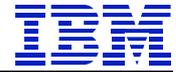


IBM SAP Technical Brief

**SAP on Power Systems
Best Practices Guide**

**IBM SAP International Competence Center
Walldorf, Germany**

**Version: 1.0
Status: November 2012**



Note: Before using this information and the product it supports, read the information in “Notices” on page 3.

Table of Contents

NOTICES	3
TRADEMARKS	4
PREFACE	5
EDITION NOTICE	5
SCOPE AND AUDIENCE	5
AUTHORS	5
SPECIAL THANKS TO.....	5
FEEDBACK	5
1 INTRODUCTION	6
2 BEST PRACTICES RECOMMENDATIONS	8
2.1 SYSTEM FIRMWARE UPDATES	8
2.2 AIX UPDATES	8
2.3 IBM I UPDATES	9
2.4 RESTRICTED AIX TUNABLES AND I/O PACING.....	9
2.5 PERFORMANCE ADVISORS	9
2.6 RUN PARTITIONS IN POWER7 MODE.....	10
2.7 RUN PARTITIONS WITH SMT ON.....	11
2.8 PARTITION PLACEMENT.....	11
2.9 CHECKING FOR LOCAL PROCESSORS AND MEMORY.....	12
2.10 DYNAMIC PLATFORM OPTIMIZER	13
2.11 PARTITION AND SYSTEM SIZING	13
2.12 VIO SERVER PROCESSOR CONFIGURATION	15
2.13 VIRTUAL ETHERNET PERFORMANCE.....	15
2.14 VIRTUAL PROCESSOR FOLDING	15
2.15 POWER7 PROCESSOR UTILIZATION METRICS	16
2.16 CAPACITY VERSUS RESPONSE TIME (LATENCY)	16
2.17 OPTIMIZED SAP KERNEL 7.21-EXT FOR AIX	17
2.18 OPTIMIZED SAP KERNEL 7.20-EXT AND 7.21-EXT FOR IBM I	17
2.19 SAP EXTENDED MEMORY MANAGEMENT ON AIX	17
2.20 SAP EXTENDED MEMORY MANAGEMENT ON IBM I.....	17
2.21 POSIX AND OLSON TIME ZONES.....	18
2.22 ORACLE DATABASE SERVER ISSUES.....	18
2.23 MAIN STORAGE POOLS ON IBM I.....	18
3 BRIEF SUMMARY OF IMPORTANT RULES	20
4 RESOURCES	21
4.1 WHITEPAPERS.....	21
4.2 WEB SITES/BLOGS.....	21
4.3 REDBOOKS.....	21
4.4 PRODUCT DOCUMENTATION.....	22
4.5 SAP DOCUMENTATION.....	22
4.6 IMPORTANT SAP NOTES.....	23

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This 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 may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

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

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

Trademarks

IBM, the IBM logo, and [ibm.com](http://www.ibm.com) are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at <http://www.ibm.com/legal/copytrade.shtml>

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

Active Memory™	Power Systems™	pSeries®
AIX 5L™	POWER4™	Redbooks®
AIX®	POWER5™	Redbooks (logo)®
DB2®	POWER6+™	System i®
IBM®	POWER6®	System p®
iSeries®	POWER7™	System x®
Micro-Partitioning™	PowerHA™	Systems Director VMControl™
OS/400®	PowerVM™	Tivoli®
POWER Hypervisor™	POWER®	Workload Partitions Manager™

The following terms are trademarks of other companies:

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

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

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

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

Novell, SUSE, the Novell logo, and the N logo are registered trademarks of Novell, Inc. in the United States and other countries.

Oracle, JD Edwards, PeopleSoft, Siebel, and TopLink are registered trademarks of Oracle Corporation and/or its affiliates.

ABAP, SAP EarlyWatch, SAP GoingLive, SAP NetWeaver, SAP, and SAP logos are trademarks or registered trademarks of SAP AG in Germany and in several other countries.

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.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both. Other company, product, or service names may be trademarks or service marks of others.

Preface

Edition Notice (November 2012)

First Edition

Scope and Audience

People reading this document should have a comprehensive background in PowerVM virtualization topics, IBM AIX or IBM i performance tuning, and SAP Basis administration and configuration. The document addresses customer personnel and IBM practitioners (from pre- & post-sales and Services, GBS/GTS, ATS, eTS) involved in the planning, implementation, and performance monitoring and analysis of SAP landscapes on IBM Power Systems. We assume that the reader is familiar with PowerVM virtualization terminology as for example logical partitions, physical processors, virtual processors, logical processors, shared processor pool, entitlements, etc. We will not explain the detailed meaning of each PowerVM term mentioned in this paper and refer to the available Redbooks (see references section) for a detailed introduction in PowerVM technologies.

Authors

Walter Orb, IBM SAP International Competence Center

Special Thanks To

- Olaf Rutz, IBM Research and Development
- Christian Bartels, IBM Research and Development
- Dirk Michel, IBM STG Cross Platform Systems Performance
- Peter M Jensen, IBM ISV Business Strategy and Enablement
- Ron Schmerbauch, IBM Rochester Development Lab

Feedback

We are interested in any feedback you have. Please send your comments to isicc@de.ibm.com.

Disclaimer

This document is subject to change without notification and will not comprehensively cover the issues encountered in every customer situation. It should be used only in conjunction with the product literature accompanying the products listed above. The information contained in this document has not been submitted to any formal IBM test and is distributed AS IS.

