Customer Experiences Running Databases with zVM

David Simpson
IBM Washington Systems Center
Oracle Certified & Open-Source Database Specialist
simpson.dave@us.ibm.com

Justice Heughan
IBM Washington Systems Center
Technical Enablement Specialist
Justice.Heughan@ibm.com

State Government Customer upgrade RHEL 7 to RHEL 9- Oracle RAC

Oracle Database 19.0.0.0.0 is certified on Linux on IBM Z Red Hat Enterprise Linux 9 Update 2+

Notes

Oracle Database 19.0.0.0.0 with Linux on IBM Z Red Hat Enterprise Linux 9

- Red Hat Enterprise Linux 9.2: **5.14.0-284.30.1.el9_2.s390x** or later
- Minimum RU: **19.23**

Checks		
Checks		
[Packages]		
Package: compat-libcap1-1.10 (s390x)	Warning	No
Package: libstdc++-devel-4.8.2 (s390x)	Warning	No
Package: libaio-devel-0.3.109 (s390x)	Warning	No
User Mask	Failed	No
resolv.conf Integrity	Failed	No
RPM Package Manager database	Info	No

Checks	Status	Fixable
Checks		
Swap Size	Warning	No
[Packages]		
Package: compat-libcap1-1.10 (s390x)	Warning	No

- Oracle 19c released 6 years ago (no SLES 15/RHEL 9 back then)
- Minimum Oracle RU 19.23 – recommend 19.27 (or the latest RU)
- Requirements for Installing Oracle Database 19c on RHEL9 64-bit (IBM Linux on System z) (Doc ID 3068629.1)
- Requirements for Installing Oracle Database 19c 64-bit IBM Linux on System Z on SLES 15 (Doc ID 2813961.1)
  - Get latest RPM Checker Installing Oracle Database 19c on IBM Linux on System z (Doc ID 2553465.1)
  - gridSetup.sh -applyRU $STAGE/37641958 -applyOneOffs $STAGE/37641958/37654975,$STAGE/37499406

Using a Test Environment to Investigate Oracle RAC VNIC Issue

Customer's RHEL 9 install was creating network interface names based on default VNIC definition
Oracle RAC requires the Linux network Interface names to be the same

- Recreated customer challenge in personal z/VM environment (JHVM)
- **Issue:** Oracle RAC interconnect interface not coming online on customer guest
- Used test system to test VNIC definitions, persistence, and network files
- **Goal:** Determine root cause and confirm proper setup process before applying in production

z/VM: VNIC Definition and Static MAC Setup

- Defined new VNIC using NICDEF in directory entry
- Attached to SYSTEM VSWITCH
- Validated interface with CP Q V NIC

```
00:
00: CP QUERY VIRTUAL NIC
00: Adapter 0600.P00 Type: QDIO      Name: HYD1G1      Devices: 3
00:   MAC: 02-15-23-00-00-67         VSWITCH: SYSTEM VSW
00: Adapter 0700.P00 Type: QDIO      Name: HYD1G1      Devices: 3
00:   MAC: 02-15-23-00-00-5B         VSWITCH: SYSTEM VSW
00: Adapter 1000.P00 Type: QDIO      Name: HYD1G1      Devices: 3
00:   MAC: 02-15-23-00-00-66         VSWITCH: SYSTEM VSW

                                                    RUNNING    JHVM
MA    a                                                        42/001
```

RHEL: Enablement and Device Visibility

- Removed NIC from ignore list: cio_ignore -R
- Verified with lszdev; enabled device: chzdev -e
- Made persistent: chzdev -e -p
- Confirmed interface appears as enc700

```
[root@rhel92vm ~]# nmcli con show
NAME            UUID                                  TYPE      DEVICE
System enc1000  e2c765dd-0096-4e11-a95c-a7923450a610  ethernet  enc1000
enc600          415ecb48-4dad-40f7-8182-4d62d209b2b0  ethernet  enc600
lo              0a9e5018-d4c8-421e-8d4e-ef324a302c2f  loopback  lo
enc700          7ee994ea-67f5-4756-aaf7-a4204b3609ec  ethernet  --
[root@rhel92vm ~]# cio_ignore -R
[root@rhel92vm ~]# lszdev
TYPE         ID                        ON   PERS  NAMES
dasd-eckd    0.0.0190                  no   no
dasd-eckd    0.0.019d                  no   no
dasd-eckd    0.0.019e                  no   no
dasd-eckd    0.0.cc66                  yes  no    dasda
dasd-eckd    0.0.cc67                  yes  no    dasdb
dasd-eckd    0.0.cc68                  yes  no    dasdc
dasd-eckd    0.0.cc80                  yes  no    dasdd
dasd-eckd    0.0.cc81                  yes  no    dasde
dasd-eckd    0.0.cc82                  yes  no    dasdg
dasd-eckd    0.0.cc83                  yes  no    dasdh
dasd-eckd    0.0.cc84                  yes  no    dasdf
qeth         0.0.0600:0.0.0601:0.0.0602  yes  no    enc600
qeth         0.0.0700:0.0.0701:0.0.0702  no   no
qeth         0.0.1000:0.0.1001:0.0.1002  yes  no    enc1000
generic-ccw  0.0.0009                  yes  no
generic-ccw  0.0.000c                  no   no
generic-ccw  0.0.000d                  no   no
generic-ccw  0.0.000e                  no   no
[root@rhel92vm ~]# chzdev -e 700
QETH device 0.0.0700:0.0.0701:0.0.0702 configured
[root@rhel92vm ~]# chzdev -e -p 700
Configuring devices in the persistent configuration only
QETH device 0.0.0700:0.0.0701:0.0.0702 already configured
[root@rhel92vm ~]# nmcli con show
NAME            UUID                                  TYPE      DEVICE
System enc1000  e2c765dd-0096-4e11-a95c-a7923450a610  ethernet  enc1000
enc600          415ecb48-4dad-40f7-8182-4d62d209b2b0  ethernet  enc600
enc700          7ee994ea-67f5-4756-aaf7-a4204b3609ec  ethernet  enc700
lo              0a9e5018-d4c8-421e-8d4e-ef324a302c2f  loopback  lo
[root@rhel92vm ~]#
```

NetworkManager: Creating and Verifying Interface Profile

- Created profile via nmcli connection add ...
- Verified status using nmcli con show
- Rebooted guest: confirmed profile and VNIC remained active

State Government Customer upgrade RHEL 7 to RHEL 9- Oracle RAC

- OK great it works in the Lab, let's try this with the customer….

```
[[root@rhel92vm ~]# ip link show
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN mode DEFAULT group default qlen 1000
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
2: enc600: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP mode DEFAULT group default qlen 1000
    link/ether 02:15:23:00:00:65 brd ff:ff:ff:ff:ff:ff
3: enc1000: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP mode DEFAULT group default qlen 1000
    link/ether 02:15:23:00:00:64 brd ff:ff:ff:ff:ff:ff
4: priv1: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP mode DEFAULT group default qlen 1000
    link/ether 02:15:23:00:00:5b brd ff:ff:ff:ff:ff:ff
[[root@rhel92vm ~]# znetconf -c
Device IDs                 Type    Card Type      CHPID Drv. Name             State
--------------------------------------------------------------------------------
0.0.0700,0.0.0701,0.0.0702 1731/01 Virt.NIC QDIO     02 qeth priv1            online
0.0.0600,0.0.0601,0.0.0602 1731/01 Virt.NIC QDIO     01 qeth enc600           online
0.0.1000,0.0.1001,0.0.1002 1731/01 Virt.NIC QDIO     00 qeth enc1000          online
[[root@rhel92vm ~]# ip link show
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN mode DEFAULT group default qlen 1000
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
2: enc600: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP mode DEFAULT group default qlen 1000
    link/ether 02:15:23:00:00:65 brd ff:ff:ff:ff:ff:ff
3: enc1000: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP mode DEFAULT group default qlen 1000
    link/ether 02:15:23:00:00:64 brd ff:ff:ff:ff:ff:ff
4: priv1: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP mode DEFAULT group default qlen 1000
    link/ether 02:15:23:00:00:5b brd ff:ff:ff:ff:ff:ff
[root@rhel92vm ~]#
```

Oracle RAC, we need exact same network interface name on all hosts,
…..Linux reboot was fine, logoff a guest did not maintain device persistence…

Interface MAC Address Alignment for Oracle RAC

- RAC Interconnect requires persistent MAC addresses
- Resolved by setting MAC address in directory entry
- Ensures interconnect recognizes the interface upon boot
- NICDEF 0700 TYPE QDIO LAN SYSTEM VSW MACID 00005B

IMPORTANT

Use a file name with a **70-** prefix to keep the file names consistent with the **udev** rules-based solution.

