z/VM shared segments implementation for Linux on LinuxONE and IBM z

Yvon-Marie Avril
IBM, IT Specialist in Montpellier Center

2019 IBM GSE UK Region
5th november, 2019

Agenda

- Large Discontiguous Saved Segment (DCSS) definition.

- DCSS building steps.
  - z/VM defintions
  - Linux definitions

- Linux exploitation of the DCSS through ***dcssblk*** driver.
  - Sharing
  - Formatting
  - Filling
  - Saving

- Shared Segments with Docker (Demo).

Session Objectives

- Explain z/VM Shared Segment technology.
- Implement z/VM Shared Segment with Linux
- Illustrate Shared Segment usage with Docker (Demo)

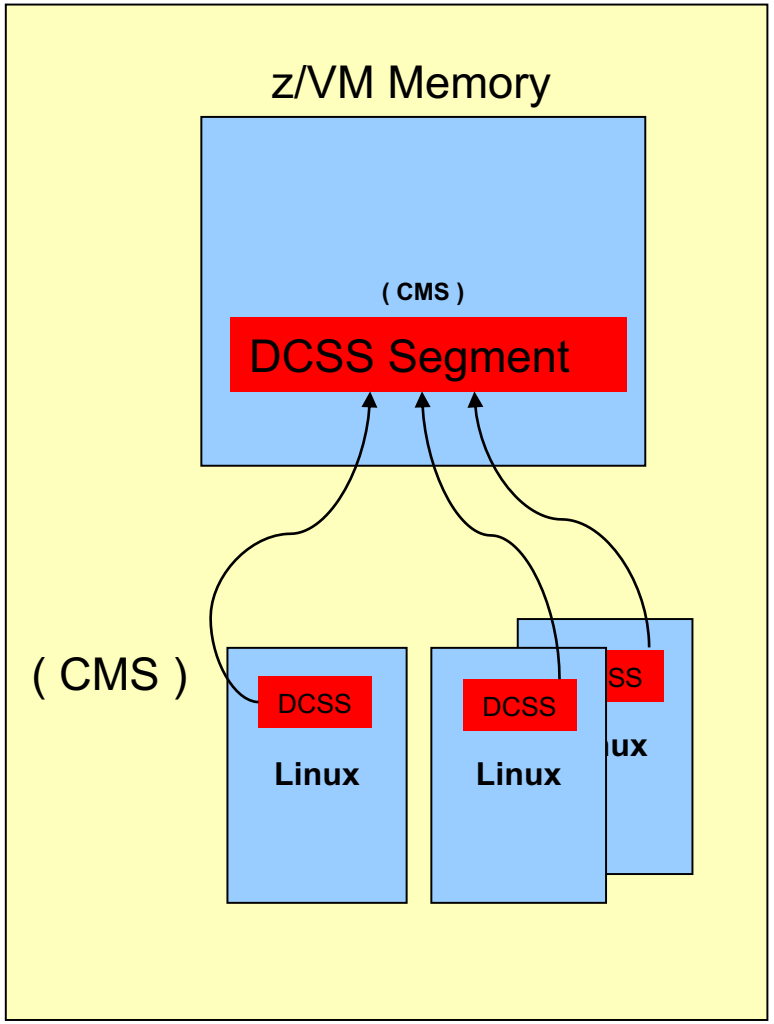
For more information on Large Discontiguous Shared Segment, see

z/VM Saved Segments Planning and Administration

Large Discontiguous Saved Segment (DCSS) definition.