1 Introduction

Many SAP customers are exploiting the benefits of the advanced virtualization technologies provided by PowerVM on IBM Power Systems. IBM Power Servers are available in a broad range of configurations starting with small blades or 1U rack mounted servers with four processor cores and a few Gigabyte of main memory all the way up to large servers with 256 processor cores and multiple Terabytes of main memory. Each of those configurations allows for the full usage of the available PowerVM virtualization technologies.¹

The most popular and in many cases preferred setup for SAP system landscapes today is to configure fully virtualized environments running with

- Dual VIO servers to virtualize I/O connectivity and provide redundancy for I/O component failures and VIO server maintenance
- Fully virtualized partitions running uncapped to fully share the available physical processor cores in a shared processor pool

When defining and subsequently activating uncapped shared processor partitions, there are only very few restrictions in place²

- Defining a desired number of virtual processors sets the maximum amount of capacity a partition can use.
- Defining the desired entitlement sets the guaranteed amount of capacity. The POWER Hypervisor will ensure that the partition will always get at least that amount of capacity (assuming that it needs it), even if the shared processor pool is fully utilized.
- For each defined virtual processor, you need to set a minimum desired entitlement of 0.1 processor cores.

The hypervisor does not provide any additional safeguards in this area. For example, it is possible to define a large production partition with 32 virtual processors (which depending on the server model already provides capacity of more than 100,000 SAPS) with an entitlement of as little as 3.2 processor cores. As long as there is still unused entitled capacity available, the hypervisor will activate such a partition and will not warn you that the system might be heavily utilized already.

Unfortunately, this approach has become a common practice, using the assumption that together with another partition attribute called “Weight”, the hypervisor will always favor this partition and allow for unanticipated (or maybe quite anticipated) growth of a factor of 10. This practice simplified the setup of new partitions, however it was never a best practice and it has certainly become more problematic with POWER7 systems. Relying on such a setup means that your most important partitions might not always get the processor cycles when they need them. This leads to potential performance problems. This is especially true for the VIO servers.

¹ Based on the purchased PowerVM license features and the size of the systems, some restrictions may apply to certain configurations.

² In this document, we do not specifically deal with partition configuration parameters like minimum and maximum virtual processors or entitlements. Those are mostly important for DLPAR operations or at the time of the activation of a partition, when you are already using a major percentage of the available resources. Therefore, when we talk about number of virtual processors or processor entitlement, we assume that this is the desired value and that the partition would get this desired value during partition activation.

The intent of this whitepaper is to provide a number of recommendations and best practices to setup POWER processor based systems for SAP landscapes. The majority of the recommendations is not really SAP specific and would apply to many other applications. Some of the recommendations are specific for POWER7 based systems, but most of the considerations will equally apply to systems based on older POWER processor generations too.

The document is a loose collection of recommendations and hints, the order of the presented topics should not imply any level of importance or priority. We recommend studying the complete document and considering all of the recommendations.

The focus of this paper is on the PowerVM related system configuration aspects. It is intentionally kept short to provide a concise summary of the most important topics, so that one can use it as a checklist for system setup and for future planning considerations. A number of topics are already explained in detail in other papers or at other places. In these cases, we provide a short summary and reference the more detailed documentation.

2 Best Practices Recommendations

2.1 System Firmware Updates

The POWER Hypervisor is integrated with all Power Servers and is part of the Power Systems firmware. The hypervisor manages system virtualization and is makes decisions about logical partition placement at activation and during dynamic LPAR operations. The first POWER7 machines shipped in 2010 introduced a number of significant changes compared to previous generations (more cores per CPU, SMT4, larger systems, etc.). Since the initial shipment, a lot of experience was gained through customer implementations especially with running large workloads in partitions with dozens of configured virtual processors and large memory footprints. Based on these experiences the POWER and OS developers have enhanced overall system performance by improving performance relevant areas in the hypervisor and the OS kernels and have fixed a number of bugs.

We strongly recommend that you check your current system firmware levels and plan to upgrade if you are running with a lower version than the current recommended minimum firmware level, which is 730_078 or higher at the time of the writing of this paper. On AIX the firmware level is viewed with the command `lsattr -El sys0 -a fwversion`, while on IBM i it is viewed with the command `DSPFMWSTS`.

One can check the firmware READMEs with a version history for the current system generation at the following web sites (unfortunately, they typically have only short one-line descriptions for each of the changes made):

Power 710 to750	ftp://ftp.boulder.ibm.com/software/server/firmware/AL-Firmware-Hist.html
Power 770/780	ftp://ftp.boulder.ibm.com/software/server/firmware/AM-Firmware-Hist.html
Power 795	ftp://ftp.boulder.ibm.com/software/server/firmware/AM-Firmware-Hist.html

We also recommend using the Fix Level Recommendation Tool (FLRT) regularly to check the current recommended firmware and software levels for your system landscape. Please find this tool at <http://www.ibm.com/support/customercare/flrt/>.