For example, create the **/etc/systemd/network/70-provider0.link** file with the following content to rename the interface with MAC address **00:00:5e:00:53:1a** to **provider0**:

```
[Match]
MACAddress=00:00:5e:00:53:1a

[Link]
Name=provider0
```

Summary and Impact

- Used lab system to test z/VM VNIC setup
- Validated persistence across reboots, profile creation, MAC settings
- Shared working steps and syntax with customer to replicate successfully
- Persistent interface across reboots

State Government Customer upgrade RHEL 7 to RHEL 9- Oracle RAC

- Thanks Justice! Let's get back to the Oracle RAC install

Oracle RAC does not restart on reboot automatically without chkconfig being installed!
What happened to chkconfig in RHEL 9????

Ohasd Process Does Not Start After Rebooting OS On RHEL 9 Due To Missing Chkconfig Package (Doc ID 3000775.1)

This is because the chkconfig package is not installed.

It is not installed by default on RHEL 9.

There is a Unpublished document defect that requires the chkconfig package to be described in the manual requirements.

Bug 36161566 : ADD 'CKCONFIG' PACKAGE AS A REQUIRED PACKAGE FOR RED HAT LINUX 9 RELEASES

chkconfig package is not present in RHEL 9 by default

SOLUTION VERIFIED - Updated June 13 2024 at 5:36 PM - English

Environment

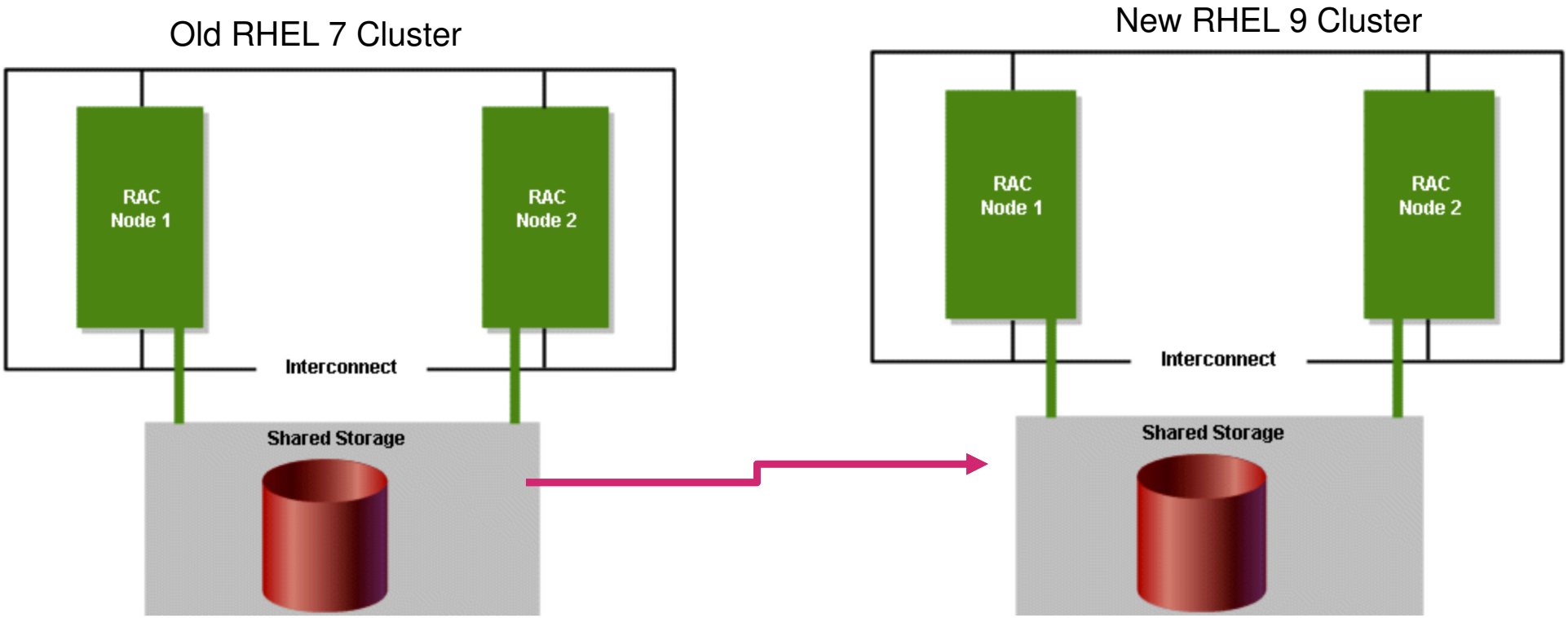
- RHEL 9

Issue

- `chkconfig` package is not present in `RHEL 9` by default.

State Government Customer upgrade RHEL 7 to RHEL 9- Oracle RAC

- OK let's try this one more time!
- Migration successful by creating new RHEL 9 cluster and "swinging" DB disk to new cluster
- Why not "in-place upgrade?" – criticality of their system if something went wrong



State Government Customer upgrade RHEL 7 to RHEL 9- Oracle RAC

- Let's plan for Production… 12 Ips available on the subnet
  - 8 are in use for Prod
  - we needed 7 new IPs for a new cluster…
  - Oh snap – we need another migration methodology!

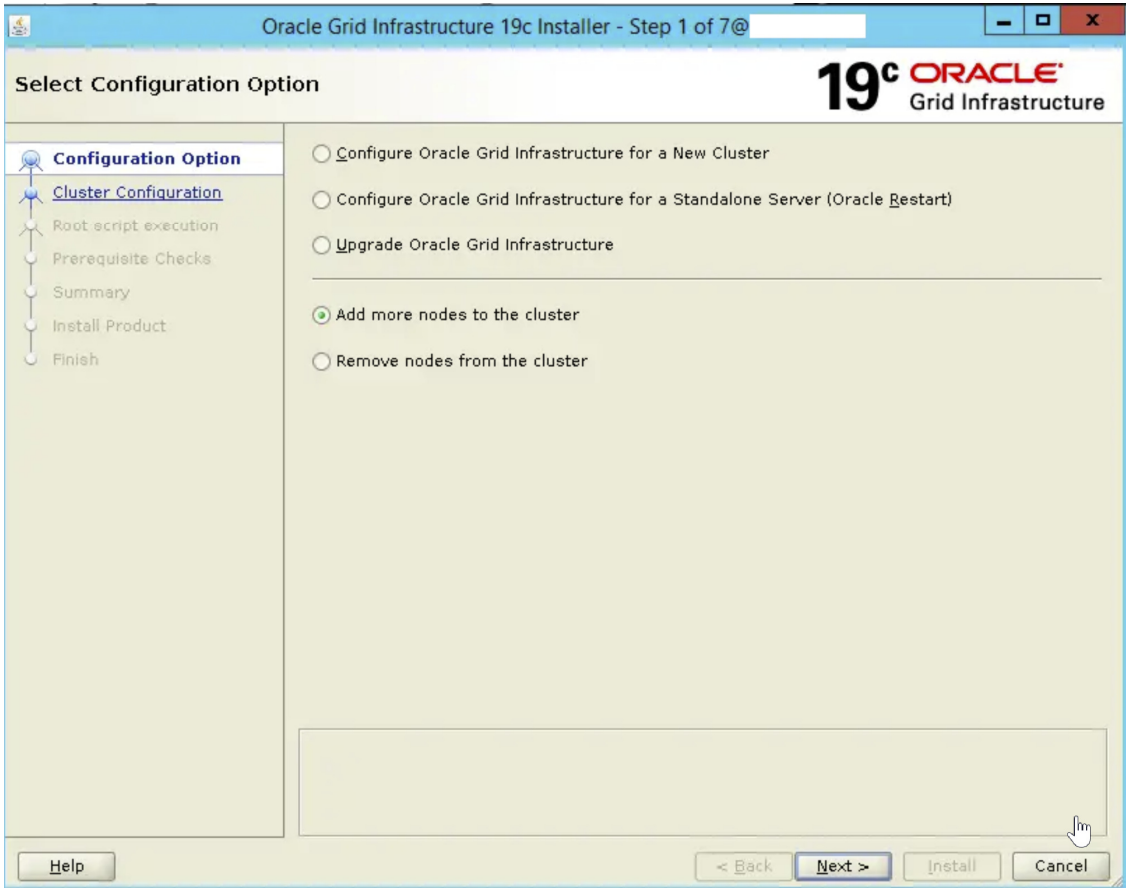
Guest	No. vCPU	RAM	IP	VLAN ID	Hostname	Network Mask
RESERVED			134.187	1387	**Network IP	255.255.255.240
RESERVED			134.187.	1387	**GATEWAY	255.255.255.240
RESERVED			134.187.	1387	**SWITCH	255.255.255.240
PDPHOR01	3	32GB	134.187.	1387	PDPHOR01	255.255.255.240
PDPHORR1	3	32GB	134.187.	1387	PDPHORR1	255.255.255.240
01 VIP			134.187	1387		255.255.255.240
R1 VIP			134.187.	1387		255.255.255.240
SCAN IP			134.187.	1387		255.255.255.240
SCAN IP			134.187.	1387		255.255.255.240
SCAN IP			134.187	1387		255.255.255.240
PDPHOR02			134.187.	1387		255.255.255.240
FREE			134.187.	1387		255.255.255.240
FREE			134.187.	1387		255.255.255.240
FREE			134.187.	1387		255.255.255.240
FREE			134.187.	1387		255.255.255.240
RESERVED			134.187.	1387	**BROADCAST	255.255.255.240

