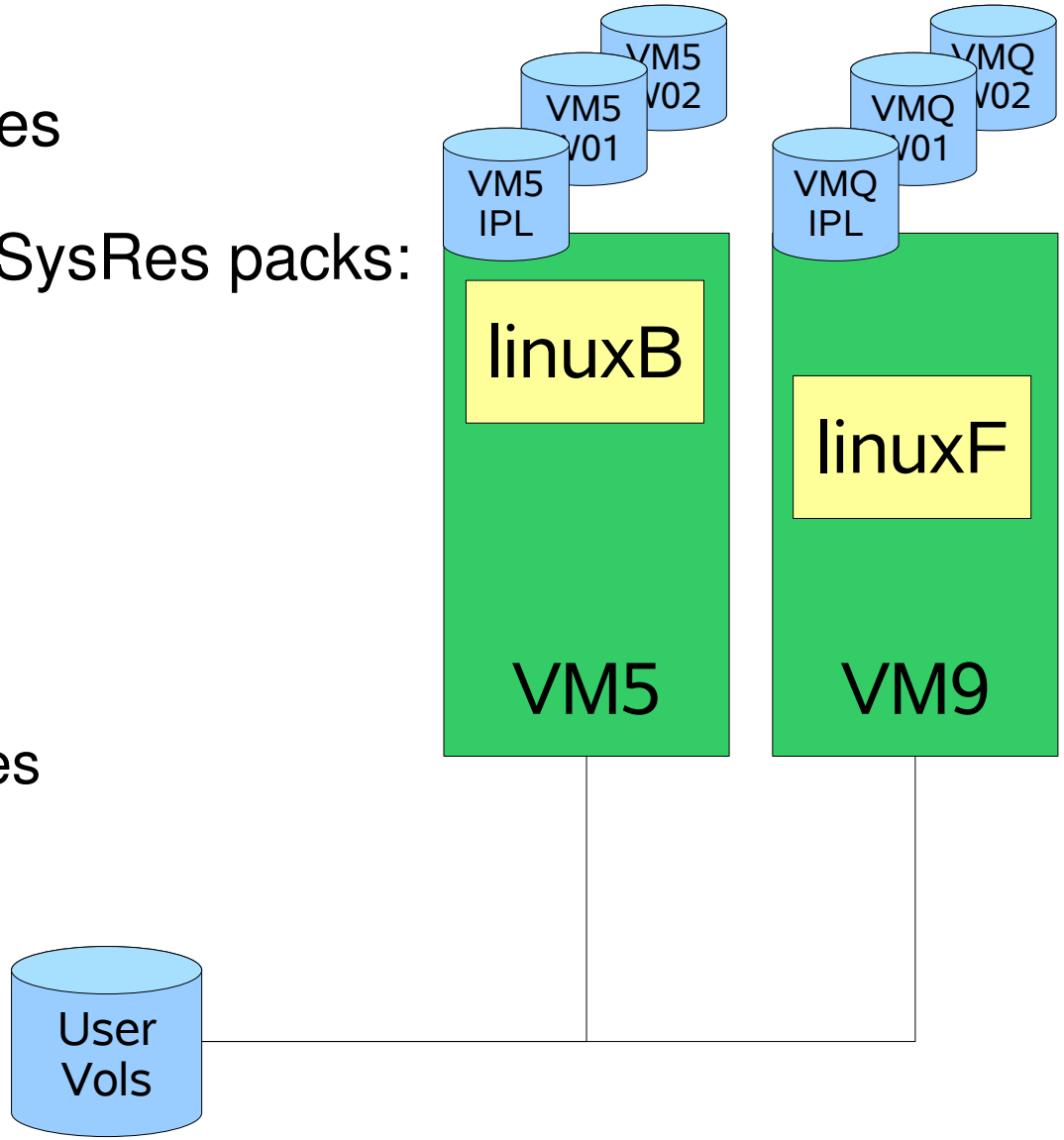


z/VM Shared SysRes



z/VM Shared SysRes

- Rebuild VM9 on VMQ volumes
- Going to move these off the SysRes packs:
  - Checkpoint
  - Warmstart
  - Directory
  - RW minis for service machines



z/VM Shared SysRes – move the checkpoint

- MAINT's CF1 is going to be shared, so:
  - in SYSTEM CONFIG on CF1

```
Imbed -SYSTEM- SYSRES
Imbed -SYSTEM- CPOWNED
```

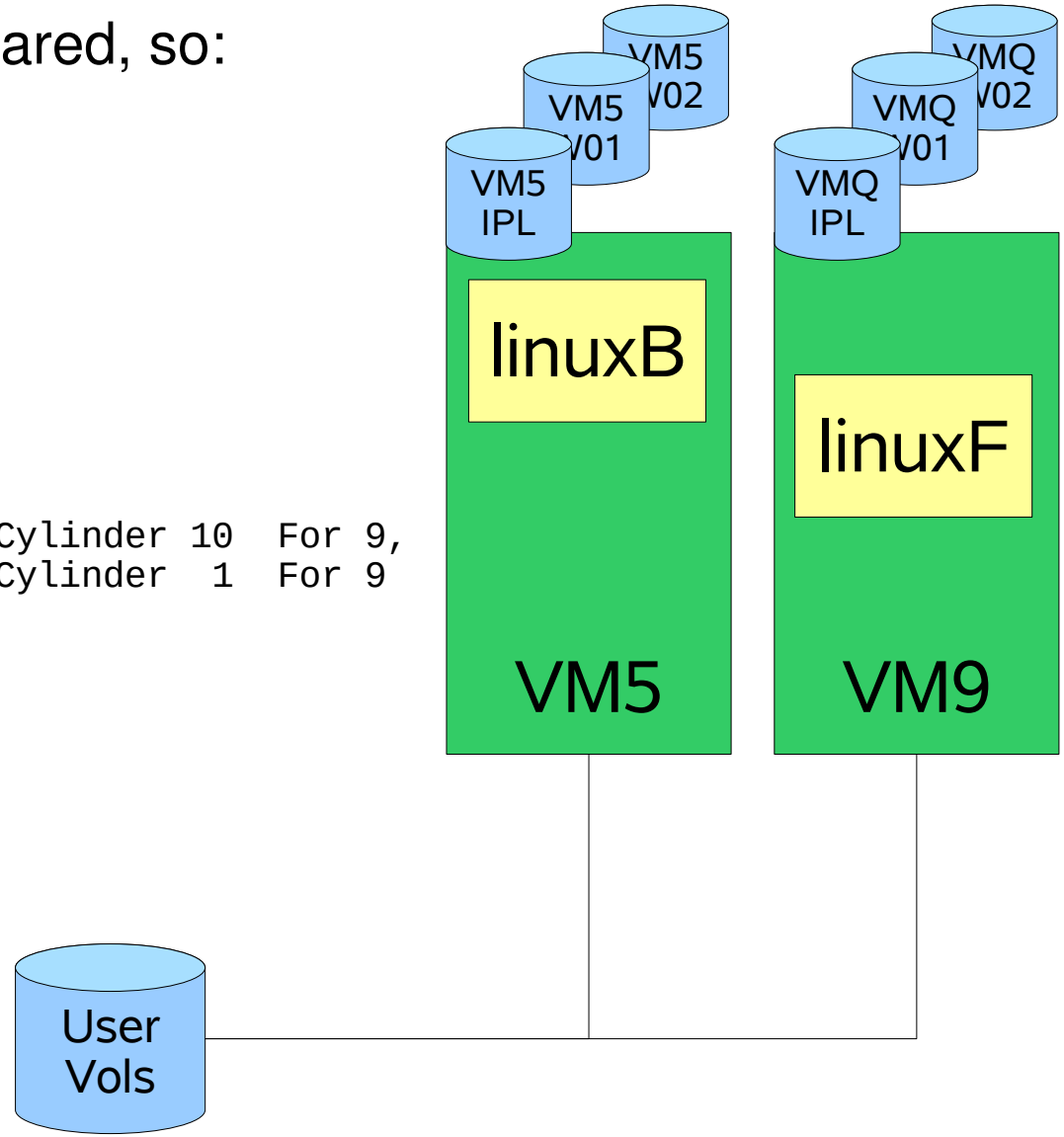
  - in VM9 SYSRES on CF1

```
System_Residence,
    Warmstart   Volid  VM9CDS From Cylinder 10  For 9,
    Checkpoint  Volid  VM9CDS From Cylinder  1  For 9
```

  - in VM9 CPOWNED on CF1

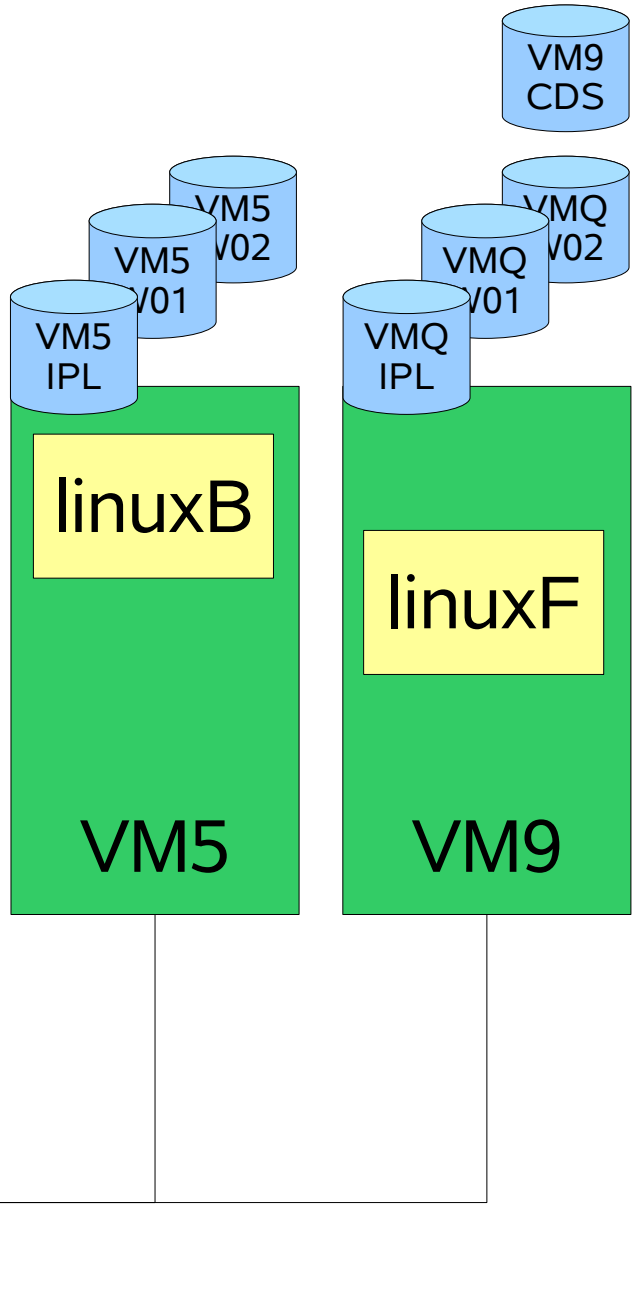
```
CP_Owned    Slot    1  &SYSRES
CP_Owned    Slot    2  VM9CDS
```

