Hints and Tips for Migrating Workload to IBM POWER9 Processor-Based Systems

Version 1.0
April 2020

Disclaimer – Hints & Tips for Migrating Workload to IBM POWER9 Processor-Based Systems

Acknowledgements

We would like to thank the many people who made invaluable contributions to this document. Contributions included authoring, insights, ideas, reviews, critiques and reference documents.

Our special thanks to key contributors from IBM Power Systems Performance:

Abraham Alvarez - Power Systems Performance

Clark Anderson - Power Systems Performance

Qunying Gao - Power Systems Performance

Bernard King-Smith - Power Systems Performance

Dirk Michel - Power Systems Performance

Rick Peterson - Power Systems Performance

Our special thanks to key contributors from IBM Development:

Juan M Casas - AIX development

Arnold Flores - AIX development

Bob G Kovacs - AIX development

Our special thanks to key contributors from IBM Global Support:

Janine A Jackson - AIX/VIOS L3 support

Evan Zoss - AIX/VIOS L3 support

1 Preface

This document is intended to provide guidance, best practice recommendations, and a checklist for migrating workloads from POWER7 and POWER8 systems to POWER9. This document by no means covers all the PowerVM, AIX or IBM i best practices and should be used in conjunction with other PowerVM, AIX or IBM i documents.

2 Introduction

When workloads are deployed on new hardware configurations, care must be given to the configuration and tuning of the new system to achieve the expected performance. This starts with good system configuration planning and continues into system setup and deployment. By spending some time to consider and implement these migration guidelines and to establish sound system tuning and configuration practices, you will be better prepared for a successful migration to a POWER9 process-based system.

3 Planning Workload Migration

The work of planning your migration should include all different areas which pertain to your applications, as well as some of the key areas that we discuss below, so that your migration runs smoothly, and you obtain the best performance from your hardware. Some of these include:

3.1 Install the latest firmware on the POWER9 system, update HMC software to latest version.

3.2 The support website [FixCentral](#) provides updates for the latest OS updates for IBM i, VIOS, AIX. Where possible, update the OS levels of the logical partitions to be migrated to the latest levels, or at the very least the minimum levels recommended for the platform.
NOTE: Be aware that installing only the minimum levels can potentially leave your partitions or workloads vulnerable to issues that have been resolved in some of the latest updates of the operating systems. See section 4 and 4.1 in this document for detail link.

3.3 POWER9 makes more efficient use of the 8 hardware SMT threads available per CPU (when running in SMT8 mode). When migrating from a POWER7 or POWER8 platform, consider the use of SMT8, in addition to

considering reducing the allocation of CPUs (in dedicated CPU LPARs), or reducing VPs and CPU entitlement (on shared CPU LPARs). Refer to [Virtual Processor and Entitlement Considerations](#) (section 6) in this document for additional information.

3.4 Placement of the partition is important in order to obtain the best performance in your POWER9 system. In order to optimize the initial placement of your partitions, it is best to create the most important and/or largest partitions first, followed by less important and/or smaller partitions. If necessary, placement can be optimized after the partitions have been created by using DPO on the HMC against either the whole managed node or the partition you are most interested in. It is also suggested to run DPO if you are constantly running DLPAR operations across partitions on your managed node; constantly running DLPAR (either add or remove) can lead to resources being placed in non-optimal locations. To check the affinity score of your system or LPARs, you can use the [lsmemopt](#) (see the HMC documentation for further information).

3.5 Capacity planning is important when considering processor migration. Consider the application behavior (e.g. highly multi-threaded workloads vs single-threaded workload) when setting performance improvement goals and expectations. See the following documents for more information:

[IBM i on Power -Performance FAQ](#)
[IBM Power Systems Performance Report](#)

3.6 EnergyScale performance can allow higher processor frequency under the right conditions. Consider the workload, the system environment, and EnergyScale configuration to better understand how performance can be affected.

POWER9 processors can be enabled to utilize a dynamic frequency mode, which provides the highest frequency possible based on the processor utilization, while also reducing the frequency of the core when the cores are otherwise idle, in order to improve the power consumption of the system. Please note that not all workloads will drive the system to use the highest possible frequency. See the following documents for more information:

[IBM EnergyScale for POWER9 Processor-Based Systems](#)
[POWER9 EnergyScale - Configuration & Management](#)

4 Software Requirements

The following lists latest software for migration.