State Government Customer upgrade RHEL 7 to RHEL 9- Oracle RAC

- Let's try another migration technique!

Oracle's Add Node procedure :

- Add a new node, then remove the old (rolling)
- Only need 4 new Ips (2 for each RAC node)
- One benefit migration could be done "rolling"
- It is supported to have RHEL 7 and RHEL 9 for a short period of time in a cluster.

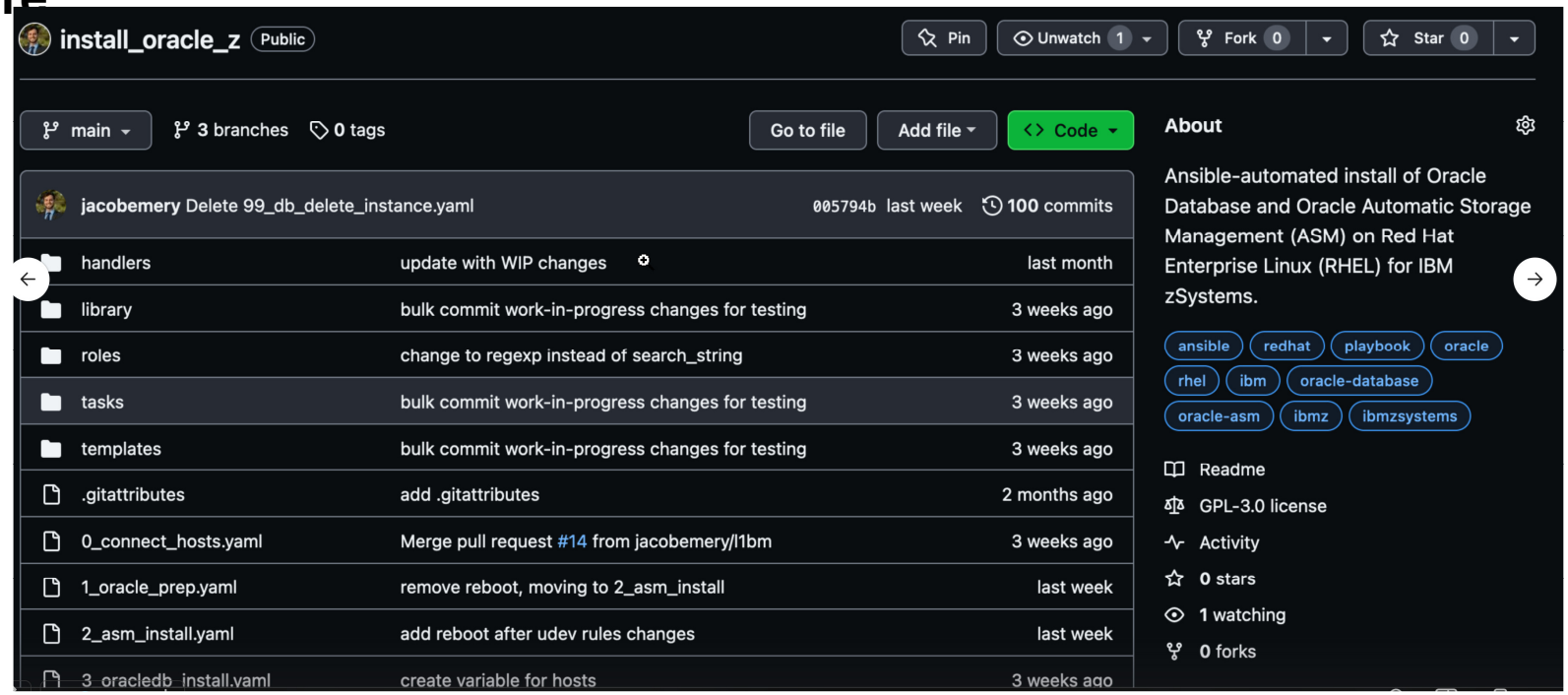


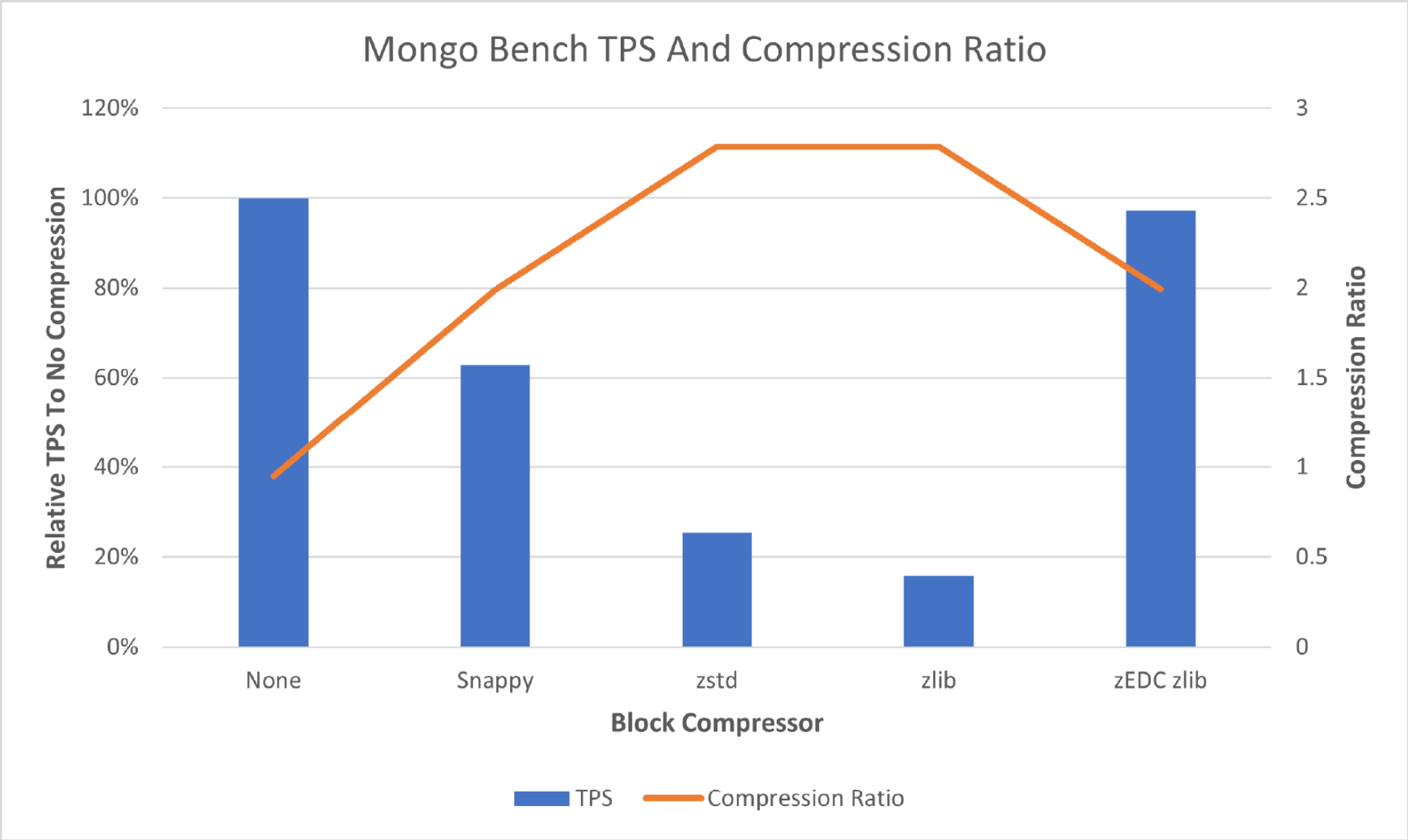
Finally this worked!!!! For real this time!