    - Don't include multiple Directory bearing volumes for other VM systems in CPOWNED



z/VM Shared SysRes

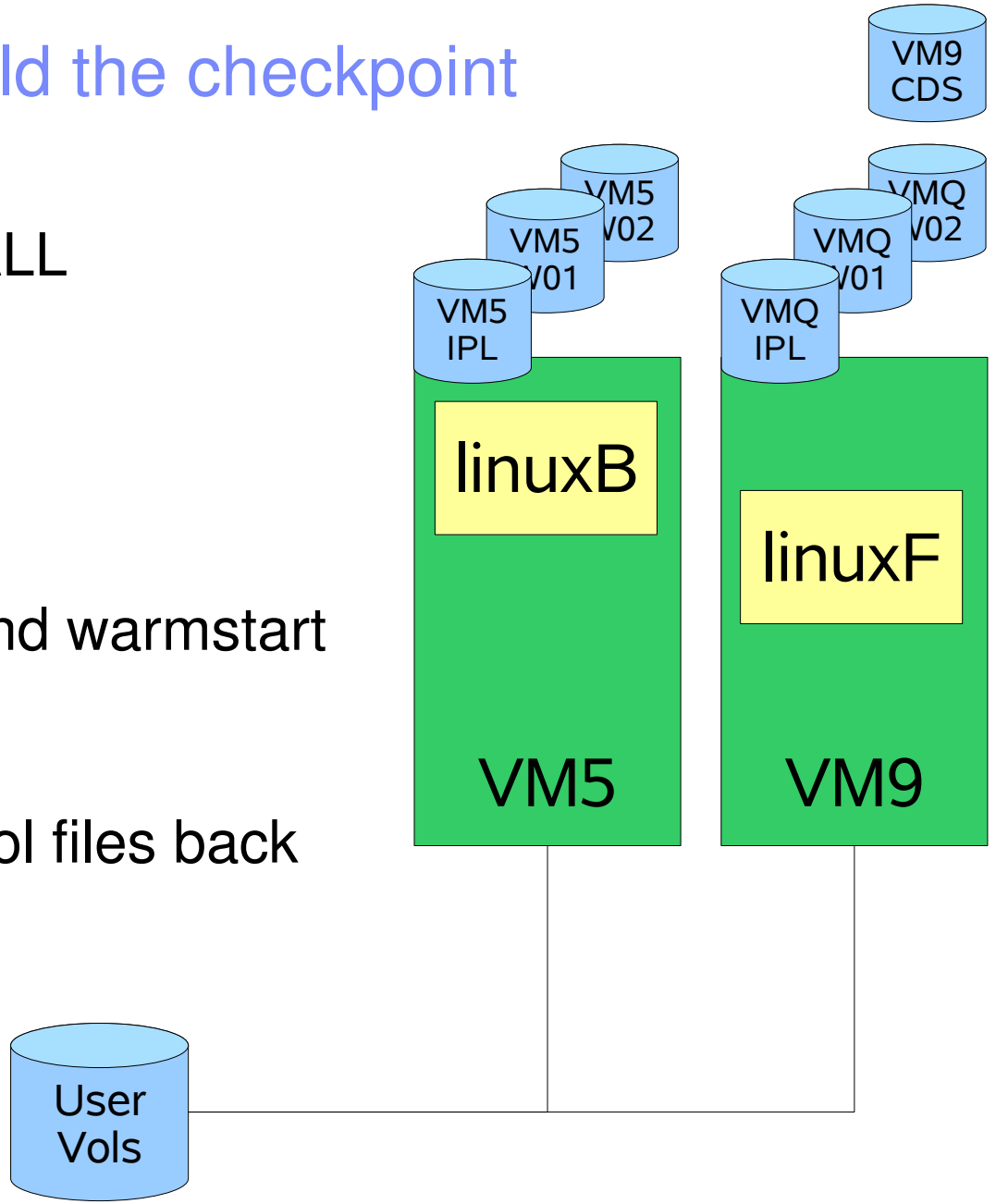
- CP format VM9CDS
  - VM9's Checkpoint, Directory, and Service Machine minis
  - Allocate checkpoint and warmstart as PERM
  - Allocate directory as DRCT
  - Allocate the space for Minis as PERM




```
CYLINDER ALLOCATION CURRENTLY IS AS FOLLOWS:
TYPE     START      END        TOTAL
----     -----      ---        -----
PERM     0          18         19
DRCT     19         58         40
PERM     59         3338       3280
```

z/VM Shared SysRes – rebuild the checkpoint

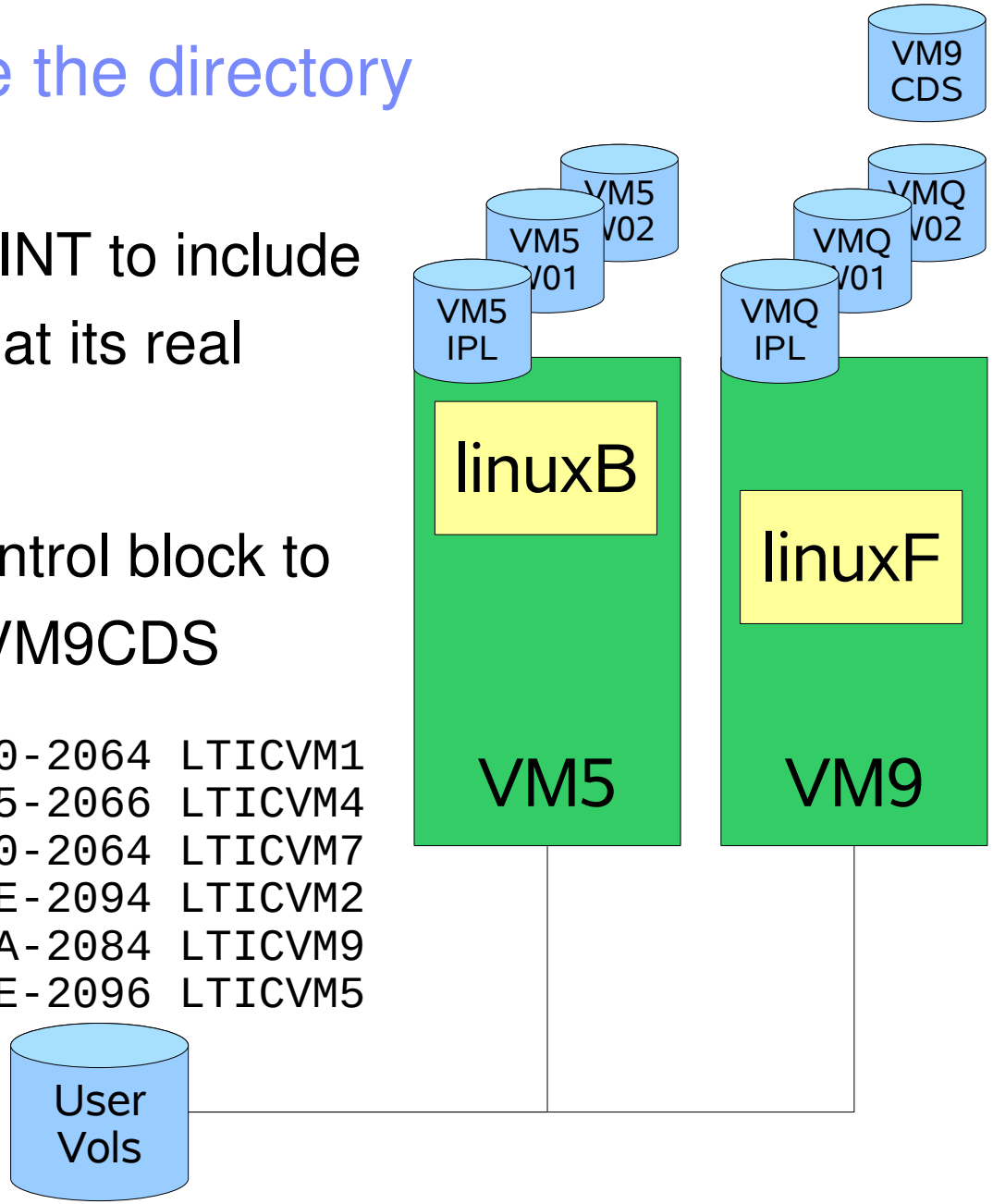
- SPXTAPE DUMP SPOOL ALL
- IPL CLEAN
  - Trashes all spool space
  - Rebuilds the checkpoint and warmstart data
- SPXTAPE LOAD to get spool files back