Principles

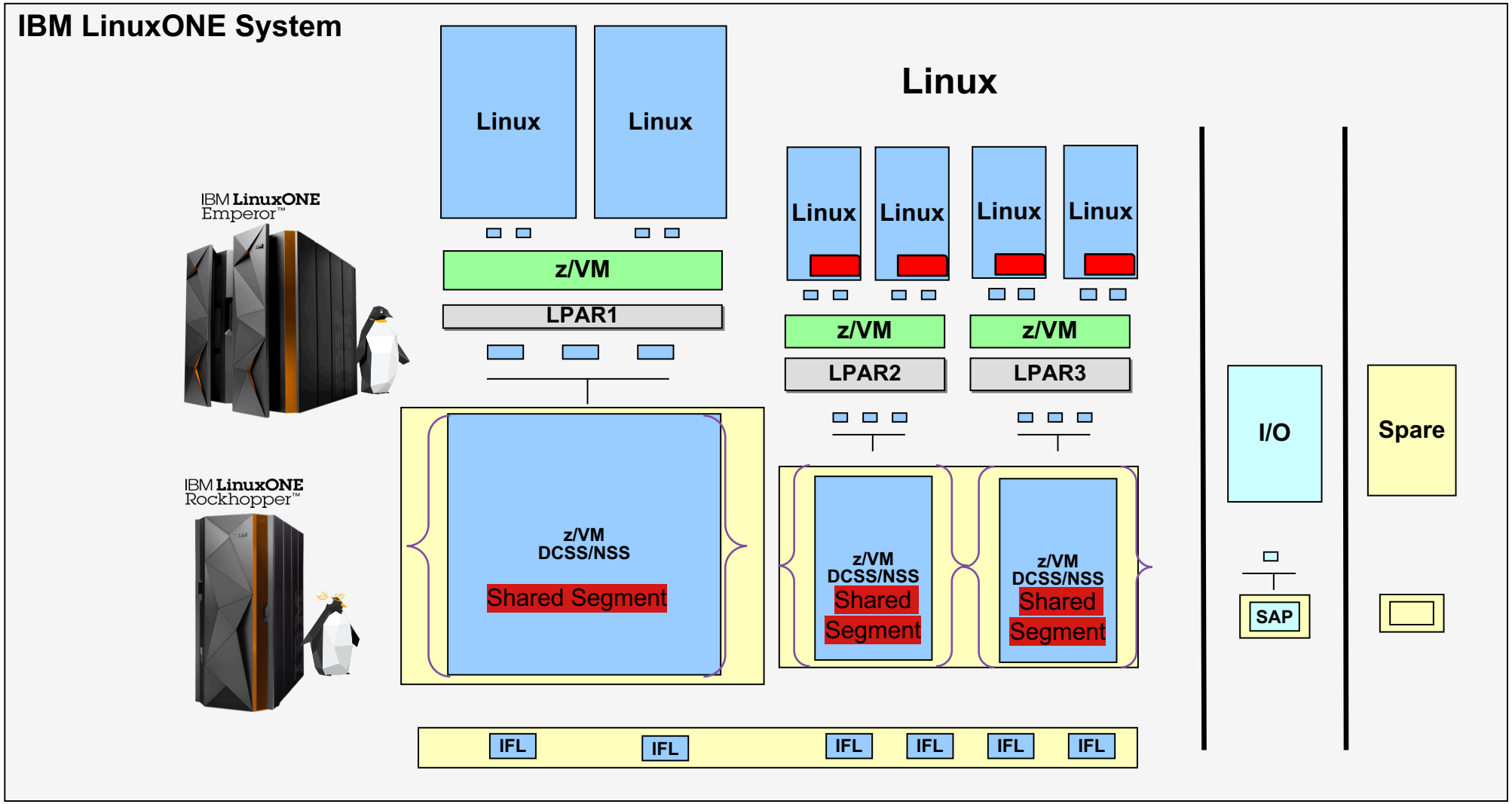
- A *saved segment* is a range of pages of virtual storage that you can define to hold data or re-enterable code (programs).



- Holds frequently used data and code.
- Can be dynamically attached to and detached from a VM.
- Permits re-enterable code to be shared by concurrently operating virtual machines.
- Using saved segments decreases the I/O rate.
- CMS : Internal subsystem for interactive interface with z/VM.

