

z/VM Startup and Shutdown Best Practices

MVMUA

October 17, 2017

John Franciscovich
francisj@us.ibm.com



Trademarks

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

BladeCenter*	GDPS*	IBM z13*	PR/SM	System z9*	zSecure
DB2*	HiperSockets	IBM z14	RACF*	System z10*	z/VM*
DS6000*	HyperSwap	OMEGAMON*	Storwize*	Tivoli*	z Systems*
DS8000*	IBM LinuxONE Emperor	Performance Toolkit for VM	System Storage*	zEnterprise*	
ECKD	IBM LinuxONE Rockhopper	Power*	System x*	z/OS*	
FICON*	IBM Z*	PowerVM	System z*		

* Registered trademarks of IBM Corporation

The following are trademarks or registered trademarks of other companies.

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

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

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

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

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

Java and all Java based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and

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

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

OpenStack is a trademark of OpenStack LLC. The OpenStack trademark policy is available on the [OpenStack website](#).

TEALEAF is a registered trademark of Tealeaf, an IBM Company.

Windows Server and the Windows logo are trademarks of the Microsoft group of countries.

Worklight is a trademark or registered trademark of Worklight, an IBM Company.

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

* Other product and service names might be trademarks of IBM or other companies.

Notes:

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

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

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

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

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

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

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

This information provides only general descriptions of the types and portions of workloads that are eligible for execution on Specialty Engines (e.g., zIIPs, zAAPs, and IFLs) ("SEs"). IBM authorizes customers to use IBM SE only to execute the processing of Eligible Workloads of specific Programs expressly authorized by IBM as specified in the "Authorized Use Table for IBM Machines" provided at www.ibm.com/systems/support/machine_warranties/machine_code/aut.html ("AUT"). No other workload processing is authorized for execution on an SE. IBM offers SE at a lower price than General Processors/Central Processors because customers are authorized to use SEs only to process certain types and/or amounts of workloads as specified by IBM in the AUT.

Notice Regarding Specialty Engines (e.g., zIIPs, zAAPs and IFLs):

Any information contained in this document regarding Specialty Engines ("SEs") and SE eligible workloads provides only general descriptions of the types and portions of workloads that are eligible for execution on Specialty Engines (e.g., zIIPs, zAAPs, and IFLs). IBM authorizes customers to use IBM SE only to execute the processing of Eligible Workloads of specific Programs expressly authorized by IBM as specified in the "Authorized Use Table for IBM Machines" provided at www.ibm.com/systems/support/machine_warranties/machine_code/aut.html ("AUT").

No other workload processing is authorized for execution on an SE.

IBM offers SEs at a lower price than General Processors/Central Processors because customers are authorized to use SEs only to process certain types and/or amounts of workloads as specified by IBM in the AUT.

Topics

Describe best practices for starting and shutting down your z/VM system:

- Things to check before IPL

- Starting your system
 - IPLing z/VM
 - After IPL
 - Validating system resources
 - Starting server virtual machines
 - Starting guest virtual machines

 - Shutting down
 - Guest virtual machines
 - Server virtual machines
 - z/VM system

IBM Solutions to Help Manage and Protect Your z/VM System(s)

- Security and User Management
 - RACF Security Server for z/VM (PF)
 - Directory Maintenance Facility for z/VM (DirMaint) (PF)
- Automation and operational monitoring
 - Operations Manager for z/VM (LP)
 - Programmable Operator Facility (PROP) (VM)
- Performance monitoring
 - OMEGAMON XE on z/VM and Linux (LP)
 - Performance Toolkit for z/VM (PF)
- Backup and recovery
 - Backup and Restore Manager for z/VM (LP)
 - Tape Manager for z/VM (LP)
 - Spectrum Protect (aka Tivoli Storage Manager) (LP)
- Interactive provisioning and system resource management
 - IBM Wave for z/VM Performance monitoring and data analysis for z/VM systems (LP)

(PF) – priced feature of z/VM
(LP) – licensed product
(VM) – included with z/VM

Before Starting Your System

Identify Your IPL Environment

- "Where and Why" are you IPLing your system?
 - Production
 - Test (LPAR or as a guest?)
 - Disaster Recovery (DR) test (local or at DR site?)
 - Actual DR