AIX Requirements:
https://www-01.ibm.com/support/docview.wss?uid=ssm1platformaix

VIOS Requirements:
https://www-01.ibm.com/support/docview.wss?uid=ssm1platformvios

HMC / Firmware Supported Combinations
https://www-945.ibm.com/support/fixcentral/main/transform?xml=https://download.boulder.ibm.com/ibmdl/pub/software/server/firmware/sfw_fixSupportedCombos.xml&title=HMC+/+Firmware+Supported+Combinations

HMC and HMC Virtual Appliance on Fix Level Recommendation Tool (FLRT) Lite
https://www14.software.ibm.com/support/customercare/flrt/liteTable?prodKey=hmc

POWER Code Matrix - Supported HMC Hardware
https://www14.software.ibm.com/webapp/set2/flrt/sas?page=mtm-supported-hmc

IBM i Technology Updates:
www.ibm.com/ibmi/techupdates

IBM i Support: Recommended fixes:
https://www.ibm.com/support/pages/ibm-i-support-recommended-fixes

4.1 Known issues

AIX

There are some issues that have been found to make an impact in performance on LPARs which are migrated to POWER9 based systems. We recommend that the LPARs which are planned to be migrated, are updated to the following AIX levels in order to prevent being exposed to performance issues after the migration to POWER9:

Description	APAR / Fix available in
P9 VPM fold threshold (refer to Appendix 10.1 for temporary fix)	IJ10664 / AIX 7100-05-03-1846 IJ10535 / AIX 7200-02-03-1845 IJ10425 & IJ10698 / AIX 7200-03-02-1846 IJ10423 & IJ10711 / AIX 7200-04-00-1937 IJ17092 / AIX 7100-05-05-1939 IJ19430 / AIX 7200-02-05-1938 IJ19016 / AIX 7200-03-04-1938 IJ16649 / AIX 7200-04-00-1937
VPM intelligent folding not supported	IJ20661 / AIX 7100-05-06-2015 IJ21390 / AIX 7200-04-02-2015 **Currently unshipped:** Request ifix for APAR IJ21390 for your AIX level.
Process vnicserver_crq uses high CPU	VIOS ONLY IJ20354 / 2.2.6.60 IJ21338 / 3.1.1.20 **Currently unshipped:** Request ifix for APAR IJ19731 for your VIOS level
Possible I/O performance problem with VIOS client LPARs	VIOS 3.1.1 ONLY IJ23222 / 3.1.1.20
Possible I/O hangs on NPIV client backed by SLI-4 adapter on VIOS	VIOS 3.1.1 or AIX 7200-04 ONLY IJ22290 / 7200-04-02-2015 & 3.1.1.20

5 Processor Compatibility Mode

To migrate a logical partition from POWER7 or POWER8 to POWER9 system, you should configure logical partition preferred processor compatibility mode to default mode. The configuration can be done through logical partition profile on HMC. If you need to migrate a logical partition from Power9 system back to POWER7 or POWER8 system, configure logical partition preferred processor compatibility mode to P7 mode or P8 mode accordingly before migration. The compatibility mode of logical partitions will be preserved across migration unless you change it. The change of the compatibility mode of the logical partition is not dynamic, it requires a shutdown and restart of the logical partition. When you restart the logical partition, hypervisor checks the configured processor compatibility mode and determines whether the operating environment supports that mode. If the operating environment supports the

configured processor compatibility mode, the hypervisor assigns the logical partition the configured processor compatibility mode. If the operating environment does not support the configured processor compatibility mode, the hypervisor assigns the logical partition the most fully featured processor compatibility mode that is supported by the operating environment. For more detail on processor compatibility mode, refer to https://www.ibm.com/support/knowledgecenter/POWER9/p9hc3/p9hc3_pcm.htm AIX command lsconf is handy to check which compatibility mode your partition is running.

