

Good practices: performance, optimization, and configuration management for z/VM and Linux on IBM Z and IBM LinuxONE

2019 VM Workshop

<http://www.vmworkshop.org>

Paul Novák

z/VM and Linux on Z and LinuxONE
Washington Systems Center (ATS)
Endicott, New York





Special Notices and Trademarks

Special Notices

This presentation reflects the IBM Advanced Technical Skills organizations' understanding of the technical topic. It was produced and reviewed by the members of the IBM Advanced Technical Skills organization. This document is presented "As-Is" and IBM does not assume responsibility for the statements expressed herein. It reflects the opinions of the IBM Advanced Technical Skills organization. These opinions are based on the author's experiences. If you have questions about the contents of this document, please contact the author at linuxats@us.ibm.com

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.

Any and 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. References in this document to IBM products or services do not imply that IBM intends to make them available in every country. 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. Any proposed use of claims in this presentation outside of the United States must be reviewed by local IBM country counsel prior to such use.

The information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice. Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM retains the title to the copyright in this paper, as well as the copyright in all underlying works. IBM retains the right to make derivative works and to republish and distribute this paper to whomever it chooses in any way it chooses.

Trademarks

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

IBM, the IBM logo, DB2, Redbooks, Tivoli Enterprise Console, WebSphere, z/OS, System z, z/VM.
A full list of U.S. trademarks owned by IBM may be found at <http://www.ibm.com/legal/copytrade.shtml>.

Microsoft, Windows, Windows NT, Internet Explorer, and the Windows logo are registered trademarks of Microsoft Corporation in the United States and/or other countries.

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

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.
UNIX is a registered trademark in the United States and other countries licensed exclusively through The Open Group.

Intel and Pentium are registered trademarks and MMX, Pentium II Xeon and Pentium III Xeon are trademarks of Intel Corporation in the United States and/or other countries.

Other company, product and service names may be trademarks or service marks of others.

Topics Covered

- Silos, towers, and other operational segments
- Understanding order of magnitude
- Maintenance & Service
- Processors
- Dispatching
- Linux Runlevels & systemd equivalents
- Unnecessary baggage
- z/VM Memory Configuration
- Linux Virtual Memory Sizing
- z/VM Paging Subsystem
- Linux Swap Space
- Disks
- Linux filesystem types and options
- z/VM Dump & Spool Space
- Networking Configuration Options
- HiperSockets Bridge - Cross CEC
- MTU Sizes, Inbound QDIO Buffer, Checksums
- SYSCTL Settings
- SSI Cluster Configuration
- CTC Subchannel Addressing
- VMSSI Live Guest Relo & MAC Addressing
- Suggestions
- Assessments, Sizing, & Capacity Planning
- Installation, Planning, & Administration
- Reference materials



With gratitude...

The following people contributed to this presentation, either directly or indirectly:

- Bruce Hayden
- Richard Lewis
- Brian Wade
- Bill Bitner
- Alan Altmark
- Jim Switzer
- Tracy Dean
- John Franciscovich
- Patty Rando
- Edi Lopes Alves
- Sandy Bulson
- Romney White
- Brian Hugenbruch
- Rob Van der Heij
- Tung-Sing Chong

Silos, Towers, Service Drawers, Competency Bins & other operational segmentation?

Collaborate!

The best and worst thing about z/VM is that you can share resources.

z/VM System Programmers and Linux Administrators may not be in the same organization and are often split apart.

If your shop is split like this, do not think of this from the perspective of “we versus them”.

Collaborate together and learn from each other!

Actively champion and support each other; find ways to automate, optimize, and attract new workloads to the platform.

Brass, woodwinds, strings, and percussion all play in the same symphony!

Experience-based viewpoints

Linux.
Unix.
Windows.
z/OS.
z/TPF.
z/VM.
z/VSE.

It is easy to over-allocate resources

Monitoring is important to examine resource usage

- hardware
- hipervisor
- virtual machines

Real-time and historical metrics demonstrate peak periods as well as average runtimes.

Be sure you are collecting z/VM MONITOR records!
There is no good reason not to.

Hardware matters!

z/VM is actual virtualization plus hardware assistance

- Hardware instructions match the physical hardware being used
- Hardware offers bespoke qualities to assist with optimizing the hipervisor
- Nearly all other hypervisor solutions are virtualization with binary translations; the hypervisor is used to translate operating system functions in a way that fully abstracts the guest OS from the underlying hardware.

Cross-platform virtualization increases challenges

- Virtualization experience on another platform does not necessarily translate to others.
- Don't assume.

“ It doesn't matter what you're doing;
you have to know what you're doing.

Steve Gracin
IBM ATS

Rough order of magnitude



Performance tuning and problem determination is typically done in the wrong order.

When done in the correct order:

- Tuning: Work MICRO to MACRO
 1. Application
 2. How implemented in the environment
 3. Middleware / Software
 4. Guest OS
 5. Hypervisor
 6. Hardware
- PD: Work BACK to FRONT
 1. Application (and databases if in use)
 2. App Server (Software)
 3. App Server's Hypervisor
 4. App Server (Hardware)
 5. Proxy / Network

Why order of magnitude is important

- No amount of time or effort spent on items lower in the list will ever fully compensate for problems or shortcomings of items above them!
- In certain (most) cases, trying to resolve or minimize a problem via changes to one of the entries further down the list will make it exponentially worse and/or create other problems, sometimes even more serious.

Application design and implementation combined represent roughly 85% of where most problems originate

Order of Magnitude

1. Architecture / Design

- User Experience Design (UxD)
CRITICAL but often ignored or overlooked
- DB use when unnecessary
- I/O blocks
- Poor performing SQL
- Fault intolerance

4. Guest Operating System

- Kernel parms
- Oversized logs
- Volumes near/at capacity
- NIC/TCP stack parms
- Scheduling method
- I/O methodology

2. Implementation

- J2EE app burdening JRE with requests for static objects from filesystem(s)
- No reverse caching proxy
- No compression
- No proper load balancing
- No keepalive parms set
- Poorly formed HTML or XHTML

5. Hypervisor

- User Directory parms
- Buffer pools
- Dispatching
- Over allocation of RAM to guests
- Improperly configured VSWITCH or QDIO NIC parms

3. Middleware

- Service/Fix level
- Version/Release level
- DBs indices
- Datasources and connectors
- .conf and/or .ini parms

6. Frame / Hardware

- Microcode
- Millicode
- Firmware
- I/O definitions
- Cabling
- Hardware fault

**Stay Updated.
Apply Service.
Install Fixes.
Don't dig yourself a hole.**

Apply z/VM Service.

z/VM Service basically means system software service in the form of updates, patches, enhancements, new features, and the like.