- Each of the above might require different definitions and actions
 - Configuration attributes
 - Devices
 - Which virtual machines to start
 - Service
 - Database
 - Networking
 - Production guests

- Automation is likely to be different for each environment
 - Monitoring and actions for various system and guest events

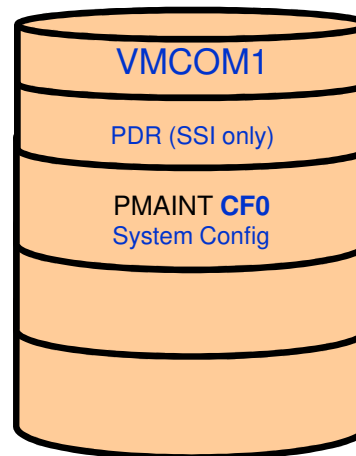
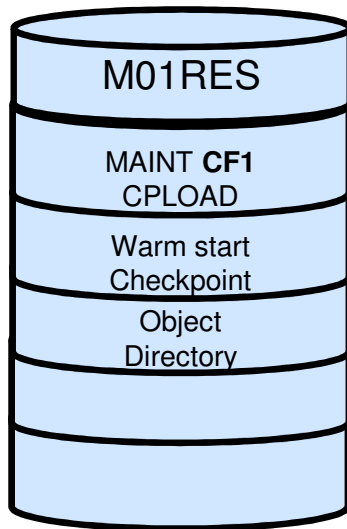
Identify Your IPL Environment ...

- Do you need to specify or change default IPL characteristics for this IPL?
 - IPL Parameters
 - Which system configuration file to use
 - Which parm disk to use
 - System/operator's console
 - Value of the CP system environment variable (CP.IPLPARMS.IPLVAR)
 - Re-initialize the PDR (SSI cluster – all members must be down)
 - etc.

- Use the CP IPLPARMS environment variable (**CP.IPLPARMS.IPLVAR**) provided with z/VM 6.4 to specify the IPL environment
 - Can be used by automation scripts to determine which actions to take and/or skip

Understand Disks Required to Start Your System

- *IPL volume*
 - Where the loader program (SAPL) is
- *"RES" volume*
 - Contains minidisk (CF1 parm disk) where the CP MODULE is
 - Warmstart and checkpoint areas (CP-formatted)
 - Object directory
- *Default IPL/RES volume:*
- *CF0 Parm disk*
 - Where the system configuration file is
- *Default CF0 Parm disk:*



- *Other system and CP-Owned volumes*
 - Page
 - Spool
 - Guest and user data

Designate Ownership of CP-owned Volumes?

- Ownership information may be specified for CP-owned volumes
 - Isolates a volume to a single system
 - Prevents CP data from being written on a volume owned by another system

- Ownership definition is mandatory in an SSI cluster, optional otherwise

- Specified with **OWNER** operand of CPFMTXA
 - Owning system name
 - Name of SSI cluster (if system is an SSI member)
 - A cluster name can be specified without a system name
 - "Cluster-wide" volumes

Verify I/O Configuration

▪ IBM I/O Exerciser for System z (ESAIO)