z/VM Shared SysRes – move the directory

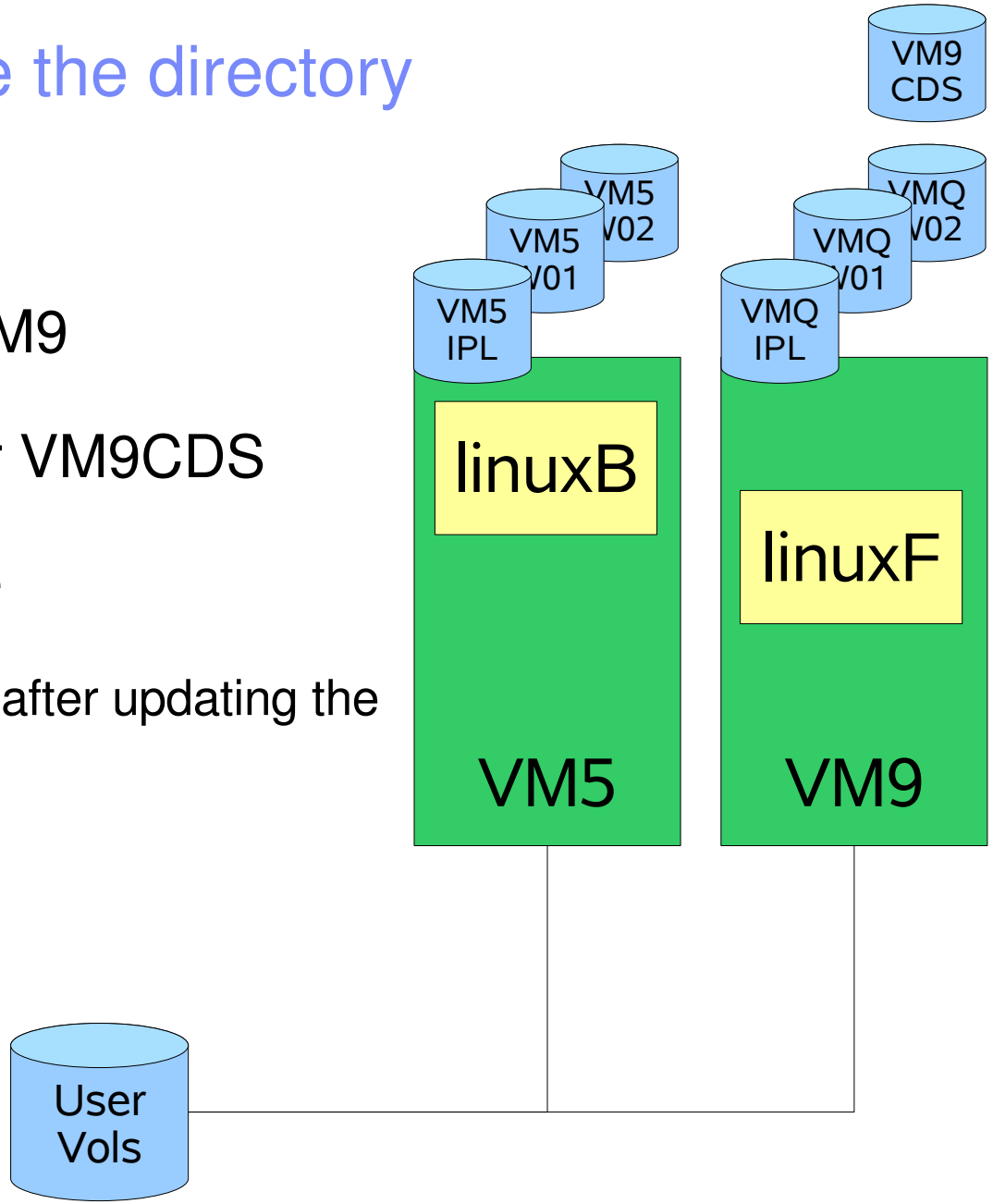
- Update the Directory for MAINT to include a fullpack mini for VM9CDS at its real device address
- Update the DIRECTORY control block to point to directory space on VM9CDS

```
DIRECTORY 0123 3390 VM1IPL *01550-2064 LTICVM1
DIRECTORY 0123 3390 VM4IPL *10D05-2066 LTICVM4
DIRECTORY 0123 3390 VM7IPL *41550-2064 LTICVM7
DIRECTORY 0123 3390 VM2IPL *4299E-2094 LTICVM2
DIRECTORY 5502 3390 VM9CDS *4B52A-2084 LTICVM9
DIRECTORY 5623 3390 VM5CDS *168BE-2096 LTICVM5
```



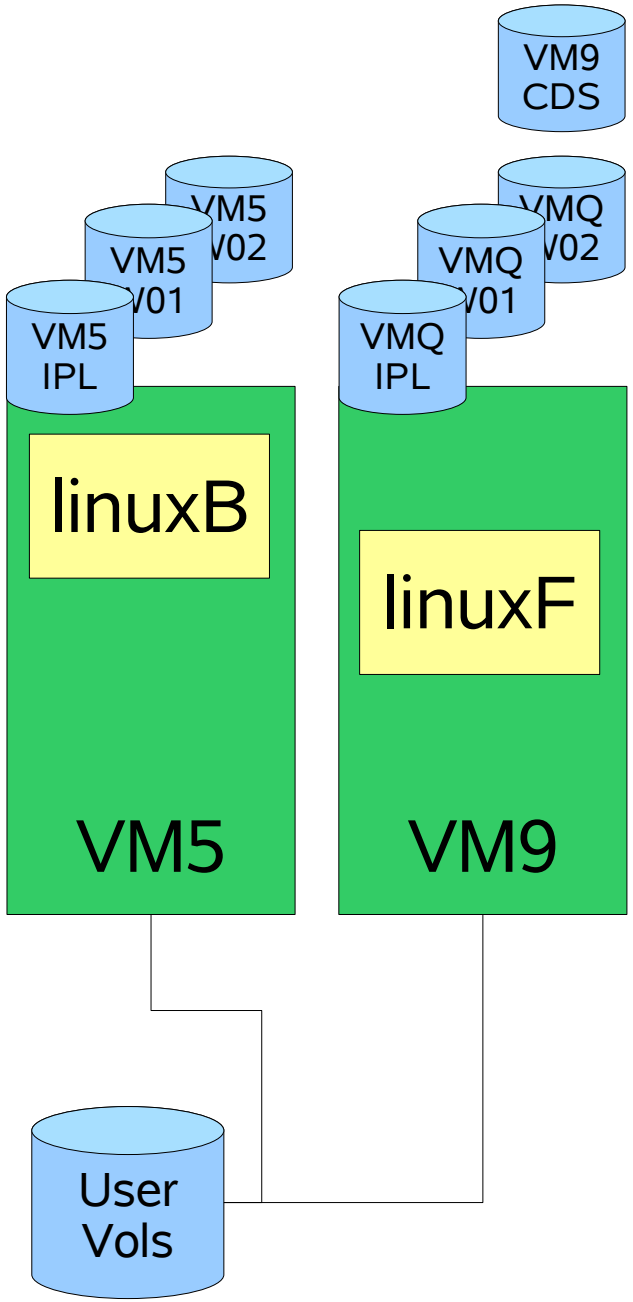
z/VM Shared SysRes – move the directory

- Get USER WITHPASS to VM9
- Link the fullpack minidisk for VM9CDS
- Run directxa with the flat file
  - Don't update the allocation map after updating the directory



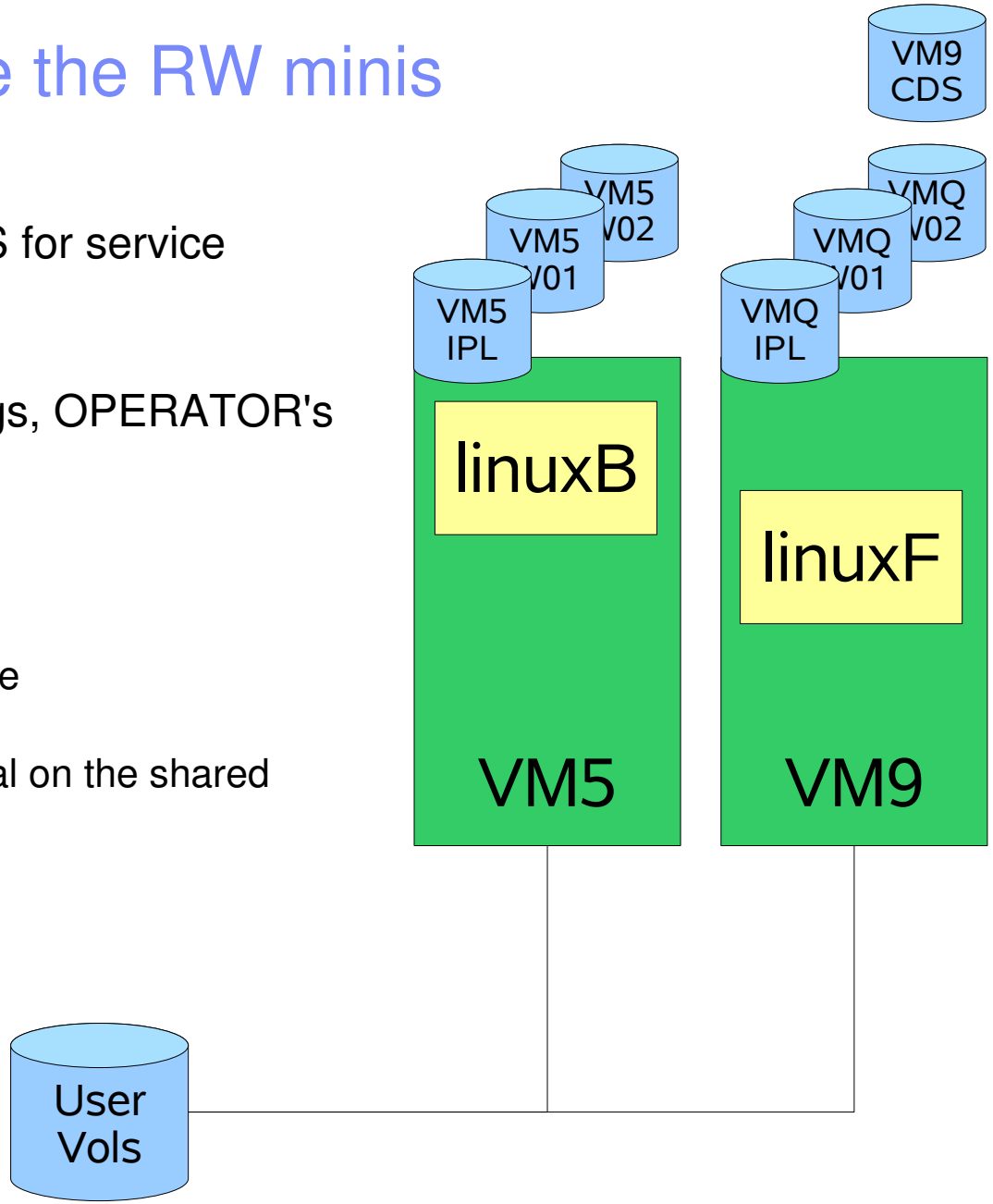
z/VM Shared SysRes – xlink format

- Backup everything here before continuing if you don't have another VM system to fix this one
- Is the VMQ SysRes a mod-9? If yes – STOP
  - You have to move everything up by 10 Cylinders first
- 1 Run XLINK FORMAT to set up the CSE track on all the VMQ SysRes volumes
- 2 Add `XLINK_Volume_Include VMQ*` to SYSTEM CONFIG
  - ANY volume labeled VMQ* without a CSE track will not be brought online, at all.
- Alter the allocation map on VMQIPL to remove the directory space
- IPL to check the changes – Should see the message that the Directory on VM9CDS is on line.