Regular Service is published in a bundle called a Recommended Service Upgrade (RSU)

- Similar in concept to an AIX™ ML/TL, Linux Service Pack, or Windows® Cumulative Fixpack
- An RSU is typically released every 3 to 6 months and contains cumulative service and includes all pre and co-requisites. No guesswork!
- Includes service for all integrated components and pre-installed program products
- Includes service required by most customer installations and is pre-tested by z/VM development to help ensure a quick, smooth, successful result!
- Easy to install, just as easy to back-out if necessary
 - Install: SERVICE and PUT2PROD
 - Back-out: SAPL – IPL from CLOAD MODULE then issue VMSES - VMFREM

More about z/VM Service

Your governance model should include applying z/VM service (updates) on a regularly scheduled basis.

- **Plan for quarterly service**
 - **Install RSUs or PTFs if available**
 - Install SPEs (New function APARs) if desired
- Systems with VMSSI and LGR can utilize these features to reduce cycle time.
Especially attractive now with z/VM 7.1!
- Use the CP command **QUERY CPSERVICE** to display the service table of all installed service.

Visit the following pages on ibm.com for the full details and subscribe:

- www.ibm.com/vm/service/news
- www.ibm.com/vm/service/redalert
- www.ibm.com/vm/service/rsu

Use the subscribe feature (“notify me”) to be automatically notified of any changes for all three pages, especially red alerts.

IBM Products Services Industries Developers Support Careers

z/VM Service: Red Alerts

This page is intended to provide information about potential issues that may occur when the end of service level reaches the end of service, we intend to remove the information from this page.

Date	Description
22 May 2018	Effective May 24, 2018, red alerts for z/VM products.
30 Mar 2018	APAR VM65943 is in error. Apply PTF for APAR VM66026.
13 Dec 2017	Avoid IPL Failures - Remind you to apply PTF for APAR VM66026.
10 July 2017	APAR VM66026 is in error. Apply PTF for APAR VM66026.

- **May 22, 2018**
Changes to uploading documentation for z/VM products.

USERS AFFECTED:
All users that upload problem related documentation.

DESCRIPTION:
Effective May 24, 2018, for GDPR compliance, all users must provide a valid IBM Support File Transfer ID and password (token), and a valid IBM ID.

RECOMMENDED ACTIONS:

1. Links to create an IBM Support File Transfer ID. If you already have an IBM ID follow these instructions.
2. Ensure you have secure transport protocols in place.
3. Change any existing EXECs, JCL, jobs, or tools to use the new password.

TIMING:
Authenticated log-in requiring an IBM Support File Transfer ID.

z/VM

News

About z/VM

Events calendar

Products and features

Downloads

Technical resources

Library

How to buy

Install

Service

Education

Site map

Site search

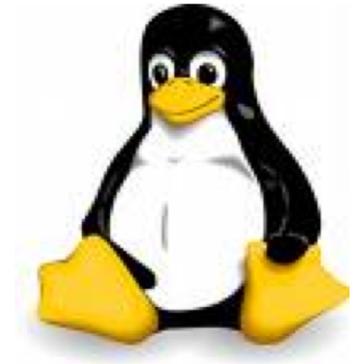
Printer-friendly

Notify me

Contact z/VM

Stay Updated – Apply Linux Updates!

- While there is occasionally some risk involved with running at the cutting edge, there is much, much greater risk when you fall behind!
- Staying down-level can expose you to zero-day vulnerabilities.
- Recommend using the most current release or version for your Linux distribution of choice. Ensure it has been tested and officially supports all required middleware and/or applications.
- Distribution service pack updates include:
 - Fixes and Security patches
 - Performance enhancements
 - New functionality
- Leverage native utilities in your Linux distribution to stay current:
 - SLES: YaST Online Update (YOU) or Zyp UP
 - RedHat: Yum update or use the RedHat Network (RHN)
 - Debian & Ubuntu: Advanced Packaging Tool (APT)



Stay Updated – Apply Linux Updates!

- Newer kernels typically include support for newer features, such as enhancements made for containerization.
- Check your kernel level easily from your Linux shell by issuing “uname -r”

Processors (Cores): Physical, Logical, and Virtual

- Physical: Actual number of purchased cores / engines (CPs, IFLs, etc...)
- Logical: Number of cores / engines defined to an LPAR in the PR/SM activation profile or DPM server config.
 - Systems development recommends a max ratio of 4 logical to 1 physical in NON-PRODUCTION.
 - Real life experience is that 3:1 is about the max.
 - Always define reserved processors in z/VM LPAR Activation Profile to enable non-disruptive activation of additional engines in the future.
 - Use the command `QUERY PROCESSORS` from a class A,B,C, or E virtual machine to see details.
- Virtual: Number of cores / engines defined to a z/VM Virtual Machine (USER or IDENTITY).
 - Do not define more virtual processors than there are logical processors in the LPAR.
- Keep in mind engine types (specifically, CPs and IFLs) do not automatically mix together inside of a z/VM LPAR or its virtual machines.
 - If you create an LPAR in “z/VM Mode” with both CPs and IFLs, z/VM will consider the CPs to be the IPL Processor Type, and all processor dispatching will use ONLY the CPs by default.

Processors – Hiperdispatch and EII

- HiperDispatch
 - Provides more efficient utilization of CPU hardware resources for dispatched work.
 - Strengthens affinity between where work is dispatched and where associated data exists, avoiding delays to retrieve
 - Recommend using fewer, larger LPARs with HiperDispatch for better performance and improved throughput
 - Use your performance monitoring tool to watch for LPAR logical CPUs being over or under configured compared to the partition's entitlement.
 - PerfKit FCX306 (LSHARACT)
- Environment Information Interface
 - Provides ability to obtain CPU resource information and to enable virtual machines to understand the characteristics of the environment in which they are running
 - Information is provided at many levels - the machine, logical partition, CPU pool, virtual machine.

Processors –Pooling

Resource (CPU) pools

- Define and limit resources a group of z/VM guests is allowed to consume as a whole among the group
- Set a boundary on lower-priority workload guests when grouped together as a pool
- Recommend defining pools to help ensure licensing compliance and better control priorities.