Ansible Automation

https://github.com/jacobemery/install_oracle_z

- Oracle Installation (using Golden Image files or applyRU, no xterm methodology)
- Spectrum Scale backup configurations
- Oracle RMAN backup configurations
- Automated server build (one node RAC) for Spectrum Scale immutable backup restore

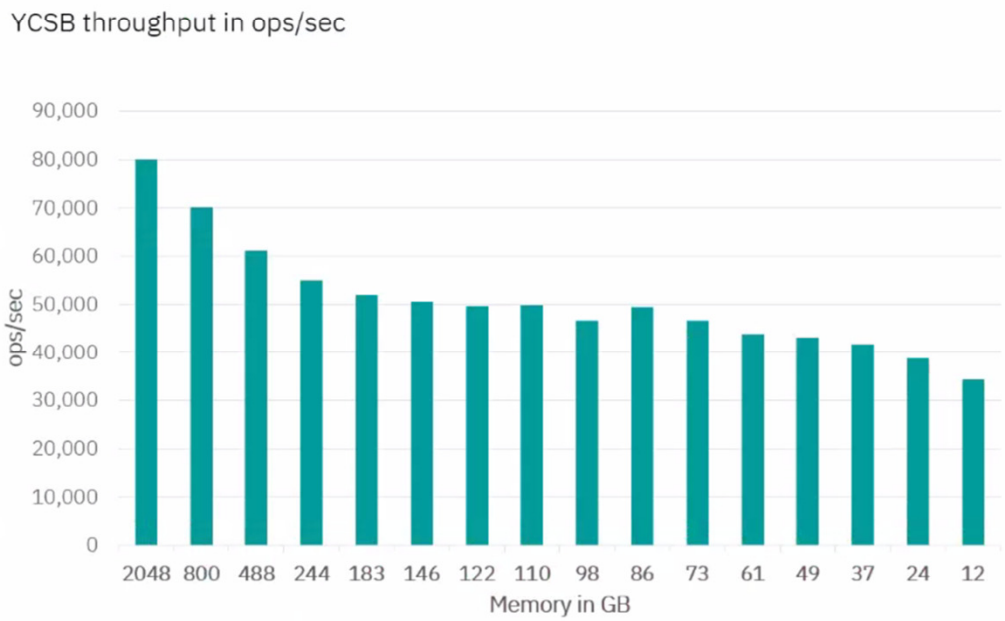



Mongo Bench TPS And Compression Ratio
120%
100%
80%
60%
40%
20%
0%
Relative TPS To No Compression
3
2.5
2
1.5
1
0.5
0
Compression Ratio
None
Snappy
zstd
zlib
zEDC zlib
Block Compressor
TPS
Compression Ratio

MongoDB Memory Sizing Testing on LinuxONE

Memory Scaling with MongoDB and YCSB - Throughput

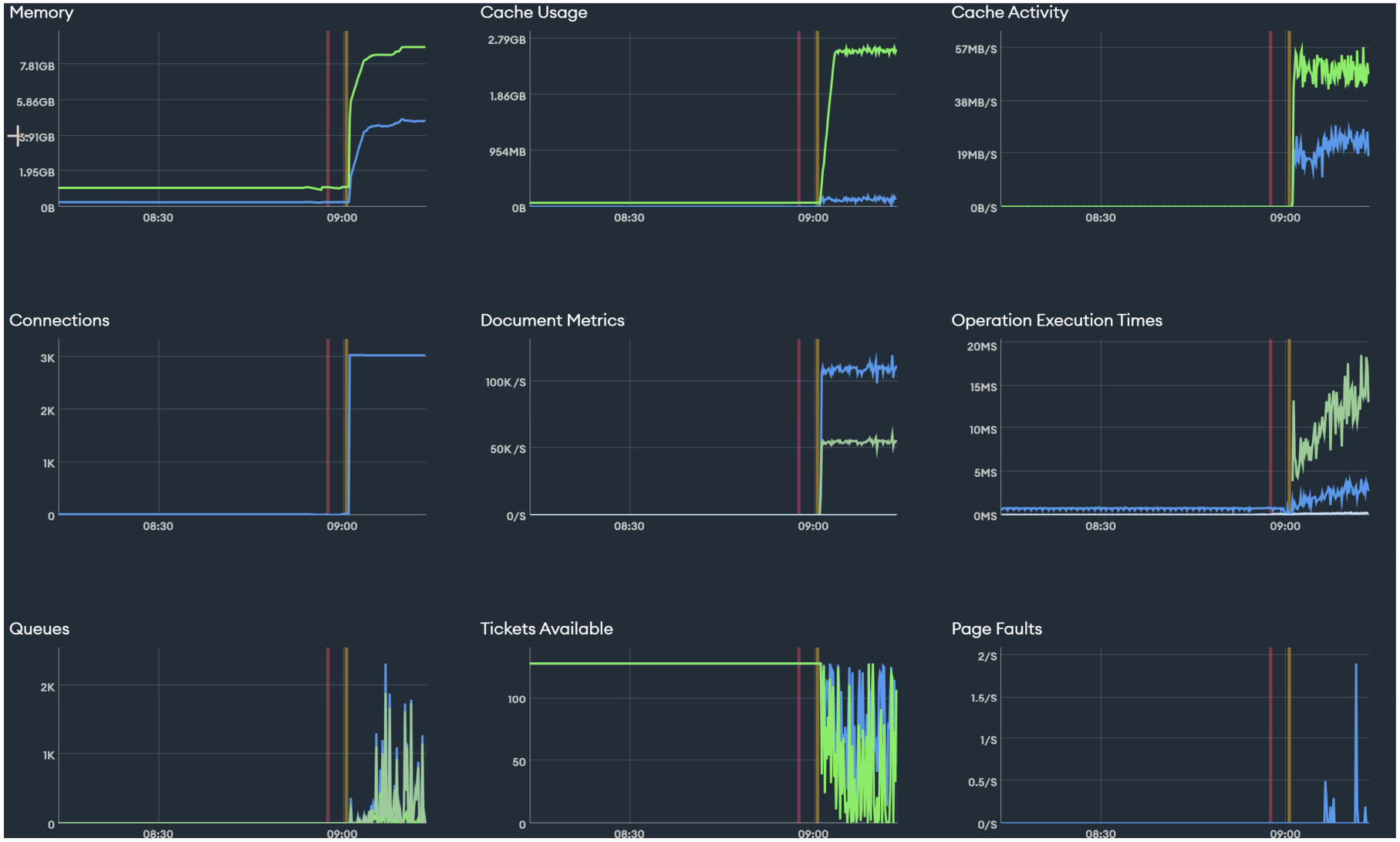
- Database setups require careful planning of memory-to-database size ratio
  - Typically Ranges from 1:1 (in-memory) to 1:10
- Varying memory from 12 GB to 2048 GB per instance while keeping the database size constant at 800 GB
- 2048 GB scenario: Entire database fits in memory with room for internal data structures.
- 12 GB scenario: Constant loading of database entries from storage
- Goal is to observe MongoDB performance and make recommendations for sizing Memory on IBM Z