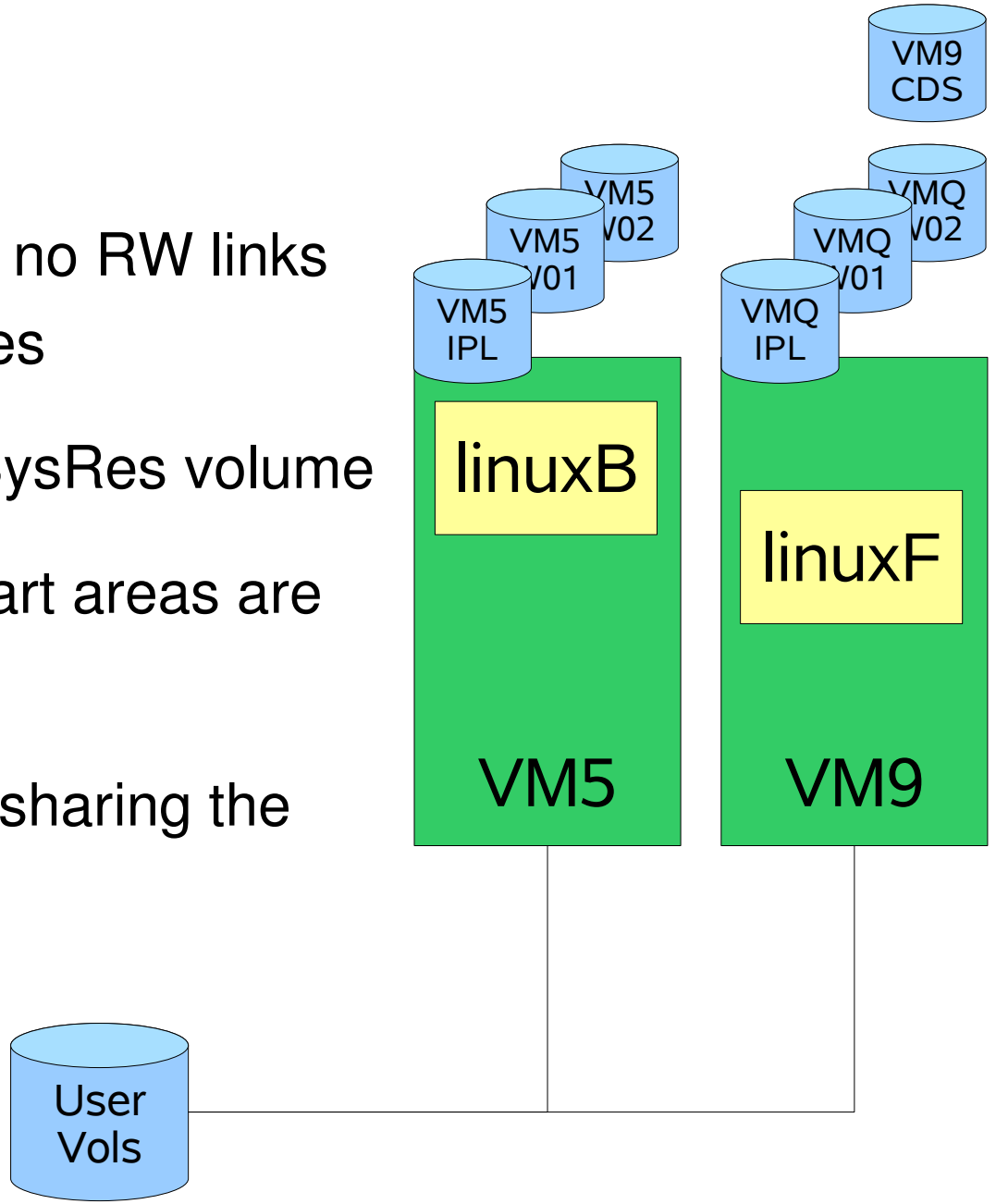
z/VM Shared SysRes – move the RW minis

- Start allocating minidisks on VM9CDS for service machines that need R/W space.
- Things like: SFS pool, RACF audit logs, OPERATOR's 191
  - Allocate them as 9xxx
  - Copy from the original to the 9xxx device
  - Flip the addresses so 9xxx is the original on the shared SysRes volume, linked RO
  - Useful for service later



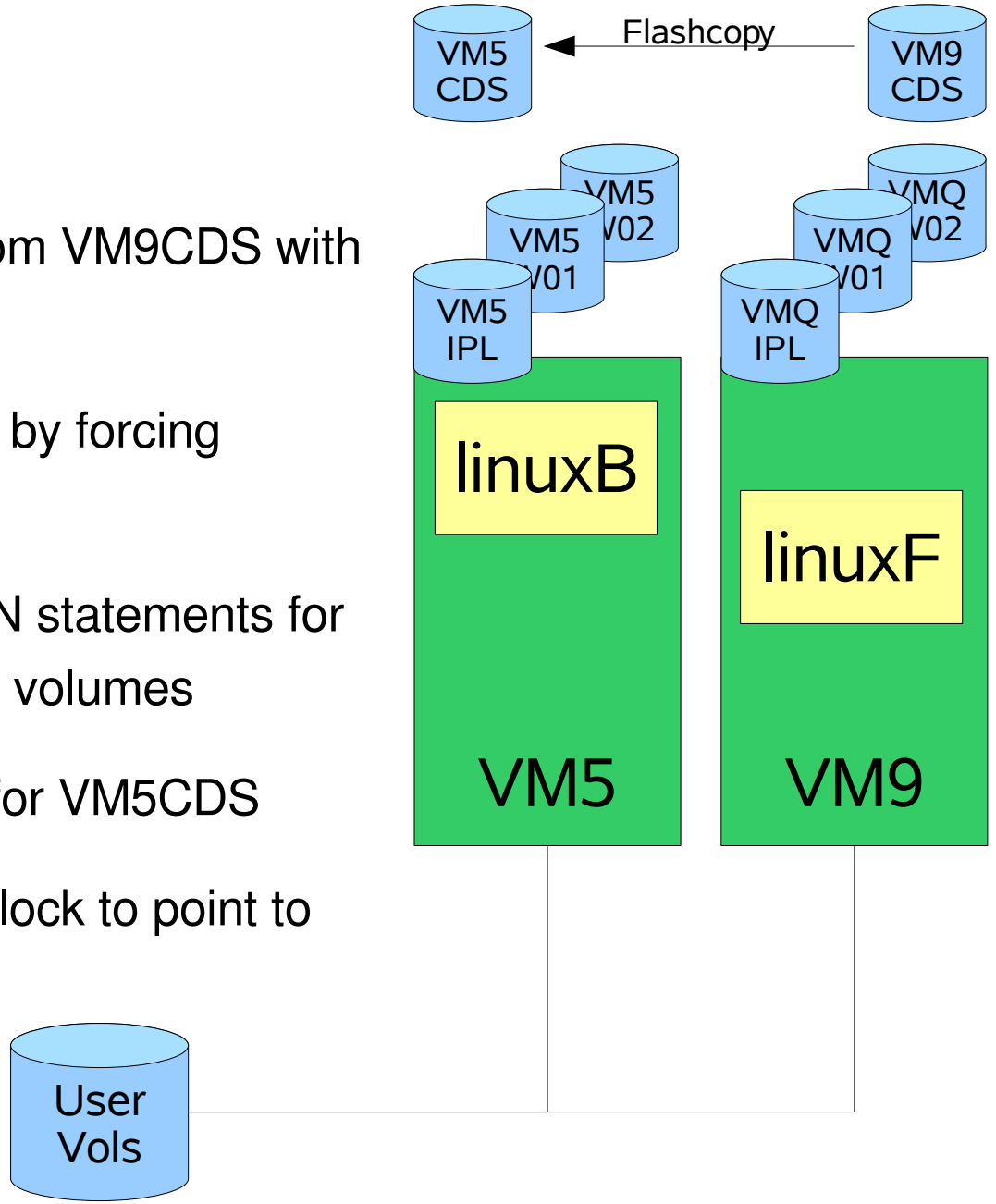
z/VM Shared SysRes

- At this point there should be no RW links to any of the SysRes volumes
- The directory is not on the SysRes volume
- The checkpoint and warmstart areas are not on the SysRes volume
- Now to add VM5 so that it's sharing the VMQ volumes