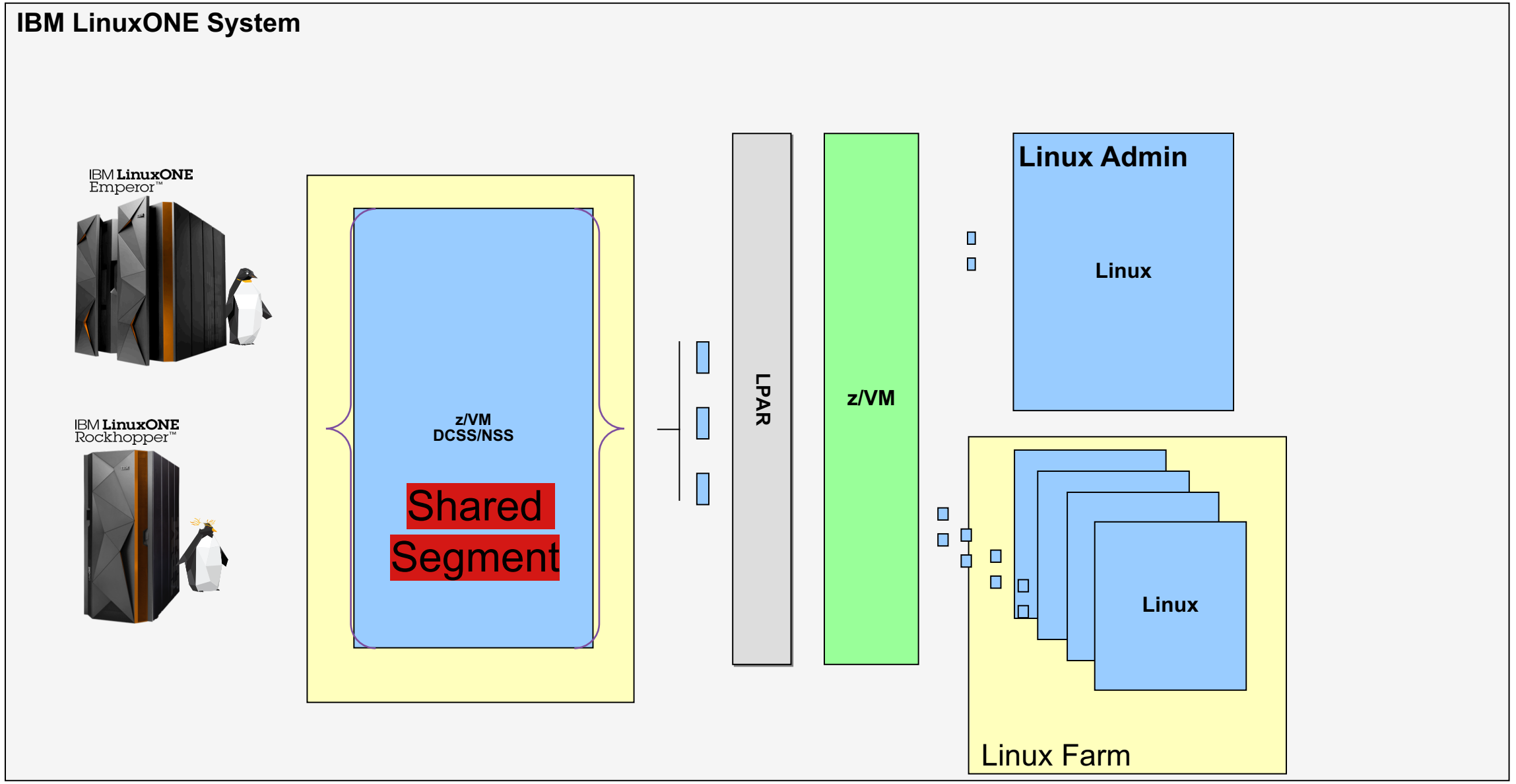
Linux OS exploits DCSS technology through a specific driver called ***DCSSBLK***

DCSS Exploitation with Linux



NSS : Name Saved System
DCSS : Load Discontiguous Saved Segment

Shared Read/Only access segment with Linux



DCSS building steps.

z/VM definitions

- A shared segment required available contiguous memory

- It must be defined from a suitable CMS virtual machine
  - Class E
  - Sufficient virtual memory in Virtual Machine

Building steps in z/VM :

- Decide for DCSS size and boundaries in z/VM memory.
- Create DCSS.
- Save the DCSS
- Modify Linux guest memory configuration.
- Add privilege class E for Linux guest (Administrator Role only) .

Commands to create the DCSS (z/VM)