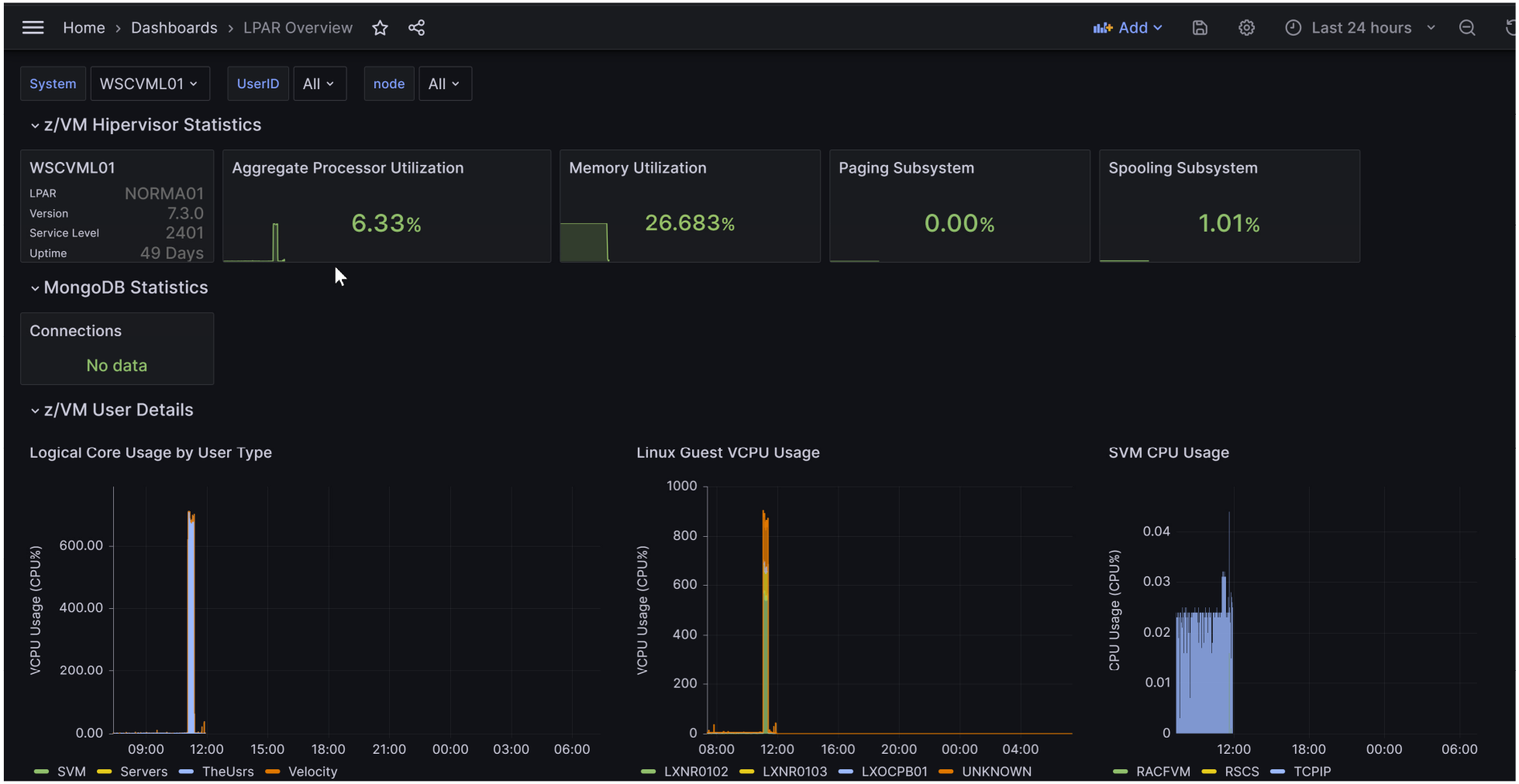
LinuxONE: MongoDB Monitoring with Grafana Demo-

Partnership with Velocity Software – see Kurt Acker's presentation for credits!

MongoDB Metrics on the Same Grafana Dashboard



z/VM & LPAR – Memory and CPU Data from Grafana Dashboard

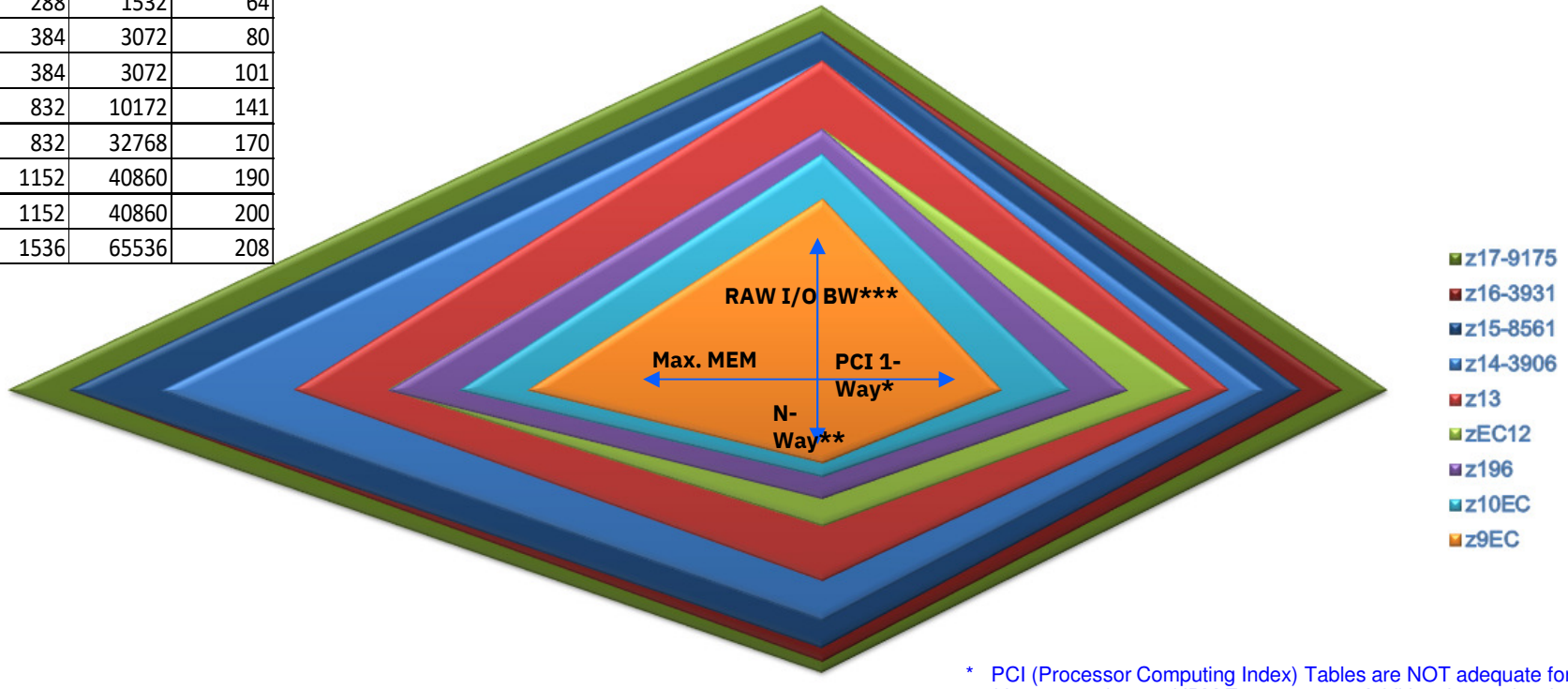


z/VM & LPAR – Memory and CPU Data from Grafana Dashboardx

IBM Z

LinuxONE 5 / z17 a Balanced system by design*

PCI 1-way	RAW I/O BW	Memory	N-way
581	172.8	512	54
920	288	1532	64
1202	384	3072	80
1514	384	3072	101
1695	832	10172	141
1870	832	32768	170
2055	1152	40860	190
2253	1152	40860	200
2477	1536	65536	208



* PCI (Processor Computing Index) Tables are NOT adequate for making comparisons of IBM Z processors. Additional capacity planning required
** Number of PU cores characterizable for customer use
*** Theoretical max. I/O Bandwidth (GB/s)