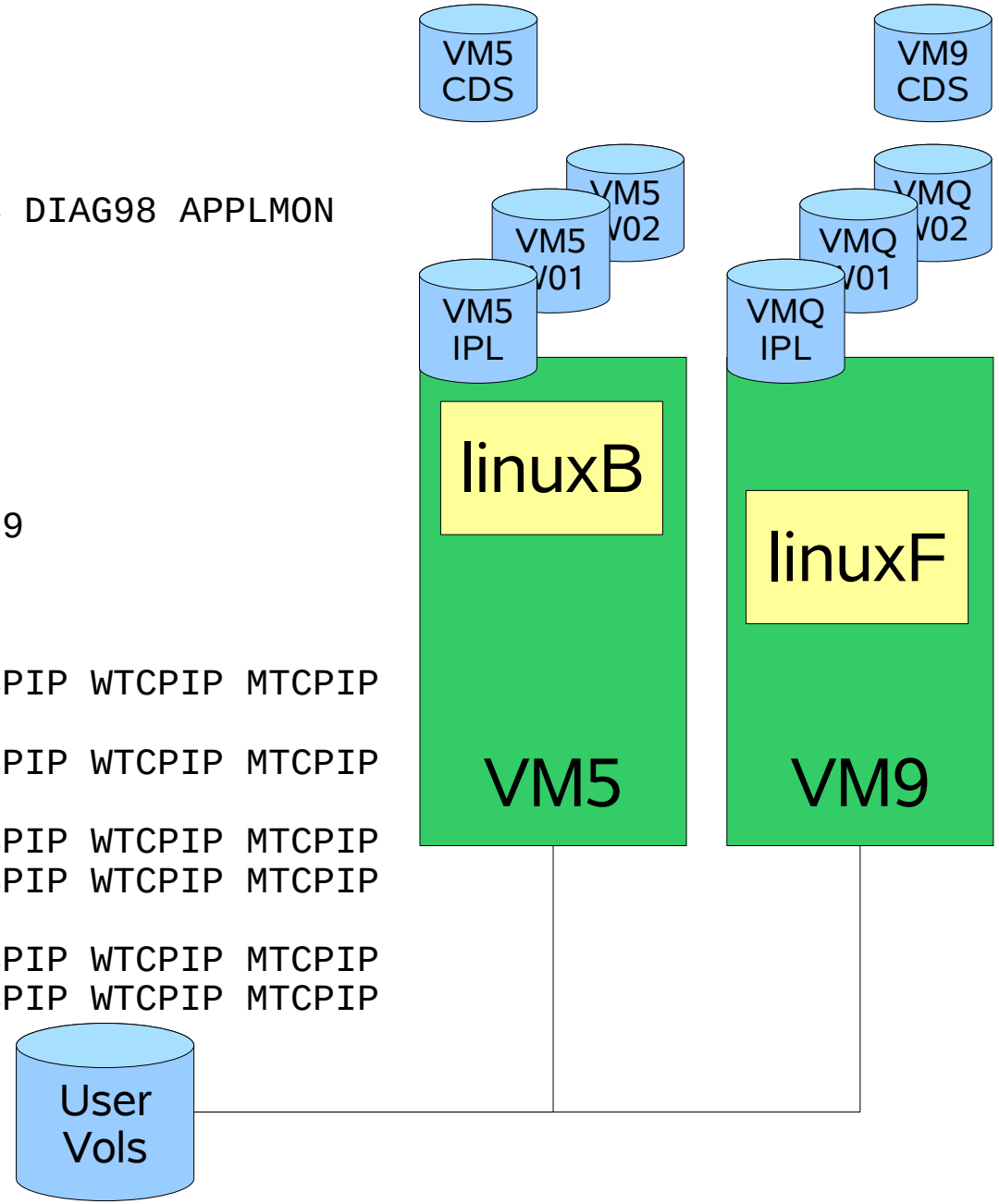


z/VM Shared SysRes

- Create a new VM5CDS volume from VM9CDS with DDR or FlashCopy
- Turn off directory updates on VM5 by forcing DIRMSAT5
- From VM9 - change the SYSAFFIN statements for VM5's system IDs to point to VMQ volumes
- Add a fullpack minidisk to MAINT for VM5CDS
- Update the DIRECTORY control block to point to the new volume



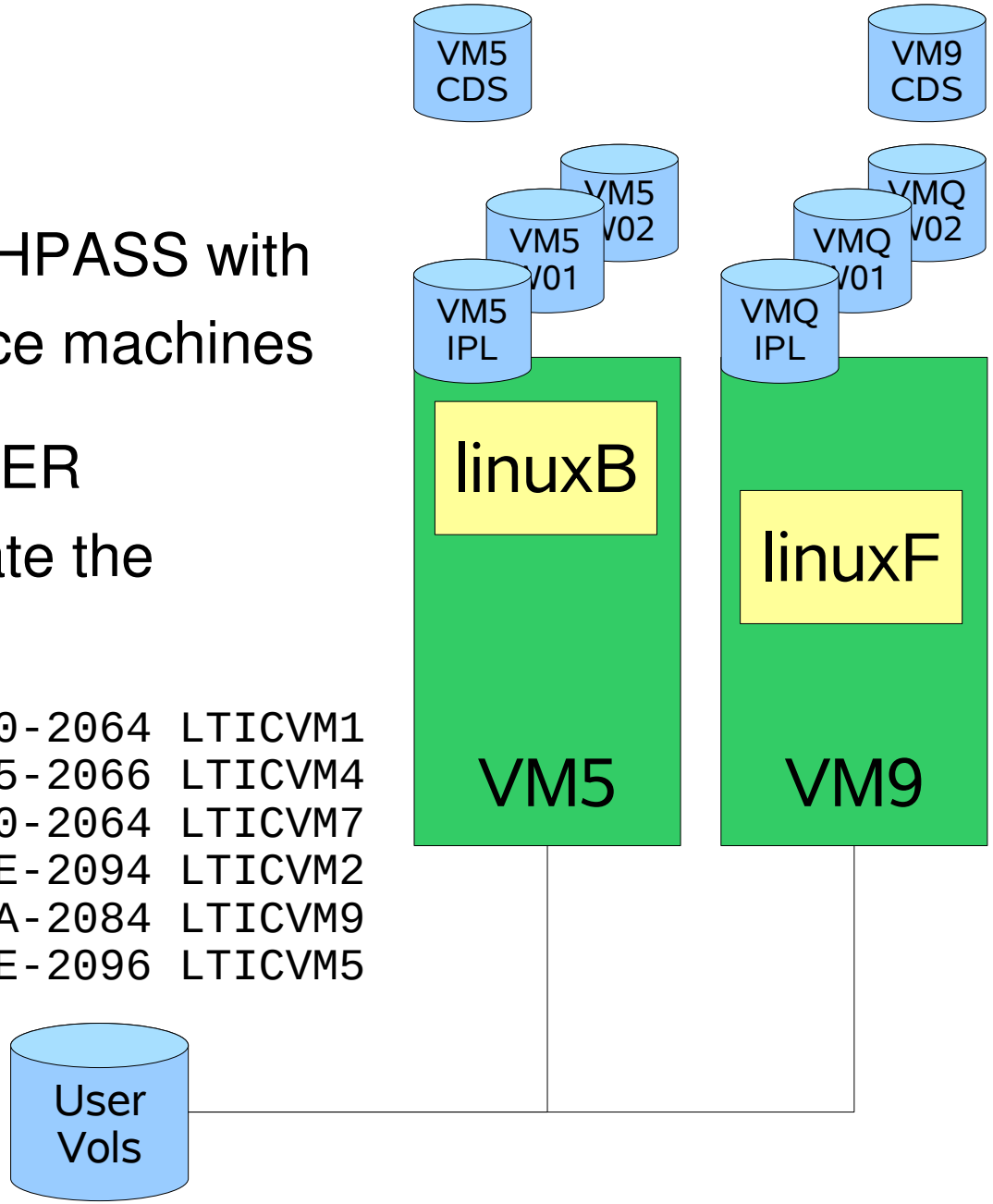
```
USER TCPIP TCPIP 64M 128M ABCG
   INCLUDE TCPCMSU
   IUCV ALLOW
   IUCV ANY PRIORITY
   IUCV *CCS PRIORITY MSGLIMIT 255
   IUCV *VSWITCH MSGLIMIT 65535
   OPTION QUICKDSP SVMSTAT MAXCONN 1024 DIAG98 APPLMON
   SHARE RELATIVE 3000
   LINK TCPMAINT 0591 0591 RR
   LINK TCPMAINT 0592 0592 RR
   LINK TCPMAINT 0198 0198 RR
SYSAFFIN LTICVM1 LTICVM4
   LINK 5VMTCP10 0491 0491 RR
   LINK 5VMTCP10 0492 0492 RR
SYSAFFIN LTICVM2 LTICVM5 LTICVM7 LTICVM9
   LINK 5VMTCP20 0491 0491 RR
   LINK 5VMTCP20 0492 0492 RR
SYSAFFIN LTICVM1 LTICVM4
   MDISK 0191 3390 2953 5 +VMRES MR RTCPIP WTCPIP MTCPIP
SYSAFFIN LTICVM2 LTICVM7
   MDISK 0191 3390 3125 5 +VMRES MR RTCPIP WTCPIP MTCPIP
SYSAFFIN LTICVM5
   MDISK 9191 3390 3125 5 +VMRES RR RTCPIP WTCPIP MTCPIP
   MDISK 0191 3390 0072 5 VM5CDS MR RTCPIP WTCPIP MTCPIP
SYSAFFIN LTICVM9
   MDISK 9191 3390 3125 5 +VMRES RR RTCPIP WTCPIP MTCPIP
   MDISK 0191 3390 0072 5 VM9CDS MR RTCPIP WTCPIP MTCPIP
```



z/VM Shared SysRes

- Generate a new USER WITHPASS with the new SYSAFFINed service machines
- Run directxa against the USER WITHPASS on VM5 to update the directory on VM5CDS

```
DIRECTORY 0123 3390 VM1IPL *01550-2064 LTICVM1
DIRECTORY 0123 3390 VM4IPL *10D05-2066 LTICVM4
DIRECTORY 0123 3390 VM7IPL *41550-2064 LTICVM7
DIRECTORY 0123 3390 VM2IPL *4299E-2094 LTICVM2
DIRECTORY 5502 3390 VM9CDS *4B52A-2084 LTICVM9
DIRECTORY 5623 3390 VM5CDS *168BE-2096 LTICVM5
```