- DEFSEG DOCK1803 40A00-49FFF SR LOADNSHR

HCPNSD440I Saved segment DOCK1803 was successfully defined in fileid <# fileid>

- SAVESEG DOCK1803

HCPNSS440I Saved segment DOCK1803 was successfully saved in fileid <# fileid>

Value are expressed as an address range in memory, in hexadecimal notation.
In this sample the segment size is 157 MB.
SR : Shared read-only access.
LOADNSHR …….

```
q nss map
FILE  FILENAME  FILETYPE  MINSIZE   BEGPAG  ENDPAG  TYPE  CL  #USERS  PARMREGS  VMGROUP
0029  DOCK1712  DCSS        N/A     5A000   61BFA    SR   A   00000    N/A       N/A
0032  DOCK1805  DCSS        N/A     4A000   51BFA    SR   A   00000    N/A       N/A
0035  DOCK1804  DCSS        N/A     53000   53EEE    SR   A   00000    N/A       N/A
0017  NLSUCENG  DCSS        N/A     02000   020FF    SR   A   00000    N/A       N/A
0016  NLSKANJI  DCSS        N/A     02000   020FF    SR   A   00000    N/A       N/A
0026  CMS       NSS     0000256K    00000   00D0D    EW   A   00016    00-15     NO
0001  CMSPIPES  DCSS        N/A     02100   022FF    SR   A   00021    N/A       N/A
0002  TCPIP     DCSS        N/A     10000   100FF    SN   R   00006    N/A       N/A
0007  DOCK1803  DCSS        N/A     40A00   49FFF    SR   A   00001    N/A       N/A
FILE  FILENAME  FILETYPE  BEGPAG          ENDPAG          TYPE  CL  #USERS
0013  DOCK1806  DCSSG     00000000080000  00000000080FFF   SR   A   00000
```

Starting address in Bytes for DOCK1803 : 1084227584 Bytes (>= 1GB)
Ending address in Bytes for DOCK1803 : 1241509888 Bytes (<= 1.3 GB)

For more information on DCSS, see
z/VM: CP Commands and Utilities Reference

Types of z/VM Shared Segments

Code	Meaning
EW	Exclusive read/write access.
EN	Exclusive read/write access, no data saved.
ER	Exclusive read-only access.
SW	Shared read/write access.
SN	Shared read/write access, no data saved.
SR	**Shared read-only access.**

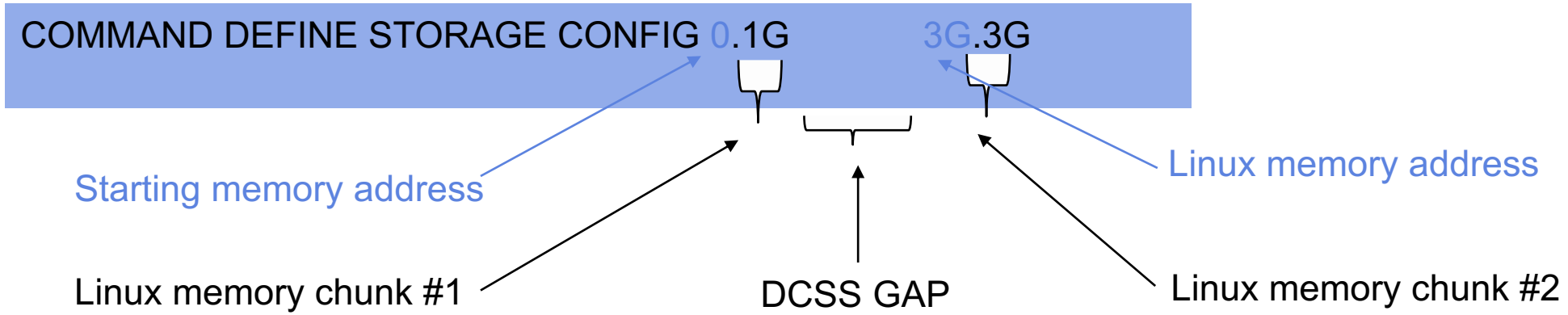
Comparaison of DCSS type SN and SR

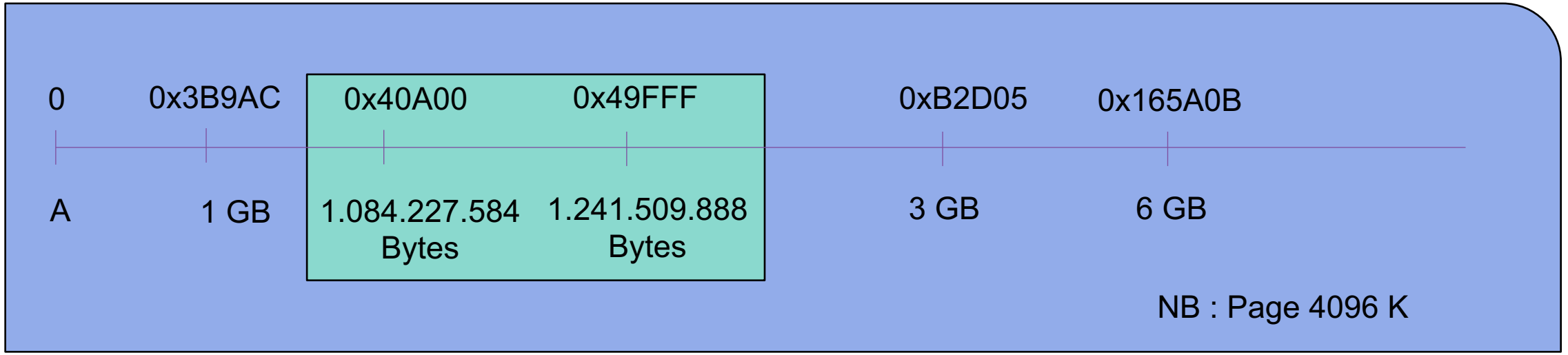
Consideration	DCSS Type	
	SN : Shared R/W	SR : Shared R/O
File system loaded into DCSS is written to z/VM spool	No	Yes
Spool processing for DCSS can delay other spool activity	No	Yes
Persistent after system restart	No	Yes