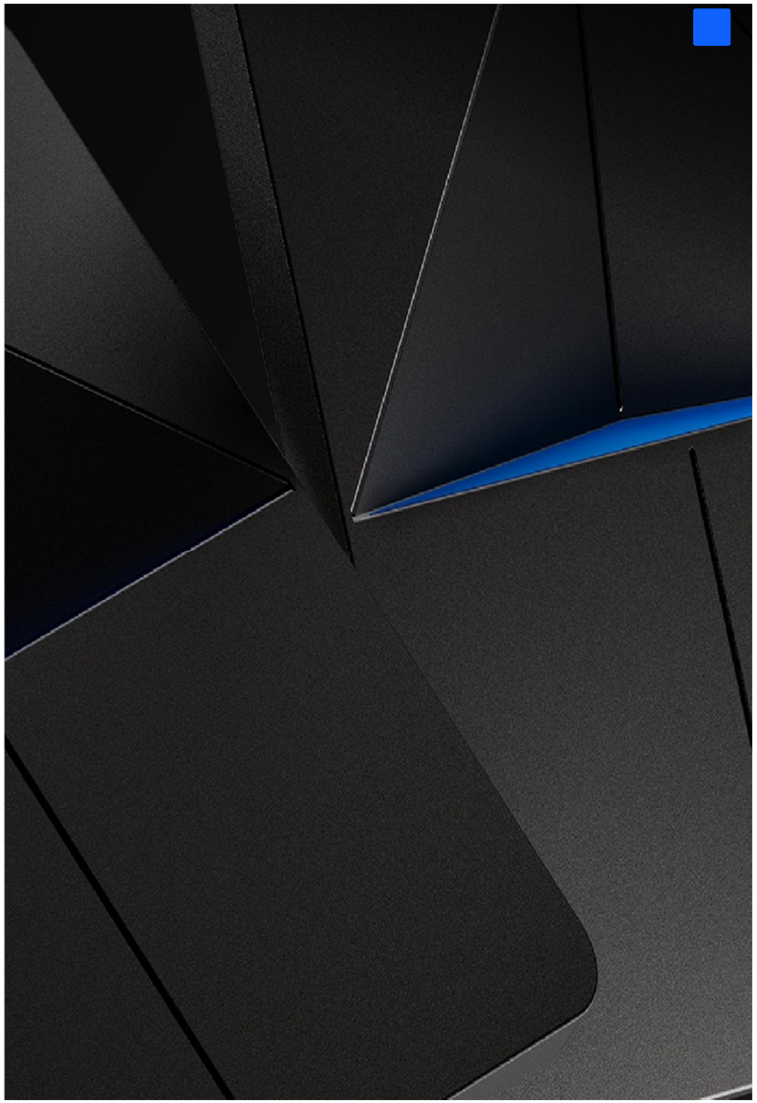
© 2024, 2025 IBM Corporation

IT Optimization Demo

During the demo, we will show:

1. “MegaBank” sample WebSphere Liberty application for a fictional bank, running on OpenShift on x86 environment
2. Visualize real-time metrics & topology with IBM Instana Observability
3. Observe environmental, application, and infrastructure metrics for MegaBank running on x86 with Grafana
4. Migrate application to OpenShift on IBM LinuxONE or OpenShift on Linux on IBM Z with **Red Hat Advanced Cluster Management**
5. In Grafana, compare metrics for x86 with IBM LinuxONE to see substantial energy savings and total cost of ownership (TCO) savings.

Special Thanks to Garrett Woodworth and Matt Mondics



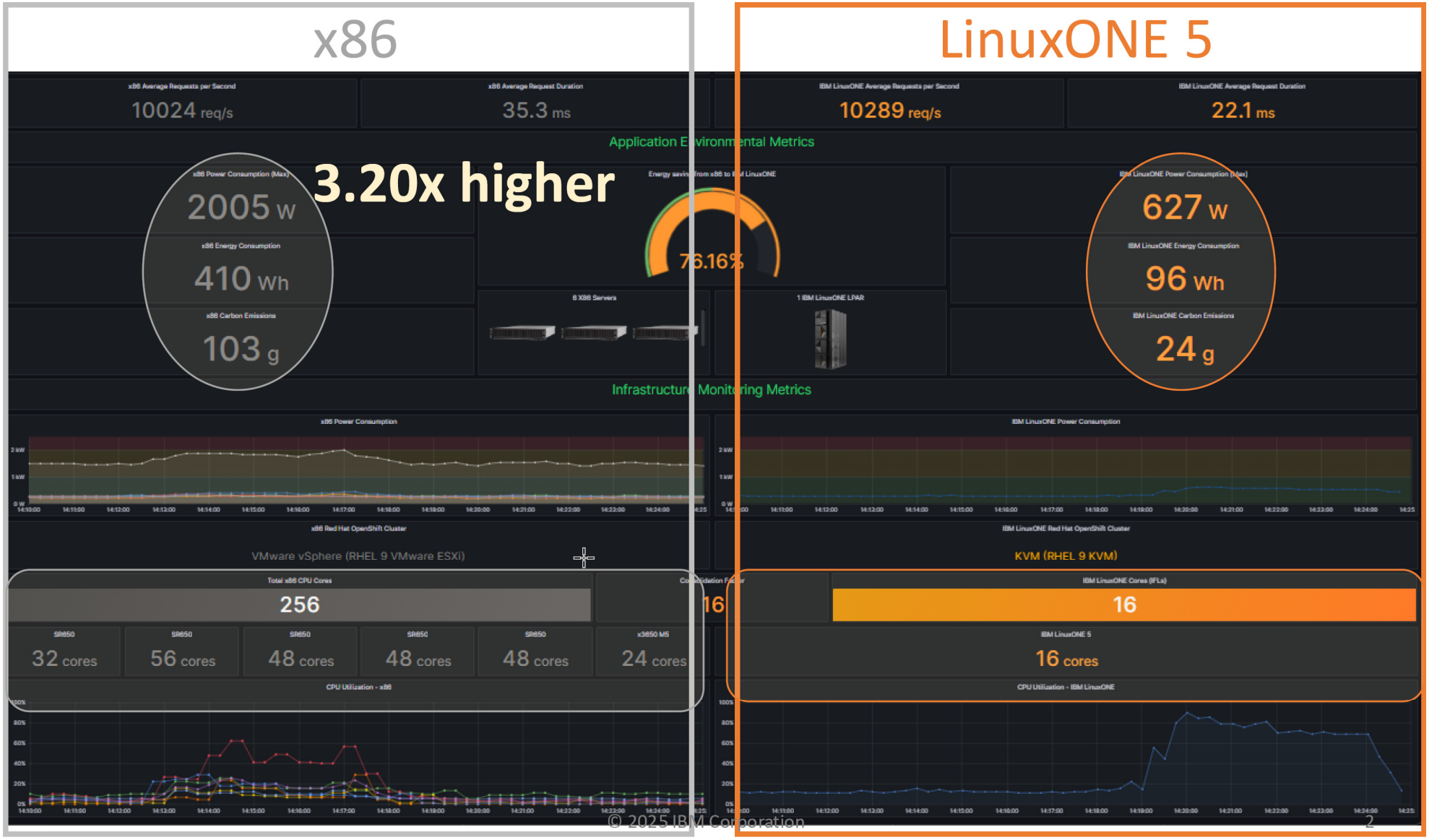
IBM internal tests:
Replacing x86 solution comprised of three-year-old servers running cloud-native, containerized workloads with IBM LinuxONE Emperor 5 can save up to $49.6 M and 76% in the total cost of ownership over 5 years[1]

TCO test results over 5 years:

49.6M savings

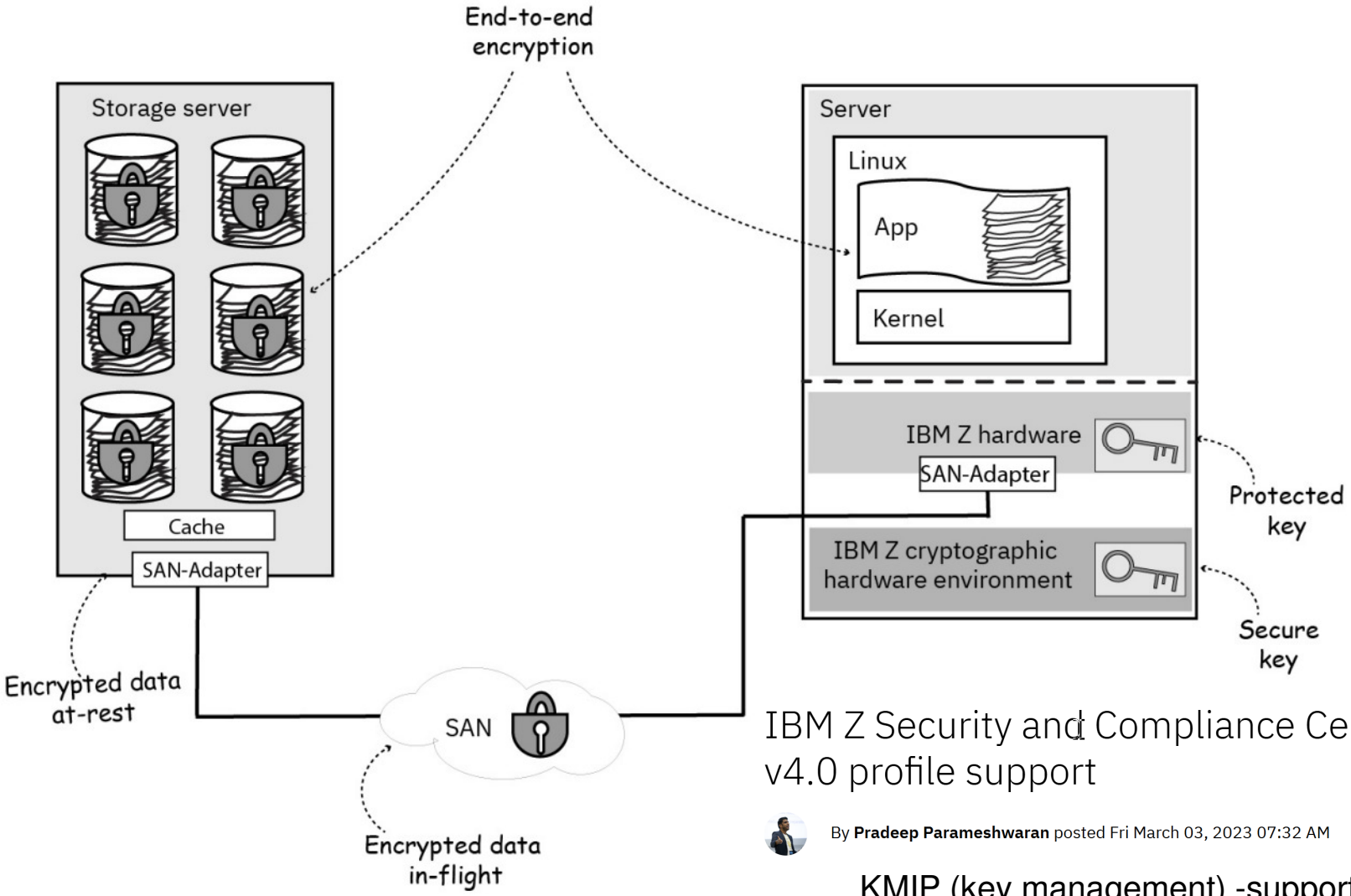
76% lower

Source: 1. Lorem source name
Source: 2. Lorem second source name



© 2025 IBM Corporation

PCI DSS (Payment Card Industry Data Security Standard) Compliance



IBM Z Security and Compliance Center : New PCI DSS v4.0 profile support

By **Pradeep Parameshwaran** posted Fri March 03, 2023 07:32 AM

KMIP (key management) -support

IBM WSC Sizing Tools:

zSCON Tool
Input:

Existing workloads:
- source make/model/GHz of Server

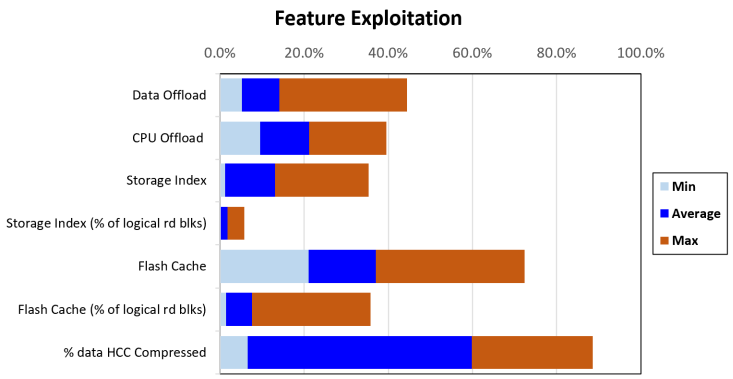
Server Hardware Description and Lookup Identification
Oracle SPARC T7-1 [2U] SPARC M7 32C 4.13GHz 4.13GHz (1ch/32co)

Output uses 3rd party QPI

LinuxONE 4 3931-LA1 IFL					
3931-7xx I1	1W IFL	3,018	299%	447%	557%
3931-7xx I2	2W IFL	5,925	153%	228%	284%
3931-7xx I3	3W IFL	8,791	103%	154%	192%
3931-7xx I4	4W IFL	11,616	78%	116%	145%
3931-7xx I5	5W IFL	14,402	63%	94%	117%
3931-7xx I6	6W IFL	17,140	53%	79%	99%
3931-7xx I7	7W IFL	19,829	46%	68%	85%

Exasizer: **Input:** Local AWR Database report

Exadata Feature Exploitation Scorecard



Target IBM Server Model:	zSeries	
Target IBM CPU Utilization:	100%	
Recommended QPI:	37235 - 37825	
Minimum Recommended Total Memory (GB)	768	Total memory should be spr

s a suggested starting point "All Flash" configuration. Alternatives involving HDDs or SDDs may be more c

Target IBM FlashSystem:	FlashSystem 9500	
ASM Redundancy (IBM):	EXTERNAL	
# of FlashSystem 9500 Systems:	1	
Module Type (TB):	19.2	
# Modules (per System):	18	
DRAID capacity (TiB) (per System):	245.1	
DRAID capacity (TiB) (total)	245.1	1.07x of required (229.4)
Est. max tot throughput capacity (GB/sec):	55.4	1.38x of required (40.0)
Est. max total configuration 8K IOPS Capacity:	2,404,471	38.8x of required (61,976.3)

Also recommend working with IBM IT economics team for TCO analysis

z/VM Large Guest Memory Support Considerations (Early Adopters)

- z/VM supported limit: 2TB per LPAR, 4 TB with z/VM 7.2 + VM66173
- z/VM virtual machine size supported: 1 TB and **now 2 TB**
- Maximum Oracle SGA (19.19 version) **1813.5 GB** (need memory for other functions) Oracle patch -> **34168505**

Live Guest Relocation

Relocating a large guest may take an extremely long time, not supported for Oracle

Reset Time

A guest reset (e.g., performed as part of LOGOFF or re-IPL processing) may take an extremely long time. Introduced minor changes that reduce the interval between guest LOGOFF initiation and subsequent re-LOGON

Helpful Commands and Tools

VMDUMP of large guest may run slowly to be viable and in any event is limited to dumping memory only up to 512GB.