For AIX >=7.2TL1, the default SMT level is SMT4 on both POWER7 and POWER8 systems, and the default is SMT8 on POWER9 system with compatibility mode of P8 and above. After migrating from POWER7 or POWER8 to POWER9 system, if you don't reboot AIX, it will continue running in SMT4 mode. After rebooting AIX, if logical partition operates in P7 mode, AIX will remain in SMT4 mode. If logical partition operates in P8 mode or above, AIX will change from SMT4 to SMT8 mode. If you want to preserve SMT4 mode across logical partition reboots, you need to run smtctl and bosboot.

With AIX >=72TL4 and firmware 940 on Power9 system, configure logical partition's processor compatibility mode to P9 mode after migrating to POWER9 system will allow the logical partitions to leverage the advanced capabilities of the POWER9 processors, taking advantage of new AIX features.

The following table describes the features supported with various compatibility modes on POWER9 system.

	XIVE	NX GZip	P9 PMU	SMT8
P7 Mode	-	-	-	-
P8 Mode	-	-	-	X
P9 Mode (<firmware940)	-	-	X	X
P9 Base Mode (firmware940)	-	-	X	X
P9 Mode (firmware940 + AIX >=7.2TL4 + IBM i >=7.4)	X	X	X	X

XIVE: eXternal Interrupt Virtualization Engine. POWER9 system supports a larger number of interrupts sources and delivers interrupts directly to virtual processors without going through hypervisor.
NX GZip: POWER9 processor-based servers support on-chip accelerators that perform various functions such as compression, decompression of data.

P9 PMU: Performance Monitor Unit. PMU is a programmable component of microprocessor core on the chip. The PMU provides a programmable interface for monitoring and collecting various hardware performance event counters.

6 Virtual Processor and Entitlement Considerations

As mentioned earlier in this document, the POWER9 processors have improvements for all SMT modes (e.g. SMT2, SMT4, SMT8). When using SPLPARs (shared processor LPARs), it is therefore recommended that you follow these guidelines in order to obtain the best results in your migration to POWER9.

With these new improvements, in general, workloads will show better performance when run in SMT8 mode, but at the same time, workloads running in SMT2 mode, will also see a significant improvement.

When planning the shared processor configuration, you need to understand well the goal that you have in mind when migrating to a POWER9 based systems.

Your VP and entitlement configurations previously determined for POWER7 and POWER8 based systems should be revisited and most likely reduced when migrating to a POWER9 based system using these guidelines to achieve your intended goal.

As a rule of thumb, it is recommended that you assign the VPs and entitlement to the LPAR as follows:
- Use the peak utilization of the LPAR as the baseline to assign VPs.
- Use the average utilization of the LPAR as the baseline to assign entitlement.

For further guidance on virtualization, see Chapter3 and 4 in the following document: IBM Power Virtualization Best Practices.

Given the higher thread strength on POWER9 processors, to maximize the processing capacity of the processor, we recommend using all the hardware threads available by using the SMT8 mode (default setting) on the partition.

In addition to using SMT8, if the goal of the migration is to reduce the processor capacity utilized by the LPAR, you must reduce the number of CPUs (dedicated LPAR) or VPs (shared LPAR) allocated to the LPAR on the new POWER9 based system. As a baseline, evaluate the capacity required to run your existing workload on a POWER9 system based the AIX capacity rating (rPerf value) which can be found on the IBM Power Systems Performance Report or, for IBM i capacity planning, the CPW rating which is listed in Section 3.5 in this document.

(AIX only)

Note that SMT4 is default on POWER8 based systems. For POWER9 based systems running AIX 7.2 TL3 and above the default mode is SMT8; the default for LPARs at older AIX levels on POWER9 is still SMT4.