- Helps identify cabling or definition errors
 - Validates all paths defined to each device
- Can be run standalone in LPAR (with SAPL) or on a running z/VM system
- See z/VM 6.4 System Operation (chapter 3) for details

```

System Check Out Tool
- NUMBER OF I/O DEVICE SUBCHANNELS: 0119C6 (HEX) - HIGHEST SCH#: 034BDD
- NUMBER OF *RESERVED* SUBCHANNELS: 017158 (HEX) - HIGHEST SCH#: 03FFFE
- NUMBER OF TYPE-1 CFG SUBCHANNELS: 000001 (HEX) - HIGHEST SCH#: 01FF00
- NUMBER OF TYPE-2 MSG SUBCHANNELS: 0000C4 (HEX) - HIGHEST SCH#: 014552
- NUMBER OF TYPE-3 ADM SUBCHANNELS: 000000 (HEX) - HIGHEST SCH#: 000000
- EXTENDED TEST (CABLING & DASD READ_VALID) STARTED... 16:14:25
* CHP=90 DEV= 0510 ERROR ON CMD=E4 CTRL=4407 STAT=0000 (PATH INOP) 16:14:26
* CHP=95 DEV= 0510 ERROR ON CMD=E4 CTRL=4407 STAT=0000 (PATH INOP) 16:14:26
* CHP=90 DEV= 0511 ERROR ON CMD=E4 CTRL=4407 STAT=0000 (PATH INOP) 16:14:26
* CHP=95 DEV= 0511 ERROR ON CMD=E4 CTRL=4407 STAT=0000 (PATH INOP) 16:14:26
* CHP=90 DEV= 0512 ERROR ON CMD=E4 CTRL=4407 STAT=0000 (PATH INOP) 16:14:26
* CHP=95 DEV= 0512 ERROR ON CMD=E4 CTRL=4407 STAT=0000 (PATH INOP) 16:14:26
* CHP=90 DEV= 0513 ERROR ON CMD=E4 CTRL=4407 STAT=0000 (PATH INOP) 16:14:27
* CHP=95 DEV= 0513 ERROR ON CMD=E4 CTRL=4407 STAT=0000 (PATH INOP) 16:14:27
* CHP=B8 DEV= 1500 ERROR ON CMD=E4 CTRL=4407 STAT=0000 (PATH INOP) 16:14:36
* CHP=85 DEV= 1500 ERROR ON CMD=E4 CTRL=4407 STAT=0000 (PATH INOP) 16:14:36
* CHP=B8 DEV= 1501 ERROR ON CMD=E4 CTRL=4407 STAT=0000 (PATH INOP) 16:14:36
* CHP=85 DEV= 1501 ERROR ON CMD=E4 CTRL=4407 STAT=0000 (PATH INOP) 16:14:36
* CHP=B8 DEV= 1502 ERROR ON CMD=E4 CTRL=4407 STAT=0000 (PATH INOP) 16:14:36
* CHP=85 DEV= 1502 ERROR ON CMD=E4 CTRL=4407 STAT=0000 (PATH INOP) 16:14:36
* CHP=B8 DEV= 1510 ERROR ON CMD=E4 CTRL=4407 STAT=0000 (PATH INOP) 16:14:36
- SELECTING DEVICE= 1510. PLEASE WAIT... 16:14:36
# SCREEN FULL... PRESS ENTER, OR TYPE (DEL) FOR AUTO WRAP MODE ==>

```

Specify Location of System/Operator Console

- Operating System Messages Panel of HMC
 - LOADPARAM **CONSSYSC**
 - **CONS=SYSC** IPL parameter
 - **SYSTEM_CONSOLE** on **OPERATOR_CONSOLES** configuration statement
- Integrated 3270 Console
 - LOADPARAM **SYSG**
 - **CONS=SYSG** IPL parameter
 - **SYSTEM_3270** on **OPERATOR_CONSOLES** configuration statement
- Device specified by address on **OPERATOR_CONSOLES** statement
- Selected system console becomes the primary system operator console when OPERATOR virtual machine is logged on during IPL

- If using an application that monitors the OPERATOR's console via SECUSER, disconnect as soon as possible so console logs can be captured
 - e.g. Operations Manager for z/VM
 - Applies to all virtual machines you are monitoring
 - Alternatively, use OBSERVER if you want to stay connected

Virtual Network Definitions

- If using Inter-VSwitch Links (IVLs), define them in the system configuration file
 - **DEFINE VSWITCH *switchname* IVL ...**
 - Must be established before any Global VSwitches or Shared Port Groups

- Global VSwitches and Shared Port Groups can also be defined in system configuration file
 - Creation will be deferred until the IVL virtual switch Uplink Port connects the host to the IVL domain ("**active**" state)
 - Define dynamically (after IPL) to allow IVL to be established

- Define virtual NICs in guests' directory entries
 - **NICDEF *vdev* LAN SYSTEM *vswitchname***
 - Will automatically authorize and couple to *vswitchname* if available when guest logs on
 - "authorize" requires Directory Network Authorization (DNA) support – APAR VM65925 (z/VM 6.4)
 - Eliminates need to authorize with a **SET VSWITCH GRANT** command