In z/VM PerfKit, check reports:

- FCX324 (CPU Pool Menu)
- FCX308 (CPU Pool Configuration)
- FCX309 (CPU Pool Activity Data)

Processors - Dynamic Simultaneous Multithreading

Dynamic SMT added to change the number of active threads per core without a system outage

- z/VM 6.4 allows one to dynamically change the number of active threads per core when SMT has been enabled in the SYSTEM CONFIG file.
 - **MULTITHREADING ENABLE TYPE ALL 1**
 - System configuration file statement enables SMT-1 (1 thread per core)
- Potential capacity gains going from SMT-1 to SMT-2: (one to two threads per core) can now be achieved dynamically
- Downgrade from SMT-2 to SMT-1 in extremely rare case that it is not optimal for workloads (response time concerns greater than capacity gains)
 - Once z/VM has started, toggle between 1 and 2 threads via CP command **SET MT TYPE ALL 2** and query status and information with CP commands **QUERY MULTITHREAD** and **INDICATE MULTITHREAD**
- Requires:
 - IPL with SMT enabled, but can vary active threads per core
 - Requires z13, z13s, LinuxONE Emperor or LinuxONE Rockhopperor newer CPC.
- If more than 32 cores are required per LPAR, can **not** use SMT -- even with one active thread per core.
- **STRONGLY recommend you read *CPU Utilization* pages on the Additional Information slide.**

Guest Virtual Processors

Be aware:

- Of the maximum per virtual machine based on VM version and your hardware.
- That various guest operating systems and workloads scale differently

Recommendations:

- Configure number of virtual processors per guest for peak workload, no more
- Do not define more virtual processors to a guest than logical processors defined to the z/VM LPAR
- Watch data from your performance monitoring tools
 - In z/VM PerfKit, check reports:
 - » FCX112 (User)
 - » FCX104 (Privileged Operations)
 - » FCX114 (USTAT / Wait State Analysis by User)

Processors – Understanding utilization

PR/SM accumulators:

- Time used by logical PUs, time logical PUs induce in PR/SM, and time PR/SM uses for itself.
- Utilization percentages calculated from these shown in Perfkit FCX126 LPAR and FCX202 LPARLOG reports.

z/VM accumulators -- for each logical PU:

- Time used by guests.
 - Time induced in the Control Program.
 - Time the Control Program uses for itself.
 - Time spent waiting.
 - Time spent parked.
-
- Utilization percentages calculated from these accumulators shown in many Perfkit reports. Recommend checking FCX126 LPAR and FCX225 SYSSUMLG.

Processors – Understanding utilization

- Guests' individual CPU utilization numbers come out in a few places, notably FCX112 USER and FCX162 USERLOG. As mentioned previously, FCX225 and FCX239 report the system T/V ratio, while FCX112 (USER) and FCX162 (USERLOG) report the T/V ratio for individual virtual machines.
- These two reports comment on two phenomena: the individual guest's own consumption, and the consumption the guest's actions induced in the Control Program.
- Almost all the time, all of these percentages are of a whole physical processor, and "100%" means "a whole physical engine's worth". The exceptions to this are FCX126 LPAR %LogId and FCX100 CPU %LOGLD.
- Again, please read the URLs under *CPU Utilization* on the Additional Information slide.

Guest Privilege Class

Most Linux guests do not require anything more than privilege class “G”.

Not a performance issue – absolutely possible to be a security and stability issue though:

- A privileged Linux guest could shutdown the z/VM system.
- A privileged Linux guest could compromise other guests or the entire z/VM system.

Not limited to the VM 3215 Linux console session either

- Linux vmcp command (October 2005 2.6 kernel stream)

Always restrict privileges to the minimum required.

- If you do grant additional classes for a valid purpose, audit your system during your quarterly service routines to ensure it is still necessary and revoke if not!

Guest Virtual Processors – What to watch for

- In z/VM PerfKit
- Report FCX225 (SYSSUMLG) and FCX239 (PROCSUM)
 - T/V ratio is a good way to gauge overhead. FCX225 and FCX239 report the system T/V ratio, while FCX112 (USER) and FCX162 (USERLOG) report the T/V ratio for individual virtual machines.
- Report FCX100 (CPU)
 - FCX100 look for Total:Virtual ratio (%CPU:%EMU). The closer to 1.00 the better.
- Report FCX104 (Privileged Operations)
 - High diagnose x'44' or x'9C' rates may be an indication of too many virtual processors. Watch for:
 - x'44' > 50,000/sec
 - x'9C' > 5,000/sec
- Report FCX114 (USTAT / Wait State Analysis by User)
 - %CPU wait should be low

Adjust the ‘SHARE’ of Virtual MP Machines

- The default SHARE setting for all virtual machines is “Relative 100”:
 - VM dispatches users by VMDBK
 - There is one VMDBK per virtual processor defined
 - A users SHARE setting is divided among the defined virtual processors
- Recommend resetting the SHARE of Virtual MP Machines as a starting point:
 - Set SHARE RELATIVE (100 * number of virtual CPUs defined)
 - This creates an initially “level playing field”
- Adjust SHARE of guest virtual machines from this point, as required:
 - Increase SHARE to prioritize
 - Decrease SHARE to penalize
- A virtual machines SHARE only comes into play when there is contention for resources, primarily CPU
- As of 6.4, CP now honors SHARE settings more accurately than previous releases.
For more details, see the SRP article listed at the end of the presentation.



More about shares

- Priorities of users (virtual CPUs) are controlled by the SHARE setting. There are 2 types:
 - Relative - valid range 1-10000
 - Absolute - valid range 0.1-100%
- Virtual machine share setting is divided by the number of virtual CPUs and assigned to each virtual CPU.
 - Minimum relative share per CPU is 1
- Set in the directory or with the `CP SET SHARE {userid}` command
 - Show the current setting with `CP QUERY SHARE {userid}`
- Absolute share users given resource first
 - If sum of absolute shares > 99%, normalized to 99%
 - Leftover share (minimum 1%) available for relative share users
- For relative share users - actual share depends on total relative shares of all virtual CPUs
 - A virtual CPU gets (vCPU relative share / total relative shares) %

Limiting shares

- There are two kinds of limits which can be set:
 - LIMITSOFT
 - LIMITHARD
- With the changes made to dispatching and SRM, if you're using either of these you may want to investigate if these are indeed producing the results you might be expecting.

Quick Dispatch

- Setting QUICKDSP traditionally would have:
 - Bypassed System Resource Management controls
 - Placed a virtual machine directly into the dispatch list
 - Made a virtual machine exempt from being placed in an eligible list
- QUICKDSP should be reserved for use only with:
 - Service Virtual Machines performing critical functions on behalf of other guests such as TCP/IP, RACF, or DTCVSW#
 - Selected high-priority or key production guests running things like DB2, WebSEAL, Edge Load Balancer, DynaCache, IBM Directory Integrator, etc.
 - If you are not already using QUICKDSP for a virtual machine, you should only do so at the recommendation of IBM VM Support.
- Interested in learning more about how SRM used to function? See the reference materials slides at the end for an excellent detailed explanation by Malcolm Beattie (IBM) .



Linux runlevel and target states

- Linux has different modes of operation.
 - Under system-v they are **runlevels**
 - Under systemd, they are **target states**

Linux runlevels

Typical system-v runlevels:

- 0 - Halt the system
- 1 - Single-user mode
- 2 - Multi-user mode (without networking)
- 3 - Multi-user mode
- 5 - Multi-user mode (display manager, GUI)
- 6 - Reboot the system

When you boot Linux, it will initialize to a pre-defined default runlevel (usually 3 or 5).