2.2 AIX Updates

Just the same as with system firmware updates mentioned in the previous section, there are a number of performance optimizations and fixes for some specific problems included in later AIX technology and support package levels. POWER7 servers were announced with support for AIX 5.3, AIX 6.1, and AIX 7.1 at certain minimum technology and support package levels. Theoretically, one could still run your POWER7 systems with AIX 5.3, but first AIX 5.3 is out of support since April 2012, and second AIX 5.3 was only supported while running the POWER7 processors in POWER6 compatibility mode. In this mode, the processors are running in SMT2 mode only meaning that you cannot exploit the full capacity of a POWER7 core. Finally, the performance enhancements mentioned above were included in AIX 6.1 or AIX 7.1 only.

So please make sure that you are using the following minimum recommended AIX software levels (once again, these are the minimum recommendations at the time of the writing of this paper and they could change again over time):

AIX 6.1	TL7 SP2 or later
AIX 7.1	TL1 SP3 or later

2.3 IBM i Updates

POWER6 servers are supported with IBM i V5R4M0 (V5R4M5 for the Licensed Internal Code), IBM i 6.1 and IBM i 7.1. POWER7 servers are supported with IBM i 6.1 (V6R1M1 for the Licensed Internal Code) and IBM i 7.1. Further enhancements to IBM i 7.1 are being delivered in certain intervals through so-called Technology Releases (TR). In particular IBM i 7.1 Technology Release 3 is containing several functional and performance enhancements for SAP.

A description of the changes in IBM i 7.1 TR3 is available in the IBM Techdocs library at: <http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/FLASH10763>

Functional errors in IBM i are corrected through Program Temporary Fixes (PTFs). IBM updates so-called “Informational APARs” on a regular basis with a list of PTFs that are required when running SAP on IBM i. Depending on your release level, you can find the list of required PTFs in the following Informational APARs:

V5R4M0:	II14126
V5R4M5:	II14319
6.1:	II14362
6.1.1:	II14535
7.1:	II14536

2.4 Restricted AIX Tunables and I/O Pacing

Do not change any restricted tunables without discussing the change with AIX support first. If you still have some systems running at AIX 5.3, do not carry forward tunables when you migrate to AIX 6.1 or AIX 7.1. A number of tunable settings that had to be adapted with AIX 5.3 have become the default settings with AIX 6.1 and AIX 7.1 (see SAP notes 1048686 and 973227) and in general, there is no need to change those tunables anymore. During review of the system when planning an AIX upgrade, one might find some other tunables not mentioned in the above SAP notes. Those might have been changed based on recommendations from IBM to resolve a performance or some other system issue. In this case, please contact IBM support prior to the upgrade to verify whether those changes would still be applicable to the new release.

In the past it was also quite common to change the I/O pacing settings for minpout and maxpout. Please verify the current settings on AIX 6.1 or AIX 7.1 using the `lsattr -El sys0 -a minpout,maxpout` command. Those values should rarely require changes to the default values of 4096/8193. Do not change them without contacting IBM support first; the older legacy values of 24/33 can result in some severe I/O performance issues.

2.5 Performance Advisors

The IBM Power Systems Virtualization and AIX performance teams developed some tools to help customers to analyze performance characteristics and get a health check report for their AIX and VIOS partitions.

Currently there are three performance advisors available:

VIOS Performance Advisor

The VIOS advisor is an application that runs within the customer's VIOS for a user specified amount of time (hours). It polls and collects key performance metrics before analyzing results and providing a health check report and proposes changes to the environment or areas to investigate further. The goal of the VIOS advisor is to enable the customer to make assessments about the status of the VIO server and implement recommendations based on the expertise and experience available from the IBM systems performance group.

Please follow the instructions at the following web site to download and run the VIO Performance Advisor for all VIOS related performance issues:

<http://www.ibm.com/developerworks/wikis/display/WikiPtype/VIOS+Advisor>

PowerVM Virtualization Performance Advisor

The PowerVM Virtualization Performance Advisor diagnoses performance issues within PowerVM partitions. The advisor runs within an AIX partition (dedicated or shared), analyzes performance characteristics, and provides a health check report. The user defines how long the tool runs. The health check report includes system configuration (available resources, hardware & software levels, etc), partition (type, mode, available resources, s/w levels), partition settings and recommendations to improve performance (e.g. placement, SMT, processor folding, processor, and memory pool utilization).

<http://www.ibm.com/developerworks/wikis/display/WikiPtype/PowerVM+Virtualization+performance+advisor>

JAVA Performance Advisor

The third performance advisor provides recommendations to improve the performance of Java applications on AIX. Since October 2012, the only supported JVM for SAP NetWeaver based systems is SAP's own developed SAP JVM. The Java Performance Advisor targets the IBM JVM only and does not work with SAP JVM. Therefore, it is referenced here only for completeness:

<http://www.ibm.com/developerworks/wikis/display/WikiPtype/Java+Performance+Advisor>

2.6 Run Partitions in POWER7 Mode

One of the available strategies to migrate running partitions from older POWER6 to POWER7 servers is to use Live Partition Mobility (LPM) to migrate your running systems without taking any downtime. When the migrated partition resumes on the target POWER7 server, it will run in POWER6 compatibility mode. This means that the processors will run only in SMT2 mode and the additional capacity provided by the other two hardware threads when running the processor in SMT4 mode is not available. The partition needs to stay in POWER6 compatibility mode if LPM will be used to move it back and forth between POWER6 and POWER7 servers. It is ok to run with this setup for a transitional period. However to fully exploit the capacity of POWER7 processors, downtime should be planned to shut down and re-activate the partition. Before activating the partition again, one should check that the *Processor compatibility mode* in the partition profile is set to *default* or *POWER7* to ensure that it starts in POWER7 mode.