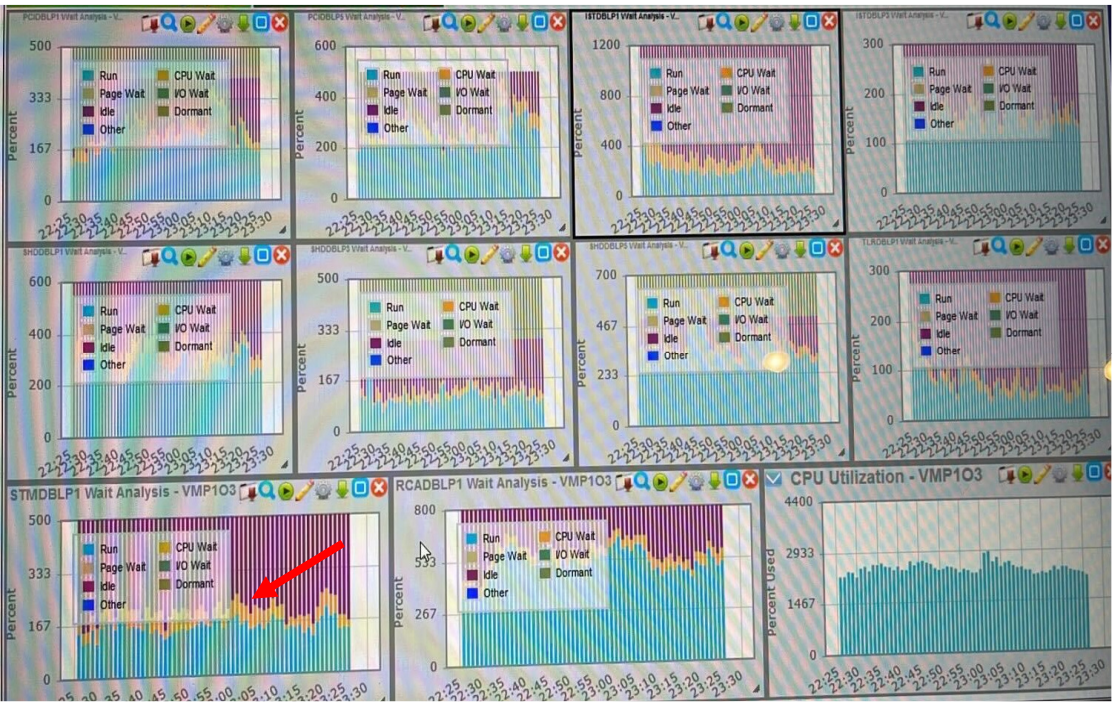
Small Guests

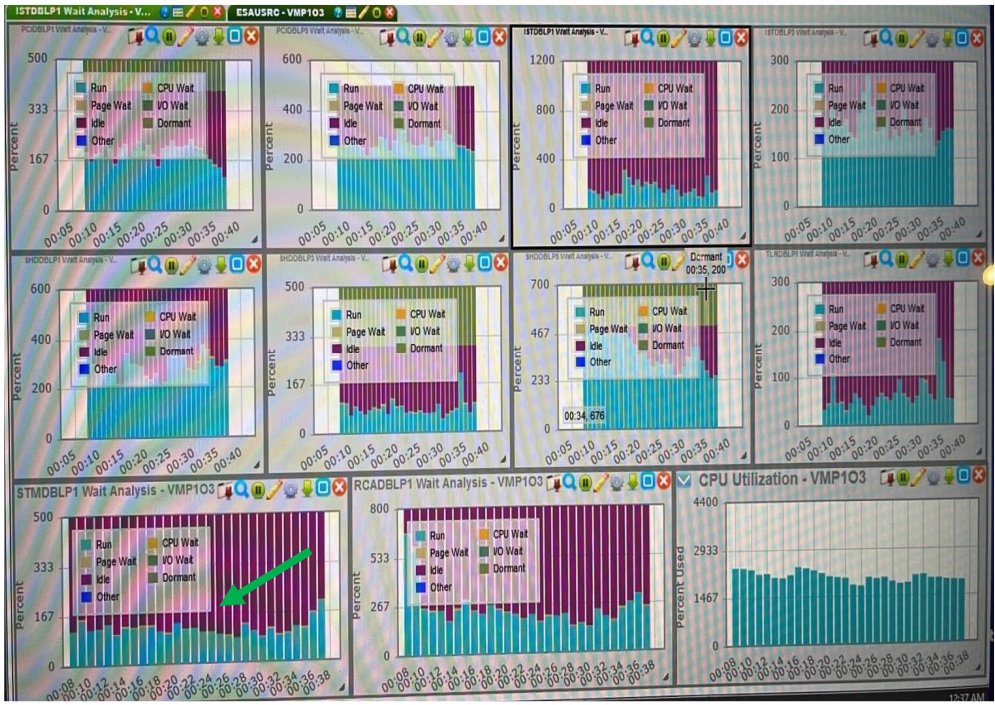
Smaller guest may be disadvantaged by larger one if memory is overcommitted.

Source: z/VM Limits https://www.vm.ibm.com/memman/gt1guest.html

Large Bank Oracle Experiences: Tuning CPU Requirements:

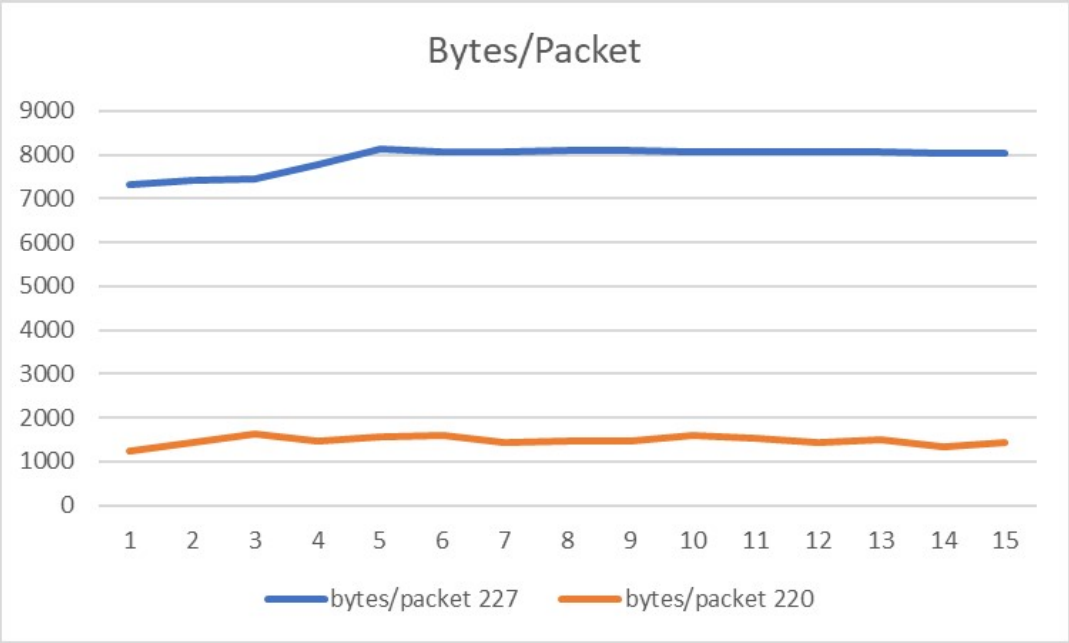
1. Make sure each guest has enough virtual CPUs to handle it's peak load (suggest cpuplugd to reduce / add cpu)
2. Check with z/VM support about changing from z/VM MODLEVEL 1 to MODLEVEL 0 is set (CP **SET SYSCONTROL DISPATCH MODLEVEL 0** Reduces CPU wait in **yellow** below)
3. Fine-tune share settings of Linux guests





Separating Backup Traffic from User Network

- Backup data rate went from 419 mB/sec to 904 mB/sec. Increased data rate by 116%.
- Bytes/pkt went from 1518 to 8218 - guest and host are more efficient
- CP CPU time per mB transmitted went from 784 us to 380 us.
- Total CPU time per mB transmitted went from 10250 us to 4368 us. 57% reduction in CPU use per mB sent. Host (CP) is operating much more efficiently.



Spectrum Scale 5.0 with Oracle Backups Test Case:

Test Case:

Seven snapshots representing everyday. Sunday, a level-0 (FULL) backup is taken. Monday to Saturday incremental backup and merge is done.

Advantages backup scheme:

- Daily backup window is reduced. Only changed blocks from previous day are backed up.
- After snapshot, incremental backups are merged back with the level-0 backup. After the merge, we have full backup.
- Changed blocks since previous incremental backup, are pushed into snapshot (**Block Push to snap**). This means we have Sundays and Mondays level-0 backup. For a week, we have seven FULL backups.
- For restore, since we always have a current level-0 backup copy, no need to merge, apply lots of archivelogs during restore (faster RTO)
- Recovery time fast and customers can meet recovery time objective (RTO) goals.
- The workload we ran against the database inserts one million rows; updates one million rows, update indexes.
- The table below shows the amount of disk storage saved, **The original DB Size is 118 GB**

SnapShots	Backup Size	Incr Backup Size	Block Push to snap
0	118 GB	None	None
1 (snp_0717@16H56M)	125 GB	50 GB	6.8 GB
2 (snp_071717H19M @)	131 GB	63 GB	6.1 GB
3 (snp_0717@17H35M)	137 GB	54 GB	6.1 GB
4 (snp_0717@18H01M)	144 GB	60 GB	6.2 GB
5 (snp_0717@18H19M)	150 GB	63 GB	6.2 GB
6 (snp_0717@19H01M)	157 GB	62 GB	6.3 GB
7 (snp_0717@19H29M)	164 GB	60 GB	6.1 GB