Starting Your z/VM System

Loading and Starting z/VM

- SAPL (Stand Alone Program Loader)
 - Used to load and start z/VM
 - Default SAPL is created during z/VM installation
 - Can rebuild and customize using SALIPL utility
 - Menu can optionally be displayed when IPLing z/VM
 - Display only when changes to IPL attributes and/or IPL parameters are required

```
STAND ALONE PROGRAM LOADER: z/VM VERSION 6 RELEASE 4.0
```

```
DEVICE NUMBER: 018B MINIDISK OFFSET: 35 EXTENT: -
```

```
MODULE NAME: CPLOAD LOAD ORIGIN: 2000
```

```
-----IPL PARAMETERS-----
```

```
cons=0080 iplvar=PRODUCTION fn=MYPROD
```

```
-----COMMENTS-----
```

```
-----
```

```
9= FILELIST 10= LOAD 11= TOGGLE EXTENT/OFFSET
```

How Do You IPL?

- Automated – no operator intervention unless there is a problem
 - **FEATURES** statement in system configuration file:
 - **ENABLE AUTO_WARM_IPL**, or
 - **AUTO_IPL WARM**
- Manual – operator prompted for type of IPL

```
Start ((Warm|Force|COLD|CLEAN) (DRain) (DIsable) (NODIRect)
      (NOAUTolog)) or (SHUTDOWN)
```

- This is required if the TOD clock needs to be set (rare)
- Can also disable autolog or terminate IPL (shutdown) in case of errors

IPL Issues

- Type of problems that could occur during IPL
 - Errors in IPL parameter processing
 - Errors in configuration file processing
 - CP-owned volumes that are offline
 - Could result in loss of data (spool files, saved systems, and saved segments)
 - Warmstart or checkpoint errors
 - Might prompt to change to FORCE start
 - Could also result in loss of data
 - Connectivity issues when joining an SSI cluster

- Decide whether to
 - Stop IPL to fix any errors
 - Continue/complete IPL and attempt to fix them dynamically

- Some errors will result in a wait state

After IPL: Validation, Servers, and Guests

Getting the Work Started after IPL

- Virtual machine "hierarchy"
 - Top priority: protecting your system and data
 - Virtual machines that create and collect system data
 - Some of these are started during IPL
 - Guests will have dependencies on various system services and virtual machines
 - Networks
 - Service virtual machines
 - Linux guests that run applications or application servers require database servers to be up and running first
- Service virtual machines that do monitoring and help with system operation can help with the above
 - Start virtual machines in the correct order
 - Verify pre-req virtual machines are available

Priority 1- External Security Manager (ESM)

- Make sure ESM is started before any other service machines
 1. Change AUTOLOG1 userid's PROFILE EXEC to autolog **only** RACFVM
 2. Move everything else in AUTOLOG1's PROFILE to AUTOLOG2
 3. RACF will start AUTOLOG2 after it is initialized

- If you are doing maintenance on RACF and don't want RACFVM or other service machines to start:
 1. Set **CP.IPLPARMS.IPLVAR** (IPL environment variable) to something like "RACFMNT"
 2. Check value in AUTOLOG1's PROFILE EXEC
 3. If "RACFMNT"
 - Autolog user RACMAINT instead of RACFVM
 - Do not start AUTOLOG2

Service Virtual Machines, etc.

▪ AUTOLOG2 (follow the hierarchy)

- Start service machines for system operation, monitoring, and management, including:
 1. Operations Manager, PROP
 - Start these first so they can monitor other service machines (consoles, etc.)
 - Use to verify service machines and manage dependencies among them
 2. TCP/IP
 3. MONITOR service, Performance Toolkit, OMEGAMON XE
 4. DirMaint
 5. Systems Management API (SMAPI) servers
 6. etc.

- Invoke scripts to
 - Verify system resources
 - Define and verify virtual networks
 - Start guests

Verify System Resources

- Create automation scripts to check system resources (and record in console file)
 - Make sure the virtual machine issuing commands has correct privilege class to query system-wide information

- Use QUERY commands to verify:
 - Available processors (**QUERY PROCESSORS**)
 - How much real memory is available? (**QUERY STORAGE**)
 - If less than expected, make sure all memory is defined as central storage (no expanded storage)
 - Page and spool volumes and available space (**QUERY ALLOC PAGE, QUERY ALLOC SPOOL**)
 - System data (EREP, Symptom, Accounting records) is being collected (**QUERY RECORDING**)
 - Userids defined on **SYSTEM_USERIDS** configuration statement
 - Usually logged on by the system during IPL