2.7 Run Partitions with SMT on

Often we get the question whether to turn SMT off, as a SAP ABAP system is not a multi-threaded application. While this is true, the NetWeaver basis environment itself is a multi-process implementation and works very well with multiple SMT threads.

We recommend to run with the default settings of SMT2 for POWER6 based systems and SMT4 for POWER7. As long as there is more than one user or batch job actively using the SAP instance, there is more than one runnable process in the system. The AIX or IBM i dispatcher can schedule these processes in parallel on separate SMT hardware threads just the same as they would schedule multiple software threads of a multi-threaded application.

Running more than one software thread on a single core can have the effect of each single software thread running slightly slower than they would when they had a complete core for themselves. It does not matter whether the application is multi-threaded or single-threaded, when there is more than one thread actively running on a core, one might notice the software threads running slightly slower.

The OS dispatchers currently favor scheduling workloads to the primary threads of each virtual processor before starting to use the secondary and tertiary SMT threads.

When running with low CPU utilization, the cores will automatically run in single threaded (or ST) mode and one should not see a major difference compared to a partition where SMT is turned off completely. Once load starts to increase on the system, the second thread is used and things might run a bit slower, but in most cases, this is not an issue. Where SMT really helps is when the system is running at utilization levels of 90% or more while still providing good response times. Running the same load with SMT disabled would most likely cause the response times to increase sharply.

Turning SMT off might improve response times. In general, it does not make a significant difference. The downside to turning SMT off is that the system would need 60-70% more processor cores to provide the same amount of capacity.

2.8 Partition Placement

Current versions of the PowerVM hypervisor, AIX, and the IBM i operating system have implemented enhanced affinity further optimizing performance for workloads running in a virtualized shared processor pool partition.

Please review the POWER7 Virtualization Best Practice Guide available at:

https://www.ibm.com/developerworks/wikis/download/attachments/53871915/P7_virtualization_bestpractice.doc?version=1

Most of this guide is applicable to all Power OS's, although at the current time it has extra AIX information.

The document explains how the hypervisor handles virtual processors and processor entitlement. It provides some hints about sizing and configuration of virtual processors, explains the processor folding feature, and covers the hypervisor strategies to achieve a higher memory affinity when placing the resources at partition activation time.

As a companion reading to this best practices guide, we also recommend to check the “Local, Near & Far Memory series” on AIXpert blog that is available on IBM developerWorks:

<https://www.ibm.com/developerworks/mydeveloperworks/blogs/aixpert/?lang=en>

Please scroll down the pages on that web site and look for the blog titles “Local, Near & Far part 1 to part 11” and “Local, Near & Far POWER7 Affinity Nine Conclusions”.

The blog series provides a detailed introduction in POWER7 affinity domains and the operating system and hypervisor strategies to schedule runnable threads in order to achieve the optimum performance.

2.9 Checking for Local Processors and Memory

The *lssrad* command allows for checking the partition placement of your partition. The *lssrad* output shows whether a partition spans multiple drawers or books. Partitions that span multiple drawers might be susceptible to some memory latency issues. If partitions span multiple drawers and could be re-balanced to run on a single drawer, then the system should show some performance improvements.

The following is an output from a well-balanced system; each scheduler resource affinity domain has 32 logical processors defined (8 cores with SMT4) and has about 28 GB of local memory available.

```
# lssrad -av
REF1    SRAD      MEM      CPU
0
      0  29889.19   0-31
      1  29868.00  32-63
      2  29865.94  64-95
      3  27242.00  96-127
1
      4  29288.00  128-159
      5  29288.00  160-191
      6  29272.00  192-223
      7  29272.00  224-255
```

The next two examples show some problematic configurations.

The first one shows some processors in SRAD 2 with no memory. Software threads running on CPU 64 to 79 will always have to access distant memory, which increases the latency for accesses to all data that is not available in the processor caches.

```
# lssrad -av
REF1    SRAD      MEM      CPU
0
      0  51618.00   0-31
      1  47808.00  32-63
1
      2     0.00  64-79
```

The second output shows SRAD 3 and 4 with some memory but no processors defined. Software threads that need to access data stored in SRAD 3 and 4 will experience some additional memory access latency as those accesses will be to distant memory.

```
# lssrad -av
REF1    SRAD      MEM      CPU
0
      0  29115.50   0-31
```

1	1	12948.00	32-47
	2	31093.00	48-79
	3	14940.00	
2	4	249.00	

In both cases the partition placement should be reviewed (refer to section 2.8).

2.10 Dynamic Platform Optimizer

To address some of the challenges associated with partition placement and memory affinity, IBM developed a new feature called Dynamic Platform Optimizer (DPO). Dynamic Platform Optimizer allows an administrator to re-balance system resources without requiring an outage. These enhancements require a minimum firmware level of 760 and the latest version of VIOS (2.2.2). The minimum HMC Code level for this firmware is HMC V7 R7.6.0. This firmware is currently available on select POWER servers.