In addition to the reduction of VPs, you can consider setting the schedo tunable "vpm_throughput_mode" to a value of 2. By default, running in raw throughput mode (i.e. vpm_throughput_mode=0) AIX spreads all available work across as many VPs as available, dispatching work **first** to the primary SMT thread of each VP, then the secondary SMT threads, and so on, thus providing the best performance in most workloads. This configuration, however, can lead to a higher PC (processor capacity) utilized by the LPAR. Depending the workload, you can reduce the PC by setting the vpm_throughput_mode tunable to 2, thus having AIX to schedule work to the primary and secondary threads equally. A more detailed discussion of the different modes can be found in the [IBM Power Virtualization Best Practices.](#)

One way you can review the CPU utilization of an AIX partition is by looking at the output of the mpstat command. In the following example you can see that mostly the primary threads of each of the VPs assigned to this LPAR are busy doing work.

cpu	min	maj	mpc	int	cs	ics	rq	mig	lpa	sysc	us	sy	wa	id	pc	%ec	lcs	Time
0	0	0	0	265	94	1	1	1	100	288	12	65	0	22	0.00	0.1	225	18:55:39
1	0	0	0	14	0	0	0	1	100	0	0	0	0	100	0.00	0.1	15	18:55:39
2	0	0	0	9	0	0	0	1	100	0	0	1	0	99	0.00	0.0	10	18:55:39
3	0	0	0	9	0	0	0	1	100	0	0	0	0	100	0.00	0.1	10	18:55:39
4	0	0	0	9	0	0	0	1	100	0	0	1	0	99	0.00	0.0	10	18:55:39
5	0	0	0	9	0	0	0	1	100	0	0	0	0	100	0.00	0.1	10	18:55:39
6	0	0	0	9	0	0	0	1	100	0	0	1	0	99	0.00	0.0	10	18:55:39
7	0	0	0	9	0	0	0	1	100	0	0	0	0	100	0.00	0.1	10	18:55:39
8	0	0	0	120	32	11	1	2	100	0	100	0	0	0	0.32	10.7	100	18:55:39
9	0	0	0	9	0	0	0	1	100	0	0	0	0	100	0.12	4.1	11	18:55:39
10	0	0	0	9	0	0	0	1	100	0	0	0	0	100	0.06	1.9	11	18:55:39
11	0	0	0	9	0	0	0	1	100	0	0	0	0	100	0.12	4.1	11	18:55:39
12	0	0	0	9	0	0	0	1	100	0	0	0	0	100	0.06	1.9	10	18:55:39
13	0	0	0	9	0	0	0	1	100	0	0	0	0	100	0.12	4.1	10	18:55:39
14	0	0	0	9	0	0	0	1	100	0	0	0	0	100	0.06	1.9	10	18:55:39
15	0	0	0	9	0	0	0	1	100	0	0	0	0	100	0.12	4.1	10	18:55:39
16	0	0	0	115	1	1	1	1	100	0	100	0	0	0	0.32	10.8	100	18:55:39
17	0	0	0	17	0	0	0	1	100	0	0	0	0	100	0.12	4.1	19	18:55:39
18	0	0	0	14	0	0	0	1	100	0	0	0	0	100	0.06	1.9	15	18:55:39
19	0	0	0	12	0	0	0	1	100	0	0	0	0	100	0.12	4.1	13	18:55:39
20	0	0	0	11	0	0	0	1	100	0	0	0	0	100	0.06	1.9	12	18:55:39
21	0	0	0	10	0	0	0	1	100	0	0	0	0	100	0.12	4.1	11	18:55:39
22	0	0	0	9	0	0	0	1	100	0	0	0	0	100	0.06	1.9	11	18:55:39
23	0	0	0	9	0	0	0	1	100	0	0	0	0	100	0.12	4.1	11	18:55:39
24	0	0	0	99	1	1	1	1	100	0	100	0	0	0	0.32	10.7	99	18:55:39
25	0	0	0	9	0	0	0	1	100	0	0	0	0	100	0.12	4.1	10	18:55:39
26	0	0	0	9	0	0	0	1	100	0	0	0	0	100	0.06	1.9	10	18:55:39
27	0	0	0	9	0	0	0	1	100	0	0	0	0	100	0.12	4.1	10	18:55:39
28	0	0	0	9	0	0	0	1	100	0	0	0	0	100	0.06	1.9	10	18:55:39
29	0	0	0	9	0	0	0	1	100	0	0	0	0	100	0.12	4.1	10	18:55:39

```
30    0    0    0    9    0    0    0    1 100    0    0    0    0 100 0.06    1.9    10 18:55:39
31    0    0    0    9    0    0    0    1 100    0    0    0    0 100 0.12    4.1    10 18:55:39
32    0    0    0    0    0    0    0    1 100    0    0 54    0   46 0.00    0.0     0 18:55:39
33    0    0    0    0    0    0    0    1 100    0    0    0    0 100 0.00    0.1     0 18:55:39
34    0    0    0    0    0    0    0    1 100    0    0    0    0 100 0.00    0.1     0 18:55:39
35    0    0    0    0    0    0    0    1 100    0    0    0    0 100 0.00    0.1     0 18:55:39
36    0    0    0    0    0    0    0    1 100    0    0    0    0 100 0.00    0.1     0 18:55:39
37    0    0    0    0    0    0    0    1 100    0    0    0    0 100 0.00    0.1     0 18:55:39
38    0    0    0    0    0    0    0    1 100    0    0    0    0 100 0.00    0.1     0 18:55:39
39    0    0    0    0    0    0    0    1 100    0    0    0    0 100 0.00    0.1     0 18:55:39
40    0    0    0    0    0    0    0    1 100    0    0    6    0   94 0.00    0.0     0 18:55:39
41    0    0    0    0    0    0    0    1 100    0    0    5    0   95 0.00    0.0     0 18:55:39
42    0    0    0    0    0    0    0    1 100    0    0    4    0   96 0.00    0.0     0 18:55:39
43    0    0    0    0    0    0    0    1 100    0    0    5    0   95 0.00    0.0     0 18:55:39
44    0    0    0    0    0    0    0    1 100    0    0    4    0   96 0.00    0.0     0 18:55:39
45    0    0    0    0    0    0    0    1 100    0    0    4    0   96 0.00    0.0     0 18:55:39
46    0    0    0    0    0    0    0    1 100    0    0    4    0   96 0.00    0.0     0 18:55:39
47    0    0    0   17    0    0    0    1 100    0    0 41    0   59 0.00    0.0    18 18:55:39
ALL   0    0    0  892  128   14    4   49 100  288   32    0    0   68 3.02  151.1   852 18:55:39
```