z/VM Shared SysRes

- On VM9: create new config files for VM5 and add them to MAINT's CF1

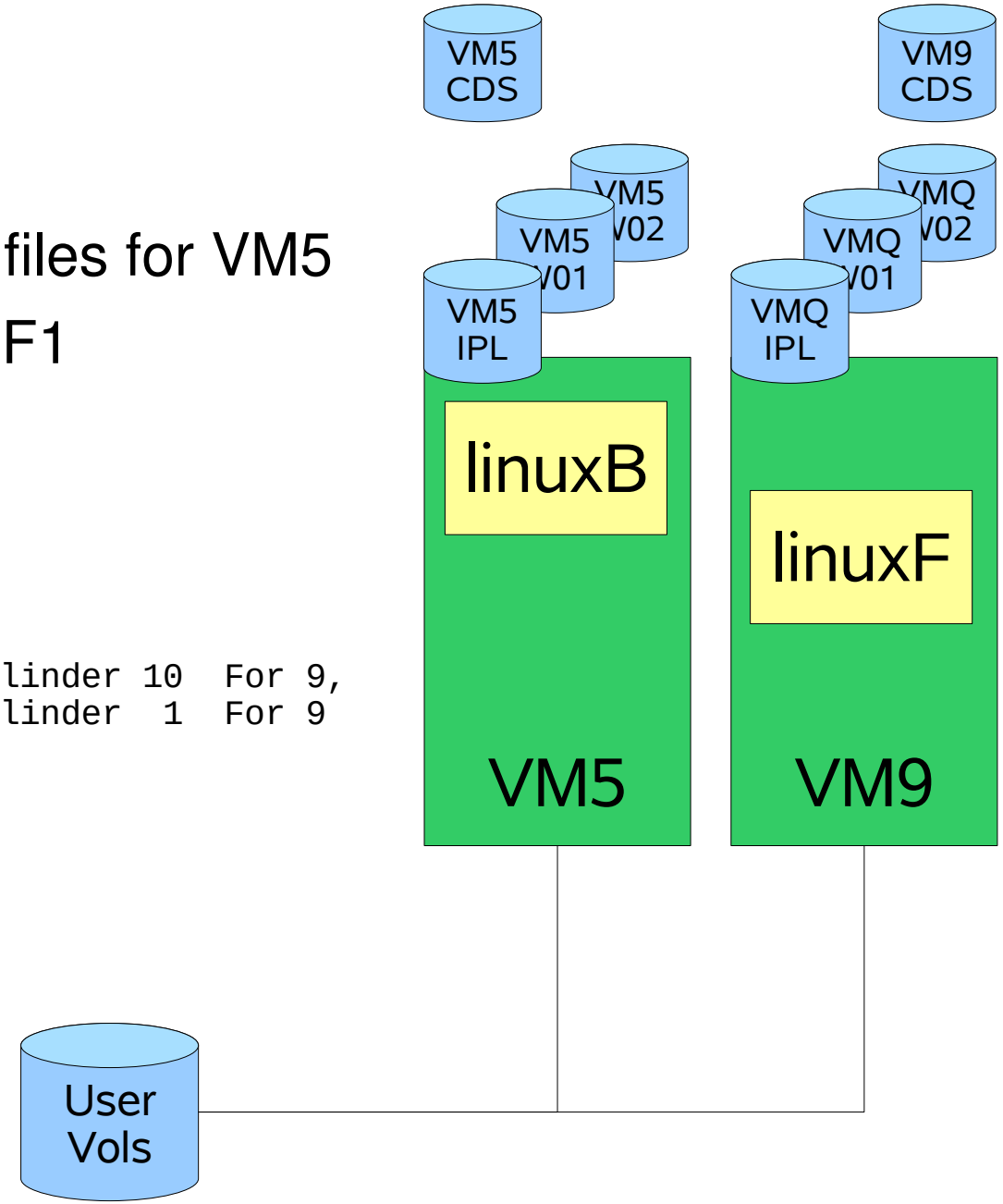
```
Imbed -SYSTEM- SYSRES
Imbed -SYSTEM- CPOWNED
```

- in VM5 SYSRES on CF1

```
System_Residence,
    Warmstart   Volid  VM5CDS From Cylinder 10  For 9,
    Checkpoint  Volid  VM5CDS From Cylinder  1  For 9
```

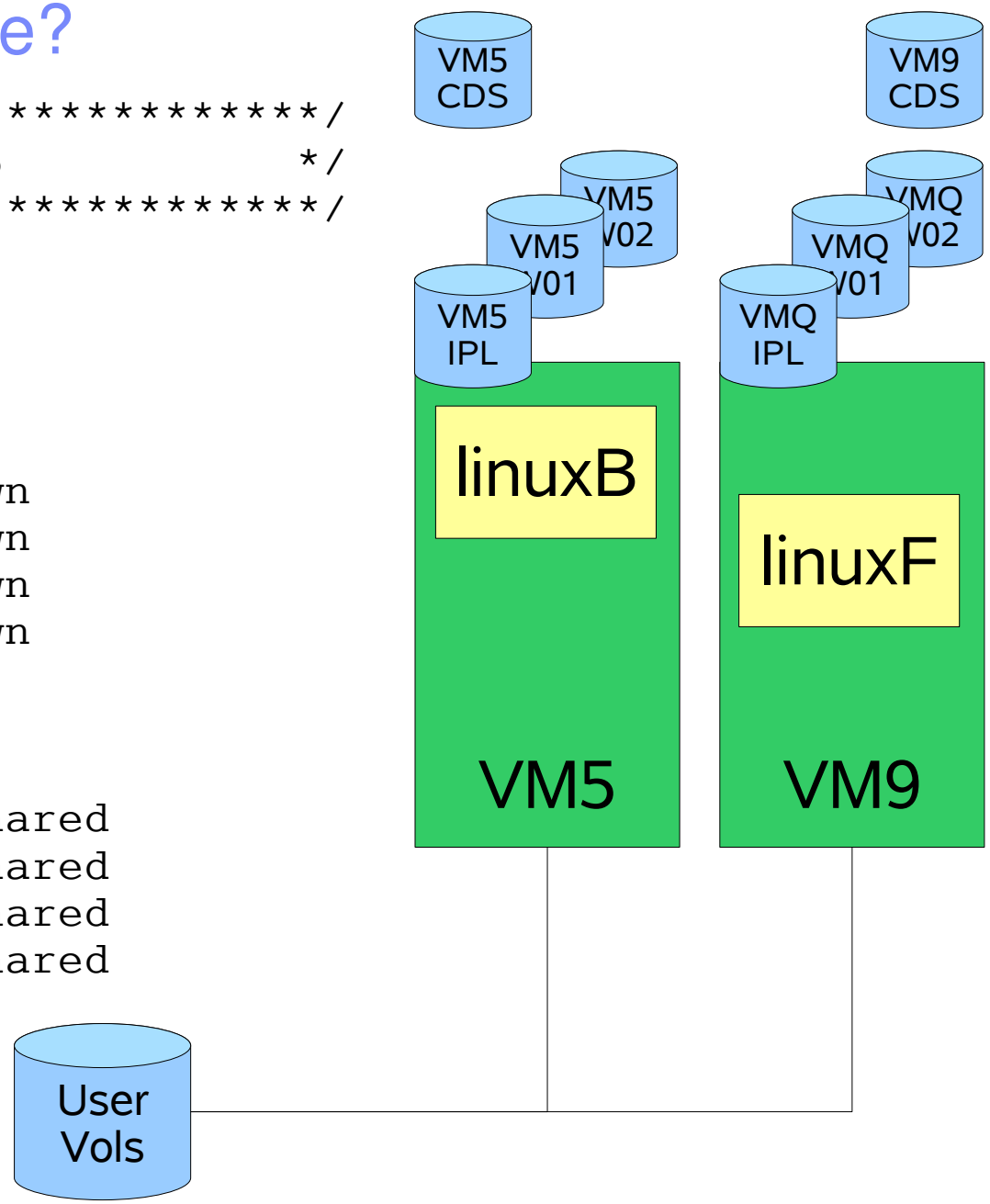
- in VM5 CPOWNED on CF1

```
CP_Owned   Slot   1  &SYSRES
CP_Owned   Slot   2  VM5CDS
```



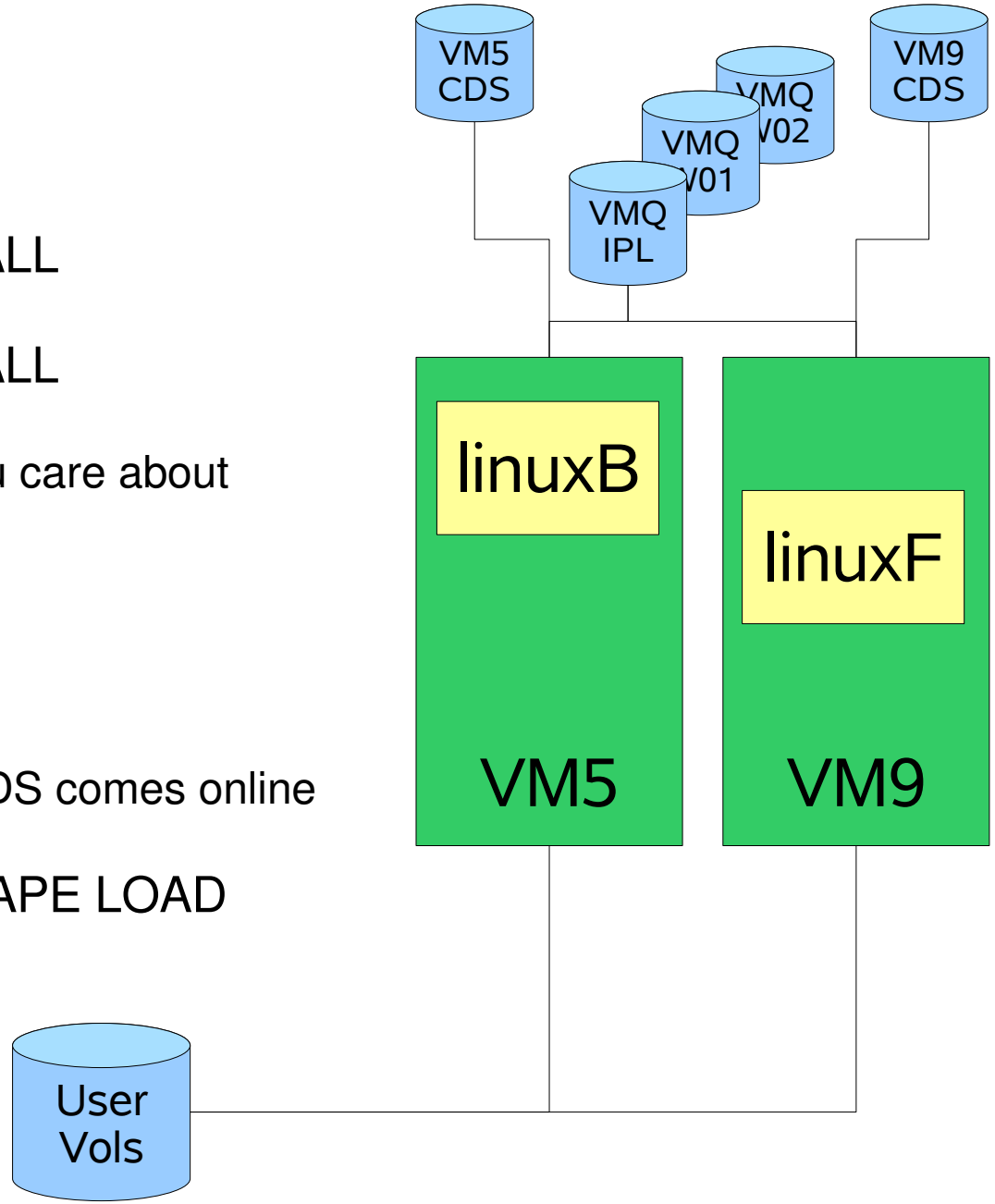
z/VM Shared SPOOL, anyone?

```
/*****************************************/
/*CP_Owned Volume Statements - VM5       */
/*****************************************/
    CP_Owned    Slot    1   &SYSRES
    CP_Owned    Slot    2   VM5CDS
    CP_Owned    Slot    3   RESERVED

    CP_Owned    Slot   37   RESERVED
    CP_Owned    Slot   38   VM5SP0 Own
    CP_Owned    Slot   39   VM5SP1 Own
    CP_Owned    Slot   40   VM5SP2 Own
    CP_Owned    Slot   41   VM5SP3 Own
    CP_Owned    Slot   42   RESERVED

    CP_Owned    Slot   53   RESERVED
    CP_Owned    Slot   54   VM9SP0 Shared
    CP_Owned    Slot   55   VM9SP1 Shared
    CP_Owned    Slot   56   VM9SP2 Shared
    CP_Owned    Slot   57   VM9SP3 Shared
    CP_Owned    Slot   58   RESERVED
```