Please check the IBM Power Systems Hardware Information Center, the firmware description, and the “POWER7 Virtualization Best Practice Guide” (see section 4.1 Whitepapers) for more information about the Dynamic Platform Optimizer:
<http://pic.dhe.ibm.com/infocenter/powersys/v3r1m5/index.jsp?topic=/p7ha2/otheradvancedfunctionsondemand.htm>
ftp://ftp.boulder.ibm.com/software/server/firmware/01AM760_034_034.html

2.11 Partition and System Sizing

Please perform thorough capacity planning and sizing exercises especially for your most important production systems. Try to focus on your daily or weekly peak hour requirements and look at the average system usage during that hour. The partition entitlement should then be set to a value that covers this average. For the virtual processor definition, round the entitlement up to the next whole number and then add a few more virtual processors to cover the occasional peak. For smaller partitions, this would probably be one or two additional virtual processors, but this could also be a larger number for partitions with a large entitlement.

Another way to look at this is to build the ratio between the capacity entitlement (CE) and the number of defined virtual processors (VP) for a partition.

(CE/VP)	Balance CE and VP
0.1	VP = 10.0 x CE
0.2	VP = 5.00 x CE
0.3	VP = 3.33 x CE
0.4	VP = 2.50 x CE
0.5	VP = 2.00 x CE
0.6	VP = 1.67 x CE
0.7	VP = 1.43 x CE

0.8	VP = 1.25 x CE
0.9	VP = 1.11 x CE
1	VP = 1.00 x CE

A CE/VP ratio of 0.2 means that for every virtual processor defined, there is a capacity entitlement of 0.2. An example for this would be a partition with an entitlement of 4.0 and 20 virtual processors. With such a setting, the partition could grow by a factor of five assuming it can get the resources from the shared processor pool.

Looking at the yellow values (CE/VP from 0.1 to 0.5), the performance for uncapped partitions might not be safe. One can use these settings for less important test and development systems, but do not use it for response time sensitive production partitions.

Performance for uncapped partitions gets better with an increasing CE/VP ratio and the most important partitions should run with a ratio of 0.8 or higher.

Running partitions with a CE/VP ratio of one looks almost like a configuration with dedicated processors. However, such a partition setup will still use the shared processor pool and give up unused capacity that other partitions can still exploit. So this can certainly make sense for some configurations.

Once the partitions are right-sized, one has to make sure that the combined entitlements fit in the available shared processor pool. On running partitions, one should monitor the values for entitled capacity and available capacity in the processor pool (for example *lparstat* reports these figures as *%entc* and *app*). Ideally, the values for *%entc* should be below 100% and *app* should not drop below 1.0.

Some customers might have occasional peak capacity requirements (monthly, quarterly, yearly) that are significantly higher than accounted for with the settings based on daily or weekly averages. In such a case, we do not recommend to run with these maximum settings during all times. It is preferable to use the maximum entitlement and number of virtual processor settings in the partition profile to allow for this additional growth and then use DLPAR operations to add the resources at the times needed. Of course, the shared processor pool must be large enough to allow for the additional required capacity. At the end of the peak, one should remove the added resources again. Please consider these requirements in continuous capacity planning exercises. Another possibility to free up capacity in the shared processor pool would be to shut down less important partitions for the duration of those peak hours or use Live Partition Mobility to move them temporarily to another server.

Very often partitions are defined with a larger number of virtual processors than needed to allow for potential future growth. Same as with the monthly or yearly peaks it is preferable to configure the system for daily or weekly peak requirements and allow some headroom in the partition profiles for future growth. Then use DLPAR operations to add virtual processors and capacity entitlements when needed. This should be part of a routine monitoring, capacity verification and planning exercises.

A word of caution: please recheck the partition placement each time after performing DLPAR operations that add or remove large amounts of resources (see section 2.8).

Finally, when configuring the number of processors for individual partitions, they should not exceed the number of physical processor cores in the system. In fact, if the partition is running in a shared processor pool, the number of virtual processors defined should not exceed the number of physical processor cores in the shared processor pool. The hypervisor will not

warn about creating such a configuration, but it is not possible to use more virtual processors simultaneously than physical processor cores available in the system or the shared processor pool. Therefore, such a configuration does not make any sense and creates unnecessary overhead.

2.12 VIO Server Processor Configuration

Even more important than the production partitions are the VIO servers as they provide the bridge to disk subsystems and the outside network. SAP implementations are typically communication intensive and higher network latencies will easily affect end-user response times by adding time to the front-end communication itself or also increasing the database request time in 3-tier implementations. For many business processes there could be also additional communication between multiple SAP systems, which will also add to the end user response time. Network traffic requires CPU on the VIO servers, which is sensitive to potentially delayed scheduling of virtual processors in cases where the entitlement per virtual processor is set too low.

Running 3-tier systems with application and database servers on different systems might amplify this problem, because each round trip network packet has to go through two VIO servers twice. Let us assume the entitlement of a VIO server is set to 0.1, which means in the worst case, the VIO server might send a network packet, gets pre-empted because it has exceeded its entitlement and then has to wait nine milliseconds before the hypervisor schedules it again and can take the network interrupt. If this happens for a database request on both VIO servers on the way to the database server and then on the way back, we would add a 36 milliseconds network delay to the database response time (typically we expect a communication latency in the range of a few hundred microseconds).

This is an extreme example and it probably would never be this bad in real life, but it should just illustrate the potential problem of using too low entitlements for the VIO servers.