- Most desktop Linux systems boot into RL 5 by default; users are presented with a GUI.
- Most server Linux systems boot into RL 3 by default; users are presented with a CLI.

Query the current runlevel with the command: `runlevel`

You change between them with the command: `telinit {target #}` e.g. `telinit 3`

Linux target states

Typical systemd target states:

- `poweroff.target` (or `runlevel0.target`) - Halt the system
- `rescue.target` (or `runlevel1.target`) - Single-user mode
- `multi-user.target` (or `runlevel3.target`) - Multi-user mode
- `graphical.target` (or `runlevel5.target`) - Multi-user mode (display manager, GUI)
- `reboot.target` (or `runlevel6.target`) - Reboot the system
- `emergency.target`

Linux will initialize to a pre-defined default boot target (usually `multi-user` or `graphical`).

- Most desktop Linux systems boot into `graphical.target` by default; users are presented with a GUI.
- Most server Linux systems boot into `multi-user.target` by default; users are presented with a CLI.

Check the default target with the command `systemctl get-default`

Set the default target with the command `systemctl set-default {choice}.target`

Change the running target level with the command `systemctl isolate {choice}.target`

Linux runlevel and target considerations

- Recommendation is RL 3 or multi-user.target for Linux guests of z/VM
- X services are very costly in terms of CPU cycles and RAM, regardless of computing platform
- When improperly configured, graphical subsystems can pose a potential risk vector such as Linux GDDM listening on all interfaces
- **Production systems should never have full graphics enabled**
- When necessary, use a lightweight X-server like VNC server instead of full GUI desktop
 - Ensure that it is fully disabled once you are done
 - Consider a cron job to notify of any X services running longer than a day or two

Unnecessary Guest Virtual Machines

- Shutting down unnecessary guest virtual machines helps to improve the overall performance of the system:
 - Linux guest virtual machines almost never go dormant
- Logoff:
 - Golden images used for cloning
 - Test machines and “sand boxes”
- Suspend:
 - Production guests not necessary during POC testing or benchmarking of another application or workload
 - See Chapter 39, Suspending and resuming Linux in the current Linux on System z Device Drivers, Features, and Commands Manual
- Reduce “SHARE” setting for virtual machines running lower priority workloads



Unnecessary Services/Applications

There are a number of services in Linux that get started at boot depending on:

- Distribution
- Linux kernel level/version
- Installed software packages

Shutting down unnecessary services and unused applications helps to improve the overall performance of the system

- Status of services can be queried/changed with the “chkconfig” command

The cron daemon is useful for scheduling events to be kicked off automatically at a specific time or at regular intervals

- Running many guests with identical schedules can cause CPU spikes and stress the z/VM paging subsystem:
 - Remove unnecessary events from cron
 - Consider switching to anacron for events which must remain
 - Stagger scheduled kick-off time of events – especially for high I/O tasks like backups

z/VM Memory Configuration – Central Storage

Max 2TB (Initial plus Reserved) – Set in HMC LPAR Activation Profile

Plan on a virtual to real (V:R) memory ratio in the range of 1.5:1 to 3:1

- Production systems will typically be closer to the low end of range
- Development/Test systems may be able to push the upper end of range

STANDBY memory can be added and removed dynamically to central storage:

- Storage must be defined as “RESERVED” in the LPAR Activation Profile
- Under z/VM 7.1, memory can be added and removed dynamically (planned feature).
- Under z/VM 6.4 and older, memory cannot be removed dynamically, only added.
- This is because Linux tends to automatically address all memory available to it for use as cache. (Cached RAM is good under x86 distributed, but undesirable for a z/VM environment)

z/VM memory configuration – considerations

z/VM 6.4 increases supported real memory from 1 TB to 2 TB at LPAR level (Central storage)

- Individual Virtual Machine (USER / IDENTITY) limit remains at 1TB

Ensure sufficient dump space

Ensure sufficient paging space

- Even if not increasing memory used, a good time to double check space guidelines

- Check out the VIR2REAL utility on the z/VM Downloads page (See Additional Information slide)

z/VM memory configuration – considerations

Support for Virtual Machine (Linux) 1 MB large pages in 6.4

- Scale better and improve guest memory management
- Be more fair across users
- Honor reserved memory settings better
- To use this from Linux:
 - Build a kernel containing large page exploitation (this is the default build), add `hugepages=<n>` kernel parameter (number of large pages to be allocated at boot time)
 - If desired, set `sysctl` variable to enable allocating large pages from moveable memory

Be proactive in writing out memory pages to disk

- Read and write blocks of pages
- Use parallel channel paths (PAV) when available.

Linux virtual memory sizing

The maximum supported virtual machine memory allocation is 1TB

The most common mistake made by customers running Linux guests under z/VM is over-provisioning virtual memory due to lack of understanding

In a commodity hardware environment – including both dedicated servers and other virtualization solutions:

- Traditional wisdom suggests installing as much memory as possible
- Excess memory is used as I/O buffer and file system cache
- This cache and buffer is usually required to overcome the architectural limitations of the most popular commodity processor cores
 - Also the reason you typically see a drop off in performance at around 80-85% CPU utilization on these cores...

Linux virtual memory sizing

- The z/VM platform is not a commodity platform
 - Using thinking and methodologies from x86 does not apply!
- In a virtualized environment under z/VM, guests with excessively allocated RAM place unnecessary stress on the VM paging subsystem and can easily make guests run VERY POORLY
 - Real memory is a shared resource. caching pages in a Linux guest reduces memory available to other Linux guests
 - Larger virtual memory requires more kernel memory for address space management
 - Live Guest Relocation takes longer to complete
- Rightsizing Linux memory requirements on z/VM:
 - Is accomplished through iterative tuning of the STORAGE value for the virtual machine
 - Monitored with the “free” or “vmstat” commands along with /proc/meminfo
- Reference the following document by Stephen Wehr (IBM):
 - <ftp://software.ibm.com/common/ssi/sa/wh/n/zsw03049usen/ZSW03049USEN.PDF>

z/VM paging subsystem

- Block paging changes eliminated the benefits of 50% page space, making monitoring **very important**.
 - No loss of efficiency above 50% page space utilization
 - Contiguous storage no longer needed for block paging
- Recommend monitoring focused on availability versus performance (avoid ABEND)
 - Monitor for rapid growth in page space as well as overall size thresholds
- Early writing's goal is to keep the bottom 10% of the global aging list prewritten.
- Whether written-on-demand or pre-written, page space is still being used. From a monitoring perspective, this is all that matters.
- The closer your monitoring threshold is to 100%, the more automation is necessary to avoid an ABEND (how quickly page space can be added).
- Do not mix page space with any other space on a volume