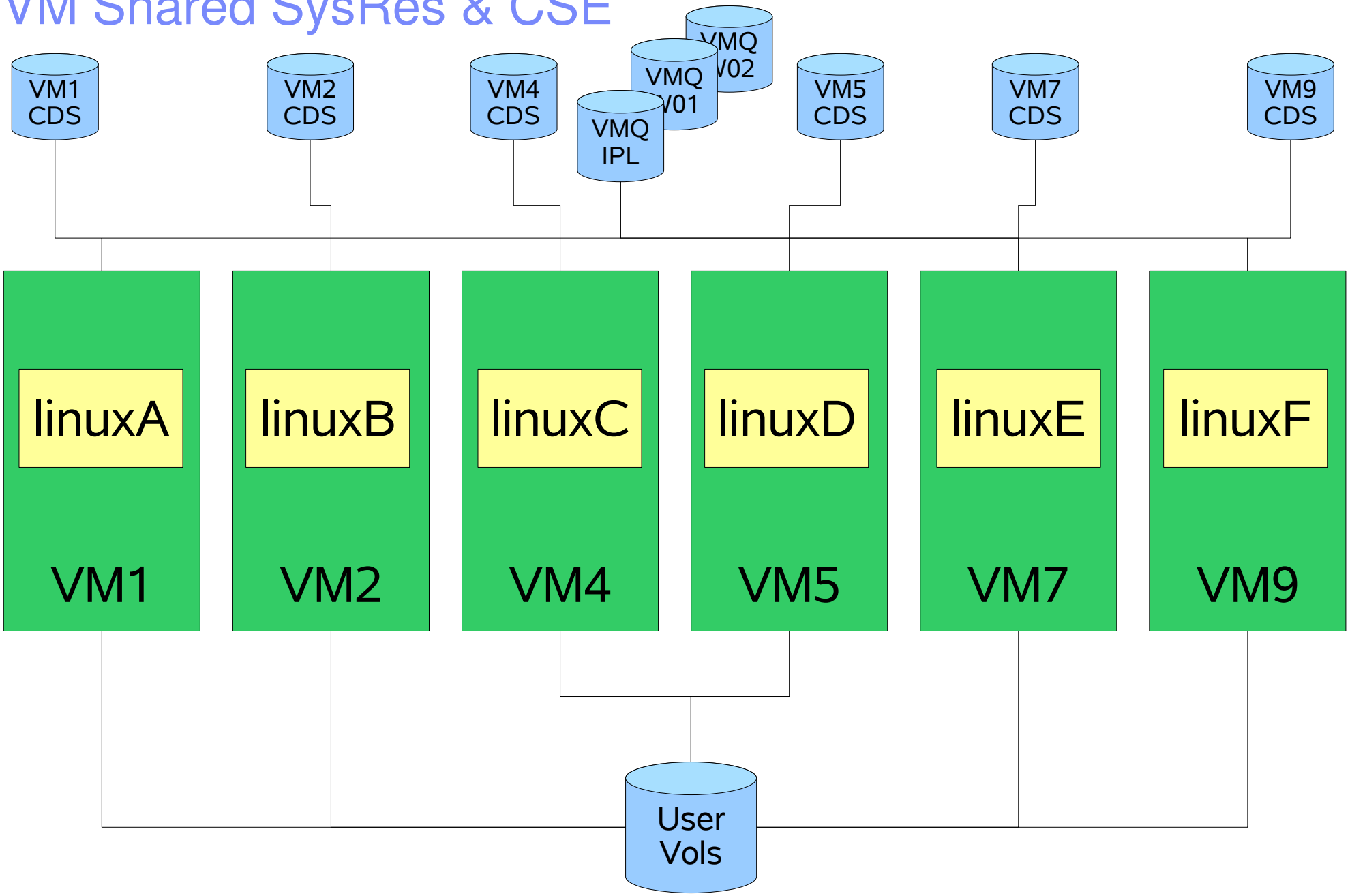


z/VM Shared SysRes

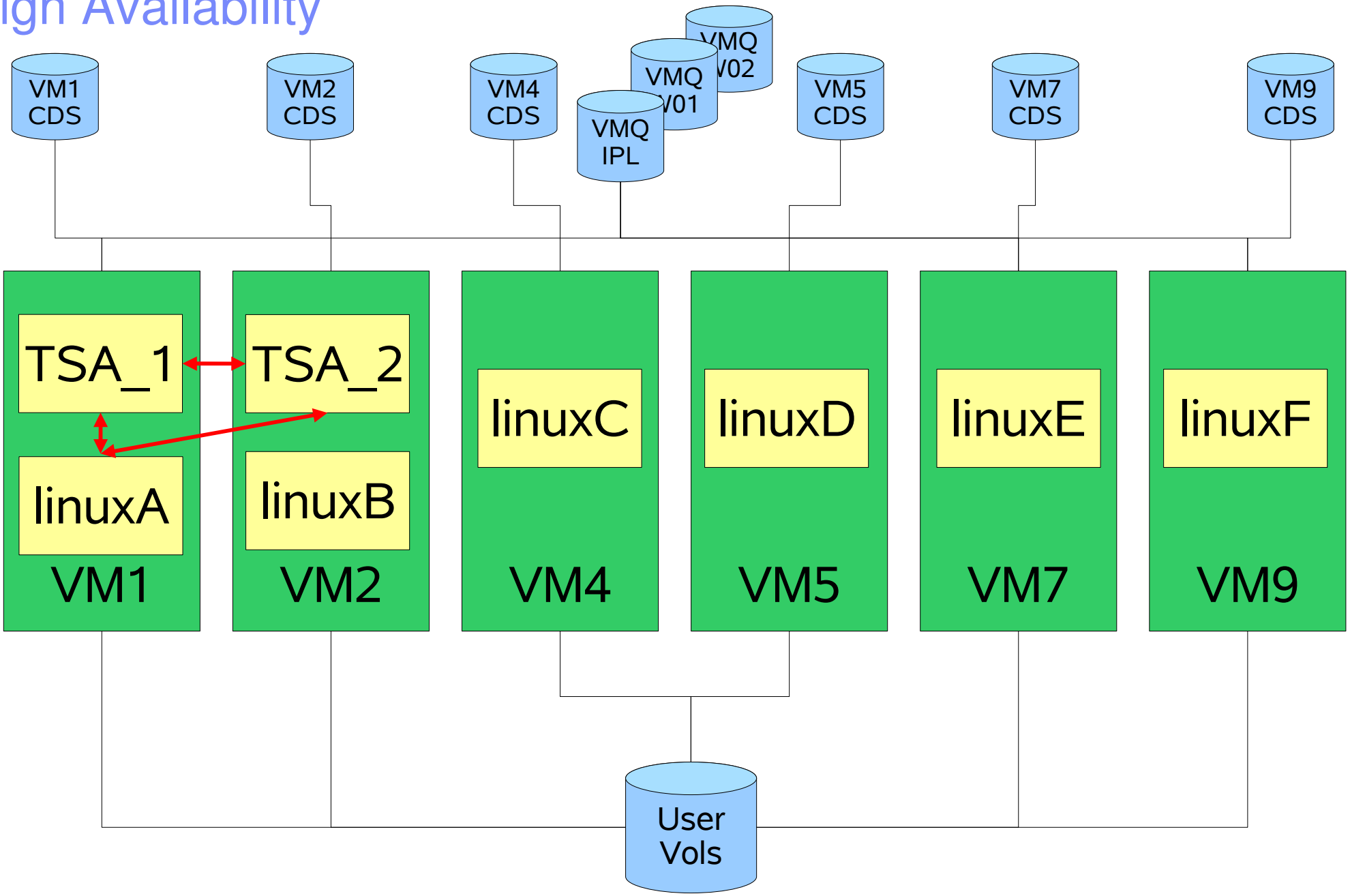
- On VM9: SPXTAPE DUMP SDF ALL
- On VM5: SPXTAPE DUMP STD ALL
  - Only if there's anything on VM5 you care about
- IPL VM5 from VMQIPL
  - CLEAN
  - Check that the Directory on VM5CDS comes online
- Restore the spool files with SPXTAPE LOAD