Therefore, we recommend configuring the virtual processors for the VIO servers as dedicated donating processors. With such a setup, the VIO server will always immediately get a physical core back when needed and it will still donate spare capacity to the shared processor pool. Customer experiences show that changing from shared processor configuration to dedicated donating processor actually reduced the number of physical cores consumed and improved the network times. For very large systems, you might even consider dedicated processors. If you still want to use virtual processors instead, then make sure they are configured with an entitlement of at least 0.8 or 0.9 per VCPU.

2.13 Virtual Ethernet Performance

Applications server and database partitions running on the same physical system communicating via virtual Ethernet adapters can also experience some unnecessary network latencies because of non-optimum settings for the virtual network adapters or virtual Ethernet adapter buffers in the hypervisor. Please review the recommendations in section 9.10 “Virtual Ethernet Performance” in the “AIX on Power – Performance FAQ” whitepaper:

<http://public.dhe.ibm.com/common/ssi/ecm/en/pow03049usen/POW03049USEN.PDF>

2.14 Virtual Processor Folding

The AIX and IBM i kernels work together with the hypervisor to dynamically increase and decrease the use of virtual processors based on the utilization of the partition. In environments with many partitions running in the shared processor pool and a large number of virtual processor configured, this feature helps decreasing the virtual context switches and the shared pool utilization.

The OS dispatchers favor to distribute the workload over the primary threads of the available cores first. From a single partition point of view, this is a good strategy because it maximizes the exploitation of available physical processor and cache resources.

With POWER7 processors, SMT4 provides more capacity on each core and this influences the way that the system utilization is calculated. Together with the strategy to favor the primary threads, this can lead to more aggressive unfolding of virtual processors as used to be the case on POWER6 or POWER5. From a single partition perspective, this still makes sense, however when running dozens of partitions in a shared processor pool one can easily end up in a situation where the ratio of active virtual processors to available physical processor cores is very high and this will have an impact on the overall system efficiency. At the same time, each partition might waste available processor resources, as they are not using the additional SMT threads.

This is another reason why one should not configure too many virtual processors for partitions. In some cases, it might actually make sense to reduce the number of virtual processors defined to force the OS dispatcher to use all available SMT hardware threads first and thus increase the overall efficiency of the system.

In addition practices that were used in the past need to be reconsidered. One common example is to look at the physical CPU consumption of a partition (some performance tools report this as *physc*) and compare this to the number of virtual processors configured to determine if a partition is CPU bound. One should never make a decision to add additional virtual processors to a partition based on the *physc* value alone and always check the actual utilization of the partition (*physb*). Please do not add virtual processors if there are still idle times available. In fact, high idle times might be an indicator to consider reducing the number of virtual processors.

Please check the AIX Information Center for some more details about virtual processor folding:

http://pic.dhe.ibm.com/infocenter/aix/v7r1/index.jsp?topic=%2Fcom.ibm.aix.prftungd%2Fdoc%2Fprftungd%2Fvirtual_proc_mngmnt_part.htm

2.15 POWER7 Processor Utilization Metrics

With previous POWER processor generations single active software threads running alone on a processor core were charged for all the resources and reported the core as 100% utilized. POWER7 introduced a change to better reflect the fact that a single software thread cannot fully utilize a POWER7 core and that there is additional capacity available on this core. A single active software thread is charged with a CPU usage of less than 100%. This change has some side effects for SAP performance statistic records. When running an ABAP intensive report in a work process that is running alone on a processor core, one will always see a difference in the reported *CPU time* and *Processing time* performance metrics.

There used to be an old rule of thumbs “If the Processing Time is two times higher (or more) than the processor time, then we likely have a processor bottleneck”. With POWER7, it is quite normal that the reported *Processing time* is lower than the *CPU time* and the above rule of thumb is not valid anymore. Please check chapter 8.6 “Processor utilization metrics” in the *SAP Applications on PowerVM* redbook (SG247564) for a detailed description of this change and its impact on SAP performance metrics.

2.16 Capacity versus Response Time (Latency)

The POWER7 processor introduced a number of architectural changes that significantly increased the available capacity on processors and systems. However, at the same time the maximum clock rate of the processors has been reduced as well (compared to POWER6).

A POWER7 with 2.5 Ghz processors blade would most likely be capable to provide an equal amount of capacity (as for example measured in SAPS) as a smaller to medium sized partition on a POWER6 595 with 5.0 Ghz processors. However, the POWER7 processors with the much lower clock rate would not be able to provide the same performance for the single software threads of that workload. The architectural changes of the POWER7 processor compensate for some of the clock rate reduction, but not for all of it. Workloads being sensitive to single thread performance criteria (for example, end user response times for critical business transactions or complex long running batch chains with certain batch jobs in the critical path, that cannot be parallelized further) should not be moved from the fastest available POWER6 processors to the slowest available POWER7 processor models. Please perform a careful analysis of capacity requirements and other critical performance criteria and then choose a target model that is able to satisfy all of those requirements.

2.17 Optimized SAP Kernel 7.21-EXT for AIX

The IBM AIX porting team has worked with SAP to improve the performance of the SAP kernel on AIX by implementing new optimization technologies for the kernel build process and adding some code changes that improve the TCP/IP communication on AIX. This kernel is available as a downward compatible kernel for NetWeaver version 7 based systems. This version can replace the kernel for most of the currently available SAP releases. Please check SAP note 1728283 “SAP Kernel 721: General Information” for further details. A number of pilot customers have already tested this kernel with a result of improved response times or a reduction of CPU consumption for existing implementations.