z/VM paging subsystem – continued

- Rule Of Thumb - Plan for a DASD page space utilization < 90%:
 - Monitor usage with Q ALLOC PAGE command and automation
 - Block page size is the key performance indicator:
 - Aim for double digits, 10 or more 4K pages per block set
 - Perfkit report - FCX109 (CP Owned Device)
- Use multiple channels to spread out I/O to paging devices
- Recommend using devices of the same size/geometry
- Leverage HYPERPAV for paging devices and use fewer, larger devices
 - Recommend enabling via command and if no surprises, update system configuration file
 - Command: **SET PAGING ALIAS ON**
 - Configuration file: **FEATURES ENABLE PAGING_ALIAS**
 - Can also be controlled at control unit level

z/VM paging subsystem – continued

- For environments where the overcommit level is low and large amounts of real memory are being used, you will want to consider disabling early writes and keep slot
 - SET AGELIST EARLYWRITES NO KEEPSLOT NO
- EDEVs as paging drives are an option:
 - Have observed 1.6 I/Os per emulated FBA volume
 - At slightly higher CPU costs
- Page space calculation guidelines are located in the CP Planning and Administration Manual
- VIR2REAL also very useful in this situation to monitor

Linux swap

- The traditional recommendation in a dedicated server environment is that swap space should be twice the memory size of a Linux machine
 - This does not apply to a z/VM Linux guest:
 - Some swap space should be defined to prevent Linux from hanging and/or a kernel panic during unexpected memory demands
 - Properly sized Linux guests should have minimal swapping under normal load
- z/VM offers multiple options for swap devices
- Recommendation:
 - One or two small V-disks (256MB - 512MB)
 - If necessary, additional minidisk(s) or dedicated volume(s)
 - Set priorities in fstab so that the V-disk(s) are used first
- See *Additional Information* slide for more details and test results for various swap device options
- Virtualization Cookbook series covers how to setup VDISK and automatically format them at boot time

SFS is your friend!

As of z/VM 6.4 the default location for components is SFS instead of minidisks

- Minimizes having to waste time to mess around resizing (move to new & copy contents) minidisks:
 - This can be disruptive (usually is)
 - Can also create fragmentation on your DASD
- You could still elect to override and use minidisks instead, but why?
 - When you install Linux, do you create separate non-LVM volumes of fixed sizes for each individual path you intend to mount?
 - Spend more of your time on optimizing your system instead of managing disks

z/VM minidisk cache

- z/VM minidisk cache is a write-through cache:
 - Improves read I/O performance, but, it's not free in terms of system resources
- Not recommended for:
 - Memory constrained systems
 - Non-shared Linux file systems
 - Linux swap file disks
- Default system settings are less than optimal. Recommended settings:
 - Disable MDC for non-shared Linux minidisks
 - Code `MINIOPT NOMDC` operands on the MDISK directory statement
 - Limit MDC in central storage, amount depends on usage
 - `SET MDC STORE 0M 256M`
 - Monitor with Q MDC command and/or a performance monitor
 - Perfkit report - FCX103 (Storage Utilization)

Disk performance

- Hardware connectivity choices:
 - FICON available up to 16GB channel speeds
- SCSI verses ECKD/FBA recommendations:
 - ECKD or FBA for z/VM and Linux “/” file system
 - SCSI LUNs for application data and databases
- Maximize hardware performance:
 - Use maximum speed channels
 - Configure maximum number of channel paths
 - Spread disks over multiple ranks within a storage subsystem
 - Use logical volumes with striping
 - Recommend exploiting PAV or HyperPAV to prevent queuing
 - Ensure High Performance FICON is enabled for your ECKD DASD.
- References and more information on the *Additional Information* slide

Linux filesystems – EXT

- EXT2 – Formerly the most widespread, falling out of favor due to lack of journal
- EXT3 – Evolution of EXT2 with the addition of file system journal. Still 32-bit, so file sizes and file systems limited in size (you can push these limits somewhat, but can introduce instability doing so)
- EXT4 – The latest evolution - 64-bit:
 - Supports HUGE sizes for both individual files and overall file systems
 - Maximum individual file sizes can be up to 16 TB!
 - Overall maximum file system size is 1 EB (exabyte).
1 EB = 1024 PB (petabyte) ; 1 PB = 1024 TB (terabyte)
 - Directories can contain a up to 64,000 sub directories (EXT3 was 32,000 max.)
 - Multi-block allocation and delayed allocation
 - Journal checksum
 - Fast fsck
 - These new features have improved the performance and reliability of EXT4, making it a clear choice when compared to EXT3.

Linux filesystems – others

- XFS - the IRIX file system, which was released in 2000 as open source. Also extremely popular in systems with large files. Max size of both individual files and FS is 8 EB. The default for Redhat as of version 7.
- Keep in mind you can not shrink an XFS file system.
- BTRFS – The default root filesystem for SLES as of version 12.
- DO NOT use BTRFS for user/application data.

Linux file systems – ZFS (Ubuntu)

ZFS filesystem is available and fully supported by Canonical: ZFS on Linux is a stable, incredibly feature-rich file system with snapshot capabilities

No, it's not the ZFS you may know from z/OS. This was born at Sun Microsystems years back

- Provided as native k-mod

Some similarities to BtrFS, but without many of the limitations and pitfalls.

- Lowest bit rot of any filesystem
- Snapshot backups
- Copy-on-write clones
- Integrity checking at multiple levels (file, tree, etc)
- Automatic repairs
- High-efficiency compression

<http://zfsonlinux.org/>



Linux file systems

- Recommend using EXT4 or XFS:
 - Journal capabilities
 - Widely used across major distributions
 - Reduced CPU load compared to other journaled file systems
- Using EXT4:
 - Use the proper journal for your workload: Three options are journal, ordered, writeback
 - Your unit file to mount an ext4 filesystem (or /etc/fstab if you're still doing it the old way) should use the flag `relatime` or flags `nodiratime, noatime` unless you really need them on because you're using cheap commodity disks

Linux file systems

- Temporary and volatile files:

- Utilize fstab to mount /tmp as a tmpfs in RAM or mount several smaller ramdisks if your CPC/LPAR are memory rich

```
tmpfs /tmp tmpfs defaults,noatime,mode=1777 0 0
```

- Leverage shared dynamic ramdisk (/dev/shm or /run/shm) if your distribution creates it for writing out many tiny files, but monitor utilization closely – try not to fill it to more than around 85% capacity.
- If none of the above work for you, create an EXT2 filesystem for temp file use. You don't need a journal, LVM, striping, mirroring, or metadata on volatile data!