This workload could be compacted to run mostly on the first two VPs, if the system were to run with vpm_throughput_mode=2, thus reducing the PC of the system.

To compare POWER9 processor-based systems to previous models, use the following documents:

[IBM Power Systems Performance Capabilities Reference](#)
or
[IBM Power Systems Performance Report](#)

For best performance, we recommend that the system be fully populated with DIMMs, rather than only partial population of DIMMs. This will allow the hypervisor a better chance to place the LPARs in an optimal location on the system.

Improving the LPAR placement will also improve the latency of most workloads, as it will make optimal use of the memory and CPU resources on the system.

For small partitions, it is best to contain the partition on a single SRAD (Scheduler Resource Allocation Domain) when possible. The hypervisor attempts to place the LPAR in an optimal location, based on:

- CPU entitlement assigned to the partition
- Amount of memory allocation to the partition
- I/O devices allocated to the partition
- Available (free) memory / CPU on the available SRADs

11

More detailed information regarding the above can be found in the [IBM Power Virtualization Best Practices](#).

An easy way to check at a high level the placement of the partition you can run the following commands:

	Command to run	Sample output

AIX	`lssrad -av`	<pre>REF1 SRAD MEM CPU
		0
		        0 63722.19 0-63
		        1 63495.00 64-127
		        2 63495.00 128-191
		        3 63495.00 192-255
		1
		        4 63495.00 256-319
		        5 63472.00 320-383
		        6 63744.00 384-447
		        7 63738.44 448-511</pre>

IBM i	`rmnodeinfo macro`	<pre>Node statistics for 2 nodes across 2 node groups Node #	0	1 |======================================= Hardware node ID	1	2 | Node group ID	0	1 | Hardware group ID	0	2 | # of Logical procs	16	16 | # of procs folded	8	16 | # main store pages	000748DB	0007B8A8 |</pre>

Placement of the LPAR can be improved, when possible, by running the optmem command from the HMC which manages this system.

7 I/O Considerations

If you are using the same network and storage adapters and the same storage subsystem configuration as your previous system, initially, the same tuning should be used on the new system. If additional performance is desired from the existing system, then normal network and storage tuning should be performed.