2.18 Optimized SAP Kernel 7.20-EXT and 7.21-EXT for IBM i

The IBM i porting team have worked together with the AIX porting team and SAP to implement new optimization technologies during the kernel build process. The optimized kernels are available on the SAP Service Marketplace for releases 7.20-EXT and 7.21-EXT and can be used as downward-compatible kernels for all systems based on SAP NetWeaver 7. They require a minimum operating system level of IBM i 7.1. Information about the 7.20-EXT kernel is available in SAP notes 1553300 and 1553301, information about the 7.21-EXT kernel can be found in SAP Note 1716826.

2.19 SAP Extended Memory Management on AIX

AIX provides a number of functions in the kernel that allows for an alternative implementation of the SAP extended memory management. Sometimes this implementation is called SHM_SEGS implementation, because it is enabled with the following instance profile parameter.

```
ES/TABLE = SHM_SEGS
```

We strongly recommend using this implementation instead of the standard extended memory implementation, because it provides much better performance. Unfortunately, the SAP installation tool sapinst does not set this as default during an installation on an AIX partition. Please check the current settings of ES/TABLE and follow the instructions in SAP notes 789477 and 856848 to enable the SHM_SEGS extended memory management.

2.20 SAP Extended Memory Management on IBM i

Like the SAP implementation on AIX, the SAP implementation on IBM i allows for two different implementations of the SAP extended memory management. On IBM i, the shipped default is ES/TABLE = UNIX_STD. On smaller servers, no significant difference in performance could be seen, but on servers with many processors, performance improvements were seen when switching the parameter to ES/TABLE = SHM_SEGS. When you switch

the parameter `ES/TABLE` to `SHM_SEGS`, you must also adjust the parameter `as4/MAXSHR64` and set the parameter `ES/SHM_BASE_ADDR` as described in SAP Note 808607 in paragraph 1.1.8.

2.21 POSIX and Olson Time Zones

The current AIX releases support two different time zone formats (POSIX and Olson).

Some examples for these formats are:

POSIX	TZ=CET-1CEST,M3.5.0/02:00:00,M10.5.0/03:00:00
Olson	TZ=Europe/Berlin

We recommend using the POSIX time zone format for SAP systems on AIX, because it provides some performance benefits. If you are using the Olson format and want to switch to POSIX, then please consult the following web site first as it describes some potential issues on certain AIX releases, which might lead to problems with the time not changing properly at the correct Daylight Savings Time start or end dates:

<http://www.ibm.com/support/docview.wss?uid=isg3T1013017>

Please refer also to SAP note 1711777.

2.22 Oracle Database Server Issues

IBM introduced a “1 TB segment aliasing” feature on later technology levels that is enabled by default on AIX 7.1 and can be enabled using the `vmo` command on AIX 6.1. This feature can improve the performance of applications using large amounts of memory. However, it can create problems with Oracle Database 11.2.0.3 causing a performance overhead and high system times. Older Oracle versions do not have the same issue.

Please have a look at the following white paper which provides recommendations to address this issue:

<http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102172>

Another document to check about running Oracle databases on POWER7 systems is:

IBM POWER7 AIX and Oracle Database performance considerations

<http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102171>

2.23 Main Storage Pools on IBM i

In general, one can run multiple SAP systems in one logical partition at the same time. However, if the SAP systems show different activity characteristics, one may observe poor performance on a system that is rarely used. Unless you configure separate main storage pools, all SAP systems on IBM i are running in the `*BASE` pool. As memory pages need to be loaded from disk into main storage, other pages that have not been used for a while are purged from the main storage. If those purged pages are needed again, they must be loaded from disk, which affects performance. Typical examples for poor performance because of the described behavior are the combination of ABAP and Java instances in one logical partition, or the combination of a frequently used development system and a rarely used test system.

To address this problem, one can configure separate main storage pools for the different SAP systems. If one system is very active in memory, it only affects pages in its own main storage pool. Other systems in other main storage pools that are idle will not be affected and can access their data directly from memory when they are used the next time. SAP Note 1023092

describes how to setup separate main storage pools for SAP instances, such as Java instances. SAP Note 49201 has general recommendations about main storage settings on IBM i.

3 Brief Summary of Important Rules

- Adjust LPAR definitions to reflect the actual demand (correct settings for entitlement and virtual processors)
- Business critical and response time sensitive production partitions should have an entitlement that covers the average utilization during the peak hour, add a few more VCPUs to cover additional peaks
- Activate the most important partitions first to allow for better partition placement
- Include VIO server partitions in sizing and capacity planning process and preferably run them in dedicated or dedicated donating processor configurations
- Boot partitions in POWER7 mode and use SMT4
- Beware of capacity versus single thread performance topics
- Do not migrate from fastest POWER6 to slowest POWER7 processor
- Check and implement the recommendations in the "P7 Virtualization Best Practice Guide" available at:

<https://www.ibm.com/developerworks/wikis/display/WikiPtype/Performance+Monitoring+Documentation>

4 Resources

4.1 Whitepapers

POWER7 Virtualization Best Practice Guide:

<https://www.ibm.com/developerworks/wikis/display/WikiPtype/Performance+Monitoring+Documentation>