z/VM dump space

- Dump Space
 - Ensure there is sufficient dump space defined to the system
 - Recommend to re-check twice annually at least. You do not want to find out you don't have enough dump space during a problem
 - Recommend defining dedicated dump volumes
 - Dump space requirements vary according to memory usage, but roughly, you should have enough dump space to back all of your actual memory plus any that might be paged out for overcommitment.
 - **QUERY DUMP** command identifies allocated dump space.
 - Calculation guidelines are located in the CP Planning and Administration Manual
 - PRACTICE!
 - Practice dump procedures once per quarter.
 - You don't want to be figuring it out or needing to spend time consulting reference materials during an emergency!

z/VM spool space

- Spool Space
 - Various uses:
 - User printer, punch, reader files (console logs)
 - DCSS, NSS, system files
 - Page space overflow
- Spool Management:
 - Monitor with `QUERY ALLOC SPOOL` command
 - Recommend using the `SFPURGER` utility on the MAINT 193 minidisk
 - Rule-based, used to clean up spool space
 - Included in the no charge CMS Utilities Feature (CUF)
 - Virtualization cookbook covers how to set this up to run automatically
 - Or, run manually if needed to quickly take action
 - Test first: `VMLINK MAINT 193 < == RR > (INVOKE SFPURGER TEST)`
 - If test looks ok: `VMLINK MAINT 193 < == RR > (INVOKE SFPURGER FORCE)`

z/VM Environmental Record Editing and Printing Program (EREP)

The Environmental Record Editing and Printing Program (EREP) is a diagnostic application program that runs under the z/OS, z/VM and z/VSE operating systems.

- Assists IBM service representatives with problem determination
- Edits and prints reports from the records placed in the error recording data set (ERDS) by the error recovery program (ERP) of your operating system.
- Records come from device or system errors as well as informational or statistical data.
- IBM service can analyze information in the EREP reports to determine if a problem exists, what it could be, and where it occurred.

EREP should be run regularly and frequently, do not disable it!

z/VM EREP (Continued)

More details and information on how to work with EREP can be found in the following publications within the z/VM Knowledge Center at the following URL:

<http://www.ibm.com/vm/library/other.html>

- Program Directory for EREP VM V3 R5 M0 (Form number GI10-8140)
- EREP User's Guide (Form GC35-0151)
- EREP Reference Manual (Form GC35-0152)

Review chapter 5 of the EREP User's Guide

For a quick creation of a history file and clearing the EREP 191, create an input parameter file such as:

```
PRINT=NO  
ACC=Y  
ZERO=Y  
ENDPARM
```

Don't forget FILEDEFS for the work files, etc.

The guide also provides information on how to setup the processing of EREP data so that it can be done automatically. This is strongly recommended.

Networking configuration options

Cross LPAR network connectivity:

- Shared OSA Express
- HiperSockets
- HiperSockets Bridge

References:

- z/VM Connectivity Manual (SC24-6174-03)
- Networking Overview for Linux on Z Redpaper (See Additional Information slide)

Three basic configurations for external network connectivity:

- Dedicated OSA (not recommended)
- Routed LAN
- VSWITCH is recommended:
 - Lower CPU costs
 - Built-in failover
 - Operates in Ethernet or IP modes (Ethernet is recommended)
 - Supports 802.1q VLANs (by port or by user)
 - Supports port isolation
 - Supports 802.3ad link aggregation

VSWITCH options

- Keep in mind that Live Guest Relocation requires the option for PORTBASED or USERBASED to be set identically on the z/VM systems you will be relocating your Linux VMs within
- Recommend using PORTBASED
 - Easier to conceptualize since physical switches use port numbers for everything
 - Easier to keep track of and manage from an SCM perspective
- Recommend using VLAN AWARE NATIVE NONE when using ETHERNET
- Example of PORTBASED ETHERNET with VLAN AWARE NATIVE NONE:
 - `DEFINE VSWITCH VSW1 ETHERNET PORTBASED RDEV xxxx xxxx VLAN AWARE NATIVE NONE`

Directory Network Authorization

- As of z/VM 6.4 RSU 1702, Directory Network Authorization (DNA) made it so that the virtual NIC for each (Linux) Virtual Machine is now fully configured by statements in the user directory entry:
 - Inclusion of NICDEF statement(s) in a directory entry will now:
 - Define the virtual NIC
 - Grant authorization for the virtual machine to couple a VSWITCH
 - Couple the virtual NIC to the VSWITCH
 - Example:

```
NICDEF F001 TYPE QDIO LAN SYSTEM SWITCH1
NICDEF F001 MACID F30006
NICDEF F001 VLAN 256
```
- DNA eliminates requirement to include directory statements (or under AUTOLOG#, etc.):
 - MODIFY VSWITCH
 - SET VSWITCH GRANT
 - COUPLE

MTU sizes matter

Set MTU to the maximum supported by all hops on the path to the final destination to avoid fragmentation:

- Use Linux command `tracepath destination` to verify the path MTU size
- If the application data is ≤ 1400 bytes, use an MTU size of 1492 instead of Linux default of 1500
- If the application is able to send larger chunks of data, use an MTU size of 8992 for jumbo frames

TCP uses the MTU for the window size calculation, not the actual application send size

MTU sizes matter

For VSWITCH, an MTU size of 8992 is recommended:

- OSA card is optimized for use with an 8992 MTU
- Synchronous operation, SIGA required for every packet
- No packing like a dedicated OSA card
- Be sure PATHMTU discovery is turned on & your network will pass ICMP type 2 packets

For HiperSockets, select an MTU size to suit the workload:

- If an app is capable of sending large packets, larger MTU will increase throughput & decrease CPU use
- An MTU size of 56K is recommended only for data streaming workloads with packets >32KB

Are you relying on deprecated network programs?

There are 2 types of programs for the low level network configuration on Linux

- Old net-tools programs (ifconfig(8), ...) are from the Linux NET-3 networking system.
 - Most of these are obsolete now and officially deprecated.
 - Stop using ifconfig, route, and other net-tools – they are going to go away at some point in the future
- New Linux iproute2 programs (ip(8), ...) are the current Linux networking system.