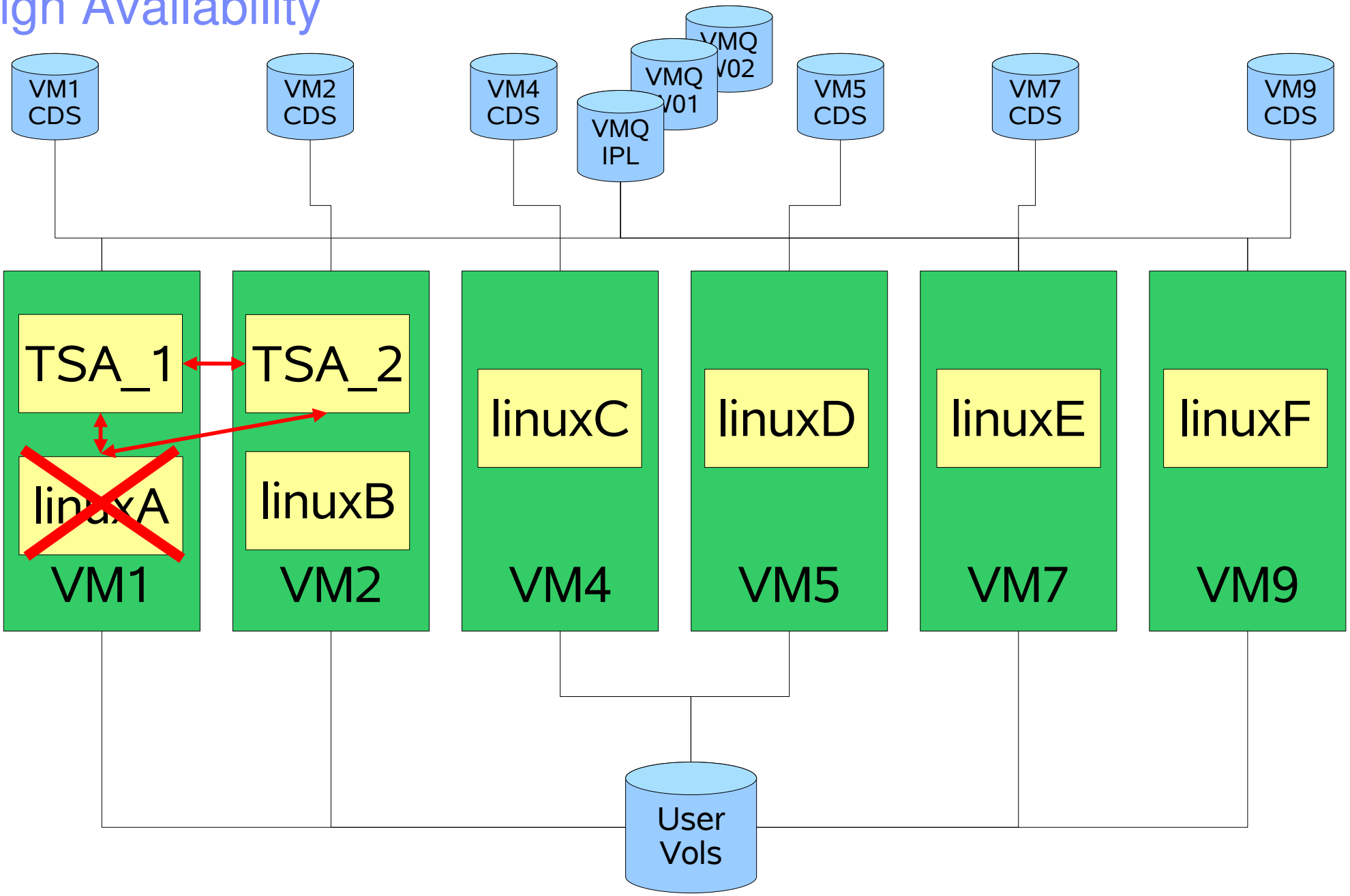
z/VM Shared SysRes & CSE



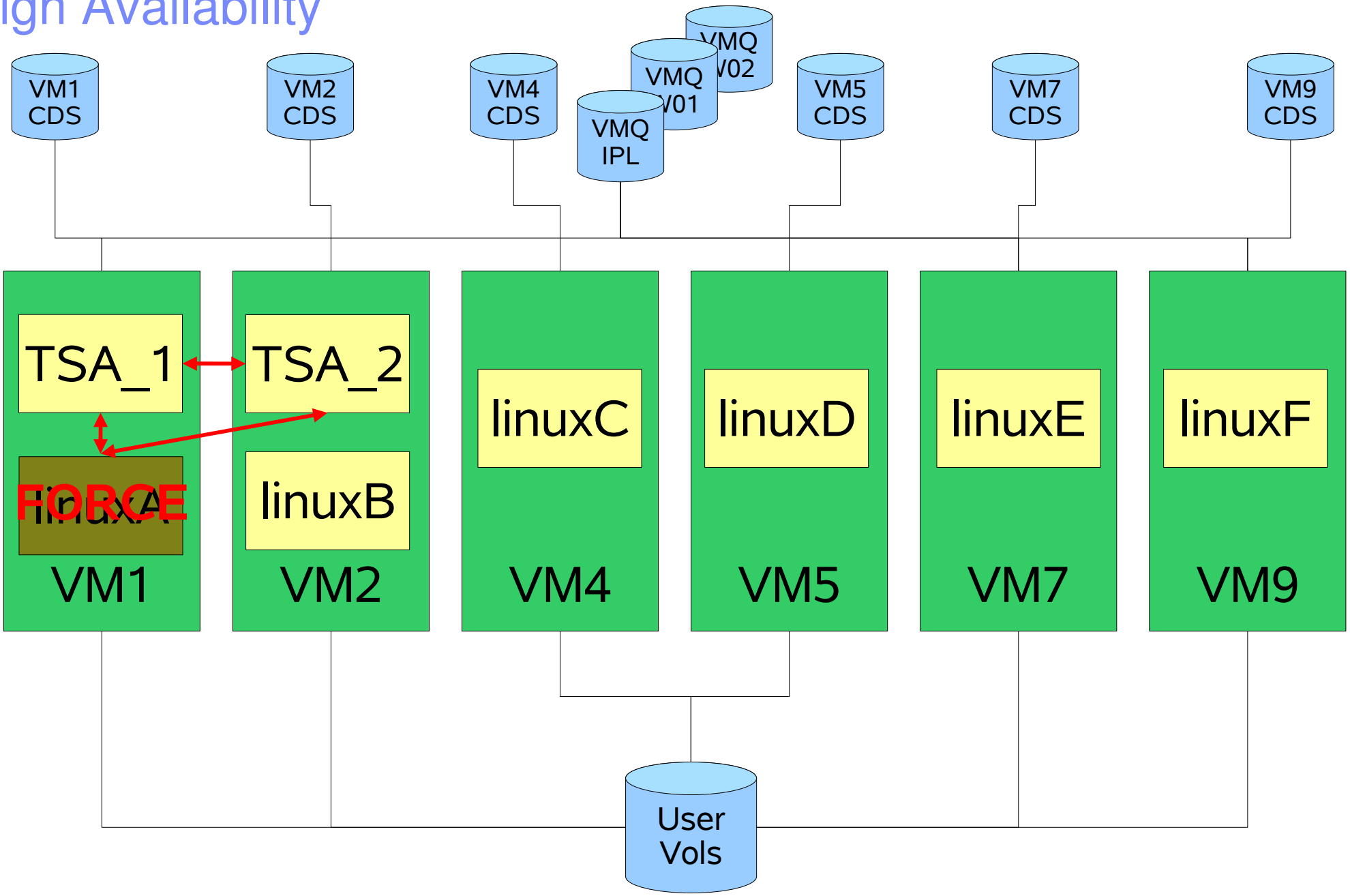
High Availability



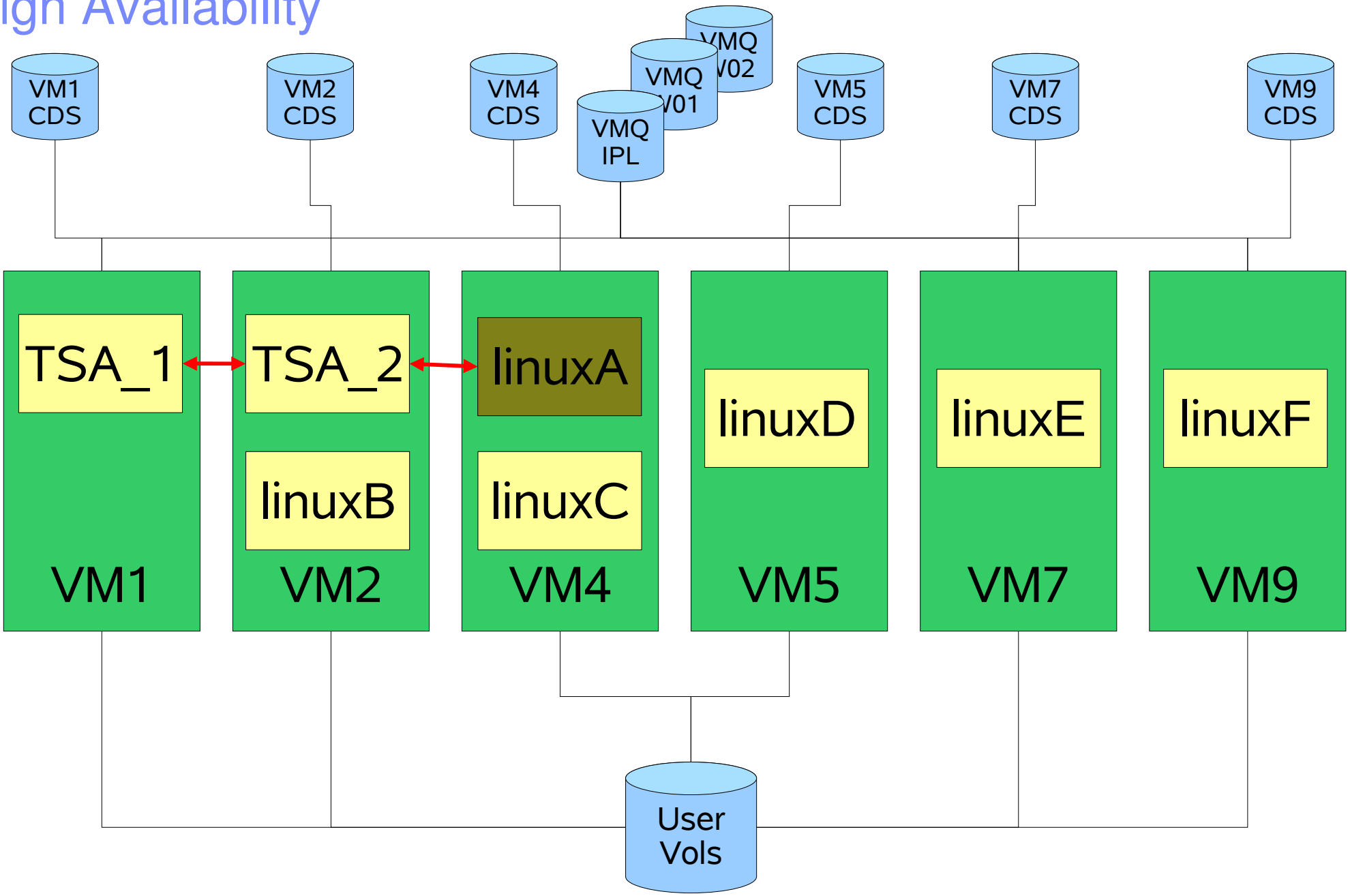
High Availability



High Availability



High Availability



Summary

- Maintaining Multiple VM systems is easier with CSE
- CSE allows greater flexibility in choosing where to run a workload
- Shared SysRes volumes will ease maintenance and provide more consistency
- CSE provides the infrastructure needed for HA