- Make sure CP dumps are enabled in case of a system failure (**QUERY DUMP**)
 - If not, you might have to allocate more dump space on your spool volumes
 - Enable dumps (**SET DUMP DASD CP IPL**)

- If running in an SSI cluster, check SSI state and mode to make sure system successfully joined the cluster (**QUERY SSI**)

Verify Virtual Networks

- Create a "networking script" in AUTOLOG2 to define and verify networks
 - Verify IVL defined in system configuration file is operational
 - **QUERY VSWITCH** shows Uplink Port state is **"active"**
 - Define Global VSwitches and verify that they are operational
 - **DEFINE VSWITCH *switchname* GLOBAL ...**
 - **QUERY VSWITCH** shows Uplink Port state is **"active"**
 - Issue **SET VSWITCH GRANT** commands to authorize guests with virtual NIC definitions in their directory to couple to virtual networks when they log on
 - Not needed if Directory Network Authorization (DNA) support is applied

Starting Guests

- "Pace" guest startups to minimize contention for system resources
 - Depends on workloads and guest requirements
 - CPU
 - Memory
 - Network connections and communication
 - Pacing can be simple or sophisticated
 - "Sleep" between guests or groups of guests
 - Monitor startup completion of one group of guests before starting the next group
 - Examine performance statistics to determine when system is ready to start another group
- Verify each group of guests have completed startup before starting next group

Automating Guest Startups

- Operations Manager for z/VM can automate management of guest startups
 - Monitor consoles
 - Monitor system events such as SSI state/mode and network status
 - Define rules to take specific actions based on the above
 - Issue commands
 - Invoke programs/scripts
 - Error handling and notification
 - The set of rules that is used for various events/messages can be changed based on current system state
 - For example, shutting down system instead of starting system

Shutting Down

Shutting Down Guest(s)

- Orderly shutdown of guests is recommended when
 - Individual guests or groups of guests are shut down
 - The system is shutting down

- Orderly shutdown helps to avoid
 - Unnecessary file system checks and repairs when guest is restarted
 - Application errors in the event a pre-req service machine is shutdown before the guest that has dependencies on it

- Enable guests to receive shutdown signals
 - Useful for orderly shutdown of both individual guests and the entire z/VM system
 - Guests that are enabled will receive a "termination" signal when
 - The system is shutting down
 - **FORCE** command is issued to log the guest off
 - **SIGNAL SHUTDOWN** command is sent to the guest

- Guests that receive signals will have a designated amount of time to shutdown and logoff
 - **SHUTDOWNTIME** interval for the system
 - Interval or time of day designated when the signal is sent

Shutting Down Your z/VM System

- If in an SSI cluster, Live Guest Relocation (**VMRELOCATE** command) can be used to relocate Linux guests to another member of the cluster
 - Enables guests running key applications to continue running through "this" system's outage

- **SHUTDOWN** command causes a shutdown signal to be sent to enabled virtual machines
 - All guests receiving the signal will shutdown independently of other guests
 - Guest virtual machines that do not receive the signal will be logged off the system without consideration of any in-progress work or dependencies with other virtual machines

- But ... remember the virtual machine hierarchy!
 - Before issuing the **SHUTDOWN** command
 - Issue **SIGNAL SHUTDOWN** to groups of guests and service virtual machines
 - Reverse the order that you started the guests and other virtual machines – this will prevent pre-reqs from being shutdown before their dependent guests
 - Wait until one group completes its shutdown before signaling the next group
 - After all of the guests have been signaled and responded, issue **SHUTDOWN** command for z/VM system

- As with startup, shutdown of virtual machines can be monitored and automated
 - For Operations Manager, a new set of rules and actions can be defined
 - Do not restart guests when they have shutdown/logged off
 - Manage signaling sequence for groups of guests

Summary

- Employing some of the best practices described here will help startup and shutdown of your z/VM systems and workload go smoothly

- Specific steps will be different for every system
 - Company policies and procedures
 - System capacity and workloads

- Using documented steps and procedures makes it easier when
 - Applying service or upgrading your system
 - Transferring processes to someone else

- Using tools to monitor and automate startup and shutdown processes is recommended
 - Detect completion of one process and start the next
 - Problem detection and notification or remedy