Inbound QDIO Buffer

- The QDIO inbound buffer queue can be increased for high volume (including chatty) workloads:
 - The default on older Linux systems was 16, but now typically is 64
 - Valid range is 8–128
 - QDIO OSA buffer size is 64K
 - IQDIO HiperSockets buffer size is equal to the HiperSockets MFS (16K, 24K, 40K, 64K)

- Current buffer count can be displayed with the Linux command `lsqeth -p`

```
-----  
devices                CHPID interface          cardtype          port  chksum  prio-q'ing  rtr4  rtr6  lay'2  cnt  
-----  
0.0.ea00/0.0.ea01/0.0.ea02 x00  encea0          Virt.NIC QDIO    0    sw      always_q_2  no   no    0      64  
-----
```

- A QDIO OSA buffer count of 128 equates to 8MB locked memory: $128 \times 64\text{KB} = 8\text{MB}$
- Set the inbound buffer queue size in the appropriate config files or commands. Changes will be effective at next reboot:
 - SLES 12: edit `/etc/udev/rules.d/51-qeth-0.0.{OSA}.rules` and add the line:
`ACTION=="add", SUBSYSTEM=="ccwgroup", KERNEL=="0.0.ea00", ATTR{buffer_count}="128"`
 - Redhat: edit `/etc/sysconfig/network-scripts/ifcfg-{NIC}` and append `buffer_count={#}` to `OPTIONS`:
`OPTIONS="layer2=1 portno=0 buffer_count=128"`
 - Ubuntu: use the `chzdev` command to set the `buffer_count` attribute in the persistent configuration:
`chzdev qeth 0.0.ea00 buffer_count=128 -p`

Disable checksum on HiperSockets (HIPERS)

- Recommend turning off checksum for HiperSockets because:
 - HIPERS are a memory-to-memory operation protected by ECC, so checksum is a total waste of CPU.
 - Overhead, slows down processing with zero benefits.
 - Can save between 7%-13% in CPU costs
 - In a “z/VM mode” (mixed engine) LPAR with GPs and IFLs where HIPERS will use GPs, savings can add up quick.
- The typical default setting is `sw_checksumming`.
- To check status, run the command:
`ethtool -k {nic}`
- To turn it off:
 - SLES 12: `/etc/udev/rules.d/51-qeth-0.0.f200.rules` add:
`ACTION=="add", SUBSYSTEM=="ccwgroup", KERNEL=="0.0.f200", ATTR{checksumming}="no_checksumming"`
 - RHEL 6/7: `/etc/sysconfig/network-scripts/ifcfg-{nic}` add:
`OPTIONS="checksumming=no_checksumming"`
 - Ubuntu Server 16.04 & 18.04 run:
`ethtool -offload {nic} rx off tx off`

SYSCTL settings

- sysctl settings can be changed
 - temporarily by the `sysctl` command, or,
 - permanently in the config file `/etc/sysctl.conf`
- If less than 2500, the processor input queue length should be increased to at least 2500
- Check: `sysctl net.core.netdev_max_backlog`
- Set: `sysctl -w net.core.netdev_max_backlog=2500`
- Adapt the inbound and outbound window size to suit the workload
 - The following values are recommended for OSA devices:
 - `sysctl -w net.ipv4.tcp_wmem="4096 16384 131072"`
 - `sysctl -w net.ipv4.tcp_rmem="4096 87380 174760"`
 - System wide window size applies to all network devices
 - Applications can still use `setsockopt` to adjust the window size

SYSCTL settings – continued

- As a general rule of thumb, the default send/receive window size should be at least twice the MTU size
 - The SAP Enqueue Server requires a default send/receive window size of four times the MTU size
- For modern kernels, the following are recommended:
 - `sysctl -w vm.dirty_ratio=10` (default is 20)
 - `sysctl -w vm.dirty_background_ratio=5` (default is 10)
 - `sysctl -w vm.swappiness=5` (default is 60)

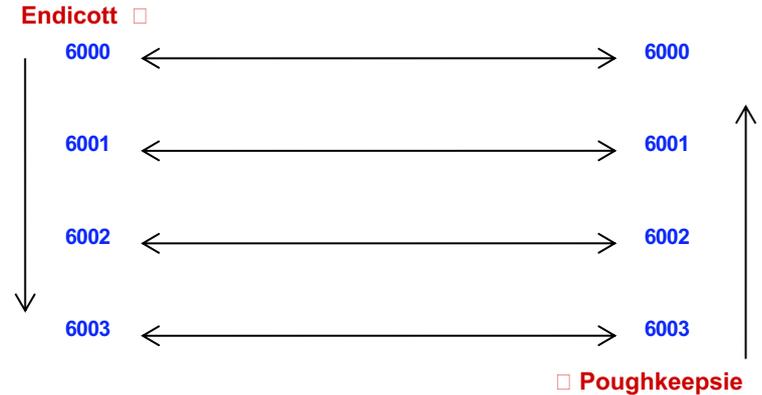
VMSSI cluster configuration considerations

- Suggested configuration for 4-member cluster is 2 LPARs on each of 2 CPCs
- Guest relocation time can be impacted by several key factors:
 - Number of ISFC Links (1 – 16)
 - Speed of ISFC Links (1Gb – 16Gb)
 - Size of guest virtual machine (memory)
 - How active the guest virtual machine is
 - Resource contention/availability on destination member
- Recommendation:
 - Minimum 4 CTCs between each cluster member: 2 FICON CHPIDs, 2 CTCs per CHPID
 - Maximum 16 CTCs between each cluster member: 4 FICON CHPIDs, 4 CTCs per CHPID
 - Testing has shown that 4 CTCs per CHPID provides the most efficient data transfer rates
 - Performance begins to degrade as the number of CTCs are increased beyond 4 per CHPID

CTC subchannel addressing

Use the same real device number for the same CTC on each SSI cluster member

- Potential performance impact
 - Algorithm does not use Round Robin
 - The more CHPIDs the greater the impact
- ISFC communications between two cluster members is done by:
 - Member name first in alphabet uses lowest subchannel address to highest
 - Member name second in alphabet uses highest subchannel address to lowest



Live Guest Relocation (LGR)

To qualify for relocation, a guest virtual machine must meet eligibility requirements, including:

- It must be logged on, but in a disconnected state
- Architecture and functional environment on destination member must be comparable to origin member
 - A relocation domain defines a set of members among which virtual machines can relocate freely
- Destination member must have the capacity to accommodate the guest
 - CPU
 - Memory
 - Paging Subsystem
- Devices and resources needed by guest must be shared and available on destination member
 - Network Connections
 - DASD

Use VMRELOCATE command with TEST operand

Recommend relocating guests serially as quiesce time is much shorter

Reserve all slots in the SYSTEM CONFIG file

CP OWNED volumes in an SSI Cluster

- There is still a total of 255 slots but:
 - DUMP and SPOOL volumes are shared and must be assigned to a unique slot number
 - RES, PAGE, and T-DISK volumes are not shared and can be assigned the same slot number on each cluster member

Recommendations:

- Adopt an easily recognizable volume naming convention that uses the RDEV in it
- Separate shared and non-shared volumes
 - DUMP and SPOOL volumes begin in slot 10 and are assigned in ascending order
 - PAGE and T-DISK volumes should begin in slot 255 and assigned in descending order:
 - Avoids interference with SPOOL volumes

– All unused slots in-between are defined as “RESERVED”

Networking: virtual MAC addressing

MAC address assignments are set through the VMLAN config statement

- MACPREFIX must be set to different value for each VM system
 - Especially if systems are on the same network segment!
 - Default is 02-00-00 for each member
 - Recommend last two bytes be replaced with a unique number for each system
 - DO NOT leave this as the default. It will come back to haunt you eventually!
- USERPREFIX must be set for SSI members
 - More on the next slide...