Change linux directory z/VM directory to redefine memory configuration

Example given for a 4GB linux virtual memory size

Linux guest of 4 Gb user definition





Linux exploitation of the DCSS through dcssblk driver.

Linux Tasks in Administrator / Linux Farm

Statements for Linux Administrator

- Load the DCSS

#echo DOCK1803 > /sys/devices/dcssblk/add

- *Share the DCSS*

echo 0 > /sys/devices/dcssblk/DOCK1803/shared

- *Format the DCSS*

mkfs -t ext4 /dev/dcssblk0

- Mount the DCSS as filesystem

mount /dev/dcssblk0 /usr/local/docker

- *Save the DCSS*

echo 1 > /sys/devices/dcssblk/DOCK1803/save

- Umount the DCSS

umount /dev/dcssblk0

- Unload the DCSS

echo DOCK1803 > /sys/devices/dcssblk/remove

Statements for Linux Farm

- Load the DCSS

echo DOCK1803 > /sys/devices/dcssblk/add

- Mount the DCSS as filesystem

mount /dev/dcssblk0 /usr/local/docker

mount: /dev/dcssblk0 is write-protected, mounting read-only

For Information :

- Umount the DCSS

umount /dev/dcssblk0

- Unload the DCSS

echo DOCK1803 > /sys/devices/dcssblk/remove

NB : The guest needs to have privclass E to save the DCSS contents

Industrialisation/Automation

rc.local script example :

```
#!/bin/bash

SEGMENT="DOCK1803"

echo ${SEGMENT} > /sys/devices/dcssblk/add

# Binaries Docker Segment Mounting
if [ -d /sys/devices/dcssblk/${SEGMENT}/ ]; then
        dcssID=`ls /sys/devices/dcssblk/${SEGMENT}/block/ | sed
          's/^.\{7\}//g'`
        mount -o ro /dev/dcssblk${dcssID} /usr/local/docker
fi
```

Advantages/Use cases/Recommendations

- Carrefully decide for DCSS content
- Heavily used binairies (Performance/optimisation)
- Farm of identical binairies (Server quantity)

- Same physical storage
- Swap disk
- A single administrator in charge of update DCSS content

Shared Segments with Docker (Demo).

Presenter demonstration

Demo Docker Demo with binary installation

docker version
docker images
docker run -it ubuntu /bin/bash
hostname
exit
docker ps | grep ubun
docker ps -a | grep ubun
docker restart baa194ad0029
docker ps |grep ubun
docker exec baa194ad0029 hostname
docker attach baa194ad0029
exit
docker ps | grep ubuntu
docker ps -a | grep baa194ad0029
docker rm baa194ad0029
docker rmi ubuntu:latest
Docker images

Linux Farm IPL log

- When the Linux farm starting, the storage configuration

```
Storage cleared - system reset.
STORAGE = 4G
Storage Configuration:
0.1G 3G.3G
Extent Specification                    Address Range
---------------------------------------- ----------------------------------------
                  0.1G                  0000000000000000 - 000000003FFFFFFF
                3G.3G                   00000000C0000000 - 00000017FFFFFFFF
Storage cleared - system reset.
```

- Boot starting

```
zIPL v2.4.0-1.el7 interactive boot menu

  0. default (4.14.0-115.2.2.el7a.s390x)

  1. 4.14.0-115.2.2.el7a.s390x
  2. linux
  3. linux-0-rescue-fbe272327ebc41a1ac2a6cb1a6dcc172
Note: VM users please use '#cp vi vmsg <input>'
Please choose (default will boot in 5 seconds):
Booting default (4.14.0-115.2.2.el7a.s390x)
```