AIX on Power – Performance FAQ

<http://public.dhe.ibm.com/common/ssi/ecm/en/pow03049usen/POW03049USEN.PDF>

Disabling unshared segment aliasing on AIX for Oracle Database 11.2.0.3

<http://www.ibm.com/support/techdocs/atmastr.nsf/WebIndex/WP102172>

IBM POWER7 AIX and Oracle Database performance considerations

<http://www.ibm.com/support/techdocs/atmastr.nsf/WebIndex/WP102171>

Tuning SAP with DB2 on IBM AIX

<http://www.ibm.com/support/techdocs/atmastr.nsf/WebIndex/WP101601>

Oracle Architecture and Tuning on AIX v2.30

<http://www.ibm.com/support/techdocs/atmastr.nsf/WebIndex/WP100883>

4.2 Web Sites/Blogs

PowerVM Virtualization Performance Advisor

<http://www.ibm.com/developerworks/wikis/display/WikiPtype/PowerVM+Virtualization+performance+advisor>

VIOS Performance Advisor

<http://www.ibm.com/developerworks/wikis/display/WikiPtype/VIOS+Advisor>

JAVA Performance Advisor

<http://www.ibm.com/developerworks/wikis/display/WikiPtype/Java+Performance+Advisor>

Local, Near & Far Memory series on AIXpert blog on developerWorks

<https://www.ibm.com/developerworks/mydeveloperworks/blogs/aixpert/?lang=en>

SAP on IBM i Wiki on IBM developerWorks

<https://www.ibm.com/developerworks/mydeveloperworks/wikis/home?lang=en#/wiki/i%20ERP1/page/SAP%20on%20IBM%20i>

SAP on IBM i Space in the SAP Community Network

<http://scn.sap.com/community/ibm-i>

4.3 Redbooks

SAP Applications on IBM PowerVM

<http://www.redbooks.ibm.com/abstracts/sg247564.html?Open>

IBM PowerVM Virtualization Introduction and Configuration

<http://www.redbooks.ibm.com/abstracts/sg247940.html?Open>

IBM PowerVM Getting Started Guide

<http://www.redbooks.ibm.com/abstracts/redp4815.html?Open>

IBM PowerVM Best Practices

<http://www.redbooks.ibm.com/abstracts/sg248062.html?Open>

IBM PowerVM Virtualization Managing and Monitoring

<http://www.redbooks.ibm.com/abstracts/sg247590.html?Open>

POWER7 Optimization and Tuning Guide

<http://www.redbooks.ibm.com/redpieces/abstracts/sg248079.html?Open>

4.4 Product Documentation

AIX 6.1 Information Center

<http://pic.dhe.ibm.com/infocenter/aix/v6r1/index.jsp>

AIX 7.1 Information Center

<http://pic.dhe.ibm.com/infocenter/aix/v7r1/index.jsp?topic=/com.ibm.aix.doc/doc/base/aixinformation.htm>

[IBM i 6.1 Information Center](#)

<http://publib.boulder.ibm.com/infocenter/series/v6r1m0/index.jsp>

IBM i 7.1 Information Center

<http://publib.boulder.ibm.com/infocenter/series/v7r1m0/index.jsp>

IBM Power Systems Hardware Information Center

<http://pic.dhe.ibm.com/infocenter/powersys/v3r1m5/index.jsp?topic=/ipha8/hwicwelcome.htm>

4.5 SAP Documentation

SAP Help Portal

<https://help.sap.com/>

SAP Community Network (SCN): SAP on AIX

<http://scn.sap.com/community/aix>

4.6 Important SAP Notes

SAP Note	Title / Description
SAP Applications on AIX	
1087498	Support for AIX 6.1
1137862	Using SAP systems with AIX 6.1
1096445	AIX 6.1 known problems for validation
1048686	Recommended AIX settings for SAP
1458918	Support for AIX 7.1
1541935	Using SAP systems with AIX 7.1
973227	AIX Virtual Memory Management: Tuning Recommendations
789477	Large extended memory on AIX (64-bit) as of Kernel
1088458	AIX: Performance improvement for ES/SHM_SEGS_VERSION
856848	AIX Extended Memory Disclaiming
1121904	SAP on AIX: Recommendations for paging space
948294	AIX JFS2: mount options to use with Oracle 10g / 11g
1289460	ORA-354 on AIX5L if agblksize is not set correctly
1728283	SAP Kernel 721: General Information
1711777	Recommended time zone setting on AIX
SAP Applications on IBM i	
49201	IBM i: Main storage settings for SAP
808607	IBM i: Memory management in a system based on PASE
853564	IBM i: Known issues with V5R4Mx
1148480	IBM i: Known issues with V6R1Mx
1432783	IBM i: Known issues with OS release IBM i V710
1687173	IBM i: Collection of notes about the 7.20 kernel
PowerVM Virtualization	
1002461	Support of IBM Dynamic LPAR and Micropartitioning
1102760	PowerVM Live Partition Mobility
1105456	SAP Installations in AIX 6.1 WPARs
1122387	Linux: SAP support in virtualized environments
1464605	POWER7 Active Memory Expansion
Monitoring	
994025	Virtualized OS environments in the operating system monitor
113169	CPU utilization metrics of IBM System p
1379855	Include VIOS Partitions into SAP Performance Monitoring