If the I/O subsystems have not changed, or the storage and network devices that connect to those adapters are capable of higher throughput or lower latency than the previous subsystem components, then the influence the I/O subsystems exert on the perceived speed of the applications will either be imperceptible or possibly show improved speed.

If the network or storage subsystems are appreciably different on the newer system than the prior system, the following list of considerations could negatively impact the perceived speed of applications.

- Devices that connect to adapters:
  - Changing from Direct Attached Storage (DAS or internal) to Storage Area Network (SAN) or Network Attached Storage (NAS) (or external storage) can increase latency. Less dedicated write cache space resulting in lower write cache efficiency, longer I/O data path lengths and the potential of other activity sharing the network and storage servers may cause increased latency.
  - Higher speed networks. Many users are moving up to 25, 50 or 100 GigE networks which require revisiting tuning.
  - Additional functions such as compression, encryption and deduplication can add latency.
  - Refer to tuning or setup guides for the new devices to understand these impacts.
- Configurations:
  - Network and storage fabric topology changes can result in slower or fewer number of paths.
  - Different storage protection levels result in differing performance impacts on certain workloads. For example, RAID 10/1 (or mirroring) has less negative impact on storage write performance than RAID5, which has less impact than RAID6.
  - Many users are implementing higher level of security which may require IPsec or other security protection. These can negatively affect network bandwidth and latency.
- Sizing:
  - Reducing the number of Storage LUNs can reduce resources in the server needed to support required throughputs. For example, do not size storage subsystems on capacity alone. Do take into account performance of both logical and physical devices, thus the number of them.
  - Storage server sizing tools such as DiskMagic may make estimates without taking into account effects of server impacts like internal bus bandwidth limitations and/or latencies. Derating their output by some percentage, 5-10%, is advisable if the server impacts are unknown.
- Virtualization:
  - Virtualization does add latency and can reduce throughput compared to native I/O. Besides the backend hardware, ensure VIOS memory and CPU

amounts are enough to provide the required throughput and response times.

- o Moving to higher speed virtualized network adapters in VIOS will require adjusting the VIOS configuration in CPUs and memory.
- o IBM PowerVM Best Practices Redbook at http://www.redbooks.ibm.com/abstracts/sg248062.html can be helpful sizing VIOS.
- AIX Specific Tuning Effects:
  - o Maximizing the number of processors used to handle I/O completion interrupts promotes higher throughput capabilities. It is advised to set the 'Desired processors' and 'Maximum processors' value in the partition's profile as defined in an HMC to power of 2 values to better enable AIX or VIOS to use more processors for I/O interrupts.
  - o Storage: Ensure the per device and per adapter queue depths and number of channels per adapter/port are sufficiently high enough to overcome the I/O path latency to be able to reach desired throughputs.
    - SSDs that attach with the NVMe protocol have large command and response queues called 'channels'. It is recommended to increase the number of channels for workloads that require high command throughput. 'High' is defined as at or approaching the device's io/s limit for the storage I/O workload that the application generates. Increasing the number of queues will not typically lower response times, nor will it increase throughput for storage I/O workloads that stress high data throughputs, where units are GB/s, and the I/O lengths are typically 16KB/io or larger.

      The number of NVMe channels, or 'nchans', can be altered via smitty menus or the command line. For smitty menus use: "Devices", then "NVMe Manager", then "Change / Show Characteristics of a NVMe Controller", then choose your drive, then the "Number of Channels" field can be selected and altered. The following command line will also work.
      - chdev -l nvmeX -a nchan=8 -P (where X = 0, 1, etc)

    - NVMe devices have a virtualization layer built into their controllers that allows the physical storage space to be partitioned into multiple logical devices called 'namespaces'. Partitioning into more logical devices can allow the host OS or VIOS to utilize more CPU and memory resources for the I/O and allow more parallelism, which can, depending upon a myriad of variables, allow the host to take better advantage of the high throughputs

> SSDs can provide. Namespaces can be created and deleted from the NVMe Manager screens.
> - o Network: For higher speed >=25 GigE adapters the default settings for number of transmit and receive queues and entries in the queues are a starting point. For 100 GigE adapter please consult the 100 GigE adapter tuning guide.
> - o IPsec: If you are not setting any IPsec rules then IPsec should be turned off. Due to the architecture of IPsec, increasing the number of transmit and receive queues on an adapter interface may not improve single client IPsec performance.