Networking: VMSSI cluster virtual MAC addressing

- MAC address assignments are coordinated across the SSI cluster through the VMLAN config statement
 - MACPREFIX **must** be set to different value for each member
 - Default is 02-00-00 for each member
 - Recommend last two bytes be replaced with the "system number" of each member
 - 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
- Examples:
 - VMSYS01: VMLAN MACPREFIX 021111 USERPREFIX 02AAAA
 - VMSYS02: VMLAN MACPREFIX 022222 USERPREFIX 02AAAA
 - VMSYS03: VMLAN MACPREFIX 023333 USERPREFIX 02AAAA
 - VMSYS04: VMLAN MACPREFIX 024444 USERPREFIX 02AAAA

Log vmsyscon onto the console at bootup

- Create the user in Linux with the standard method (e.g... `useradd -g root,wheel vmsyscon`)
- This can be a lifesaver for times when you need to quickly get onto the system with the authority to fix an issue like a full filesystem, an out of control process, or otherwise.
- With the user being automatically logged on, you will not be shut-out from logging in (like you would via SSH or otherwise) in case of the dreaded “Fork: unable to fork”
- Pre-requisites:
 - In production environments, you should be using an ESM anyway; so when this is setup correctly using the ESM, there is a full audit trail of who has logged onto the console to access this and when
 - Ensure LBYONLY and LOGONBY is set for each Linux VM, and eliminate use of individual logon passwords for each Linux VM

More on LOGONBY

- Avoid shared passwords on IDs as much as possible by leveraging LOGONBY.
- Consider enabling LBYONLY also

```
DIRMAINT FOR MAINT710 LOGONBY
-----DirMaint LOGONBY-----
Query or update the list of users on the current LOGONBY directory
statement.

Select one of the following:
  _ ? (Query)
  X ADD
  _ DELETE

For ADD or DELETE, fill in one or more Userids:
===> JWSMITH
===> RXJONES
===> PJWILSO
===> TJWATSO
===>

DIRMAINT FOR MAINT710 SETPW LBYONLY
```

General

Adopt automated build methodologies

- Plan for the future
- Eliminate a differentiator in new to z scenarios

Centralize & Standardize

- Linux code
 - Create a local installation repository
- IAM
 - Disable root SSH access & use IUCVtty for alternate root access.
 - Use an ESM with z/VM that provides LDAP functionality, such as RACF
 - Configure Linux login IDs to use LDAP
 - » ONE password to manage!
 - » Audit trail!
 - » Easier systems management
- Setup the z/VM IUCV Terminal Server
 - Central access point to Linux virtual servers on the same LPAR.
 - Works even if TCP/IP on the targets is not working!
 - Full-screen terminals (iucv tty, HVC) help those new to LoZ by eliminating a differentiator & help speed up admin and situation recovery for experienced.
 - Included in every commercially supported s390x Linux distro, as well as many others. If not included, free download from IBM DeveloperWorks.

Installation, planning, and administration

The following documentation can be extremely helpful for Installation, Planning and Administration:

- z/VM CP Planning and Administration (SC24-6178)
- Getting Started with Linux on System z (SC24-6194)

Subscribe to the list servers

List servers, commonly referred to as "listservs", are a form of communication via discussion forums hosted by education institutions and customer sites.

Why?

For the virtually instantaneous response to your questions from members of the community. And, also for you to respond and share your experience and help others. Most list servers publish their archives of past and ongoing discussions, so take a look at their archive link to see what's being talked about.

Which lists?

Most people will want to subscribe to both of the following to start out with:

- The IBM z/VM Operating System discussion
- LINUX-390 discussion



The z/VM download library

www.ibm.com/vm/download

The z/VM Download Library is a clearinghouse or repository for tools, documentation, and other nifty gadgets of interest specifically for VMers. We have set up the library so that both IBMers and non-IBMers can submit content and so that anyone can take content.



ADDITIONAL INFORMATION

Web Sites

- z/VM Performance:
www.ibm.com/vm/perf/
- z/VM Library:
www.ibm.com/vm/library/
- Linux on Z Performance:
www.ibm.com/developerworks/linux/linux390/perf/
- SRP Article:
www.ibm.com/vm/perf/reports/zvm/html/640srp.html
- STORCONF page:
www.ibm.com/vm/perf/tips/storconf.html
- CPU Utilization in an SMT world:
www.ibm.com/vm/perf/tips/smtutil.html

Disk & Disk Performance

- www.ibm.com/vm/perf/reports/zvm/html/scsi.html
- public.dhe.ibm.com/software/dw/linux390/perf/ECKD_vsus_SCSI.pdf
- public.dhe.ibm.com/software/dw/linux390/perf/disk_performance_optimizing.pdf

Networking

- www.ibm.com/redbooks/reports/abstracts/redp3901.html

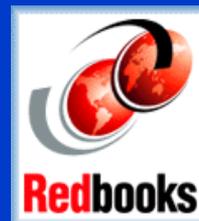
Linux swap

- www.ibm.com/redbooks/abstracts/sg246926.html

IBM Redbooks

www.ibm.com/redbooks

- The 2-volume series: An Introduction to z/VM SSI and LGR (SG24-8006) and Using z/VM SSI and LGR (SG24-8039)
- Performance Toolkit for VM (SG24-6059)
- Performance Measurement and Tuning (SG24-6926)
- The 4-volume series: The Virtualization Cookbook for z/VM and Linux on IBM Z Systems and IBM LinuxONE (SG24-8345)



Suggestions

- [Managing Memory with VMRM Cooperative Memory Management](#)

For the curious:

- [Detailed explanation by Malcolm Beattie of IBM on how SRM used to function before z/VM 6.4](#)
- [Old VM/ESA Storage Management with Tuning Guidelines book. Most of this is not current, but still excellent background info](#)
- Linux on Z interest site maintained by Mark Post of Suse: www.linuxvm.org

Thanks for listening!



Paul Novák
Senior IT Specialist, z/VM and Linux on Z
[Washington Systems Center](#)
[IBM Plant #1, Endicott NY](#)

धन्यवाद

多謝

감사합니다

Traditional Chinese

Korean

Спасибо

Russian

Gracias

Spanish

شكراً

Arabic

Thank You

English

Obrigado

Brazilian Portuguese

Grazie

Italian

多谢

Simplified Chinese

Danke
German

Merci

French

நன்றி

ありがとうございました

Japanese

ขอบพระคุณ

Thai