- DCSS DOCK1803 loading

```
[  [32m  OK  [0m] Started System Logging Service.
[    7.087968] extmem: DCSS DOCK1803 of range 0000000040a00000 to 0000000049fff
```

- rc.local script executing
- DCSS DOCK1803 mounting

```
ff and type SR loaded in shared access mode
[    7.088045] dcssblk: Loaded DOCK1803 with total size 157286400 bytes and capa
city 307200 sectors
[  [32m  OK  [0m] Started Permit User Sessions.
[  [32m  OK  [0m] Started Command Scheduler.
[  [32m  OK  [0m] Started Job spooling tools.
[  [32m  OK  [0m] Started OpenSSH server daemon.
[    7.255525] EXT4-fs (dcssblk0): mounted filesystem with ordered data mode. Op
ts: (null)
[  [32m  OK  [0m] Started /etc/rc.d/rc.local Compatibility.
[  [32m  OK  [0m] Started Serial Getty on ttysclp0.
[  [32m  OK  [0m] Started Serial Getty on ttyS0.
[  [32m  OK  [0m] Started Getty on tty1.
         Starting Terminate Plymouth Boot Screen...
[  [32m  OK  [0m] Started Serial Getty on hvc0.
[  [32m  OK  [0m] Reached target Login Prompts.
[    8.434317] Bridge firewalling registered
[    8.641075] IPv6: ADDRCONF(NETDEV_UP): docker0: link is not ready

CentOS Linux 7 (AltArch)
Kernel 4.14.0-115.2.2.el7a.s390x on an s390x

berlin-user login:
```

Session summary

- See DCSS exploitation of the Read Only Shared Segment
- Spoke about general recommendation
- Heavily used binary
- Linux farm that uses the same things
- No need to connect another hardware
- Relatively easy to implement
- Maybe only thing is you need a calculator

Follow Us

External platforms

- Website ibm.com/ibm/clientcenter/montpellier/
- Twitter @IBMCCMPL
- Virtual 360° Tour
- Videos YouTube - Client Center Montpellier Channel
- Images Flickr flickr.com/ibmccmpl

Internal IBM

- IBM Connections Community Client Center Montpellier

Notices and disclaimers

- © 2019 International Business Machines Corporation. No part of this document may be reproduced or transmitted in any form without written permission from IBM.
- **U.S. Government Users Restricted Rights — use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM.**
- Information in these presentations (including information relating to products that have not yet been announced by IBM) has been reviewed for accuracy as of the date of initial publication and could include unintentional technical or typographical errors. IBM shall have no responsibility to update this information. **This document is distributed "as is" without any warranty, either express or implied. In no event, shall IBM be liable for any damage arising from the use of this information, including but not limited to, loss of data, business interruption, loss of profit or loss of opportunity.** IBM products and services are warranted per the terms and conditions of the agreements under which they are provided.
- IBM products are manufactured from new parts or new and used parts.
  In some cases, a product may not be new and may have been previously installed. Regardless, our warranty terms apply."
- **Any statements regarding IBM's future direction, intent or product plans are subject to change or withdrawal without notice.**
- Performance data contained herein was generally obtained in a controlled, isolated environments. Customer examples are presented as illustrations of how those
- customers have used IBM products and the results they may have achieved. Actual performance, cost, savings or other results in other operating environments may vary.
- References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business.
- Workshops, sessions and associated materials may have been prepared by independent session speakers, and do not necessarily reflect the views of IBM. All materials and discussions are provided for informational purposes only, and are neither intended to, nor shall constitute legal or other guidance or advice to any individual participant or their specific situation.
- It is the customer's responsibility to insure its own compliance with legal requirements and to obtain advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulatory requirements that may affect the customer's business and any actions the customer may need to take to comply with such laws. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the customer follows any law.

Notices and disclaimers continued

- Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products about this publication and cannot confirm the accuracy of 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. IBM does not warrant the quality of any third-party products, or the ability of any such third-party products to interoperate with IBM's products. **IBM expressly disclaims all warranties, expressed or implied, including but not limited to, the implied warranties of merchantability and fitness for a purpose.**
- The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents, copyrights, trademarks or other intellectual property right.
- IBM, the IBM logo, ibm.com and [names of other referenced IBM products and services used in the presentation] are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at: www.ibm.com/legal/copytrade.shtml