- IBM i Specific Effects:
  - o Be aware that storage devices that support IBM i's native block length, 4160 or 520 bytes, can result in more efficient I/O (less CPU usage per I/O) than storage devices that only support 4096 or 512 byte block lengths.

8 Consider Lab Services engagement and/or benchmark

The IBM Systems Lab Services organization https://www.ibm.com/it-infrastructure/services/lab-services is available to assist you with resolving system, application, and database performance problems. Formal and informal training opportunities are also available where you learn how to use the performance tools and resolve performance problems on your own.

If you need additional help in assessing the potential impact of a system migration, benchmarking a system environment, or identifying ways to improve the performance of your environment, please contact IBM Lab Services at ibmsls@us.ibm.com.

9 Migration checklist
- Plan the migration.
- Install latest required software, apply the available fixes.
- Set appropriate processor compatibility mode for logical partitions before and after migration.
- Plan the virtual processor and entitlement for logical partition to best fit your operation and performance requirement.
- Follow I/O consideration guide.
- Consider engagement with IBM Systems Lab Services as described in Section 8.

10 Appendix

10.1 VPM Folding Threshold Validation Script

If you are using LPM to migrate the LPAR and are unable to update the LPAR with the CPU folding fixes shown in section 4.1, or are unable to reboot the LPAR after applying the fixes, you can use the following script after the LPAR has been migrated via LPM to the POWER9 system:

```
#!/usr/bin/ksh

curr_fold_threshold=`echo "dw schedp+20"|kdb -script | awk '/schedp+/ {printf "%d\n", "0x"$2}' |
tail -1`
curr_fold_delta=`echo "dw schedp+44"|kdb -script | awk '/schedp+/ {printf "%d\n", "0x"$2}' | tail
-1`
curr_normalized=$(( curr_fold_threshold + curr_fold_delta ))

curr_smt_mode=`smtctl | awk '/^proc0/ {print $3}'`
PROC_TYPE=`lsconf | awk '/PowerPC/ {print $3}'`

printf "%-30s: %s in SMT%s\n" "The Power platform is" $PROC_TYPE $curr_smt_mode
printf "%-30s: %s %s %s\n" "Kernel values"  $curr_fold_threshold  $curr_fold_delta
$curr_normalized
printf "%-30s: %s %s %s\n" "schedo current" `schedo -o vpm_fold_threshold`

case $PROC_TYPE in
    PowerPC_POWER8 )
        if [[ curr_smt_mode -eq 8 ]]; then
            rec_fold_delta=4
            rec_fold_threshold=45
        else
            rec_fold_delta=0
            rec_fold_threshold=49
        fi;;

    PowerPC_POWER9 )
        if [[ curr_smt_mode -eq 8 ]]; then
            rec_fold_delta=23
            rec_fold_threshold=26
        elif [[ curr_smt_mode -eq 4 ]]; then
            rec_fold_delta=13
            rec_fold_threshold=36
        elif [[ curr_smt_mode -eq 2 ]]; then
            rec_fold_delta=9
            rec_fold_threshold=40
        elif [[ curr_smt_mode -eq 1 ]]; then
            rec_fold_delta=0
            rec_fold_threshold=49
        fi;;
    * )
        printf "%-30s: %s\n" "Unknown platform, exiting"
        exit;;
esac

rec_normalized=$(( rec_fold_threshold + curr_fold_delta ))

printf "%-30s: %s %s %s\n" "schedo recommend" "vpm_fold_threshold = $rec_normalized"

if [[ curr_fold_threshold -ne rec_fold_threshold ]]; then
    printf "%-30s: %s %s %s\n" "Recommend" "schedo -o vpm_fold_threshold=$rec_normalized"
else
    printf "%-30s: %s %s\n" "Recommend" "No change.  VPM fold threshold already" $rec_normalized
fi
```

IBM ®