

2009 System z Expo
October 5 – 9, 2009 – Orlando, FL



z/VSE Performance Update

zEG05

Ingo Franzki, IBM



© 2009 IBM Corporation

Trademarks

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

Not all common law marks used by IBM are listed on this page. Failure of a mark to appear does not mean that IBM does not use the mark nor does it mean that the product is not actively marketed or is not significant within its relevant market.

Those trademarks followed by ® are registered trademarks of IBM in the United States; all others are trademarks or common law marks of IBM in the United States.

For a complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml:

*, AS/400®, e business(logo)®, DBE, ESCO, eServer, FICON, IBM®, IBM (logo)®, iSeries®, MVS, OS/390®, pSeries®, RS/6000®, S/30, VM/ESA®, VSE/ESA, WebSphere®, xSeries®, z/OS®, zSeries®, z/VM®, System i, System i5, System p, System p5, System x, System z, System z9®, BladeCenter®

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.

Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. 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.

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.

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

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

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.

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

* All other products may be trademarks or registered trademarks of their respective 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.

Disclaimer

The information contained in this document has not been submitted to any formal IBM test and is distributed on an "AS IS" basis without any warranty either express or implied. The use of this information or the implementation of any of these techniques is a customer responsibility and depends on the customer's ability to evaluate and integrate them into the operational environment. While each item may have been reviewed by IBM for accuracy in a specific situation, there is no guarantee that the same or similar results will be obtained elsewhere. Customers attempting to adapt these techniques to their own environments do so at their own risk.

In this document, any references made to an IBM licensed program are not intended to state or imply that only IBM's licensed program may be used; any functionally equivalent program may be used instead.

Any performance data contained in this document was determined in a controlled environment and, therefore, the results which may be obtained in other operating environments may vary significantly. Users of this document should verify the applicable data for their specific environments.

It is possible that this material may contain reference to, or information about, IBM products (machines and programs), programming, or services that are not announced in your country. Such references or information must not be construed to mean that IBM intends to announce such IBM products, programming or services in your country.

Agenda

- § Hardware support
- § z/VSE V 4.2 Considerations
 - More tasks
 - PAV Performance Considerations
 - CPU Balancing
- § Sizing a system for z/VSE
- § Miscellaneous Considerations

z/VSE V4.1 and V4.2 Hardware support

§ z/VSE V4.1 and V4.2 runs on the following machines

- IBM System z10 BC or z10 EC
- IBM System z9 BC or z9 EC (formerly z9-109)
- IBM zSeries: z800, z900, z990, z890
- z/VM V5.2 (or later) is a prerequisite for running z/VSE V4 under VM.

§ z/VSE 3.1 and VSE/ESA 2.7 runs on the following machines

- IBM System z10 BC or z10 EC
- IBM System z9 BC or z9 EC (z9-109)
- IBM zSeries: z800, z900, z990, z890
- 9672 Parallel Enterprise Server (G5/G6)
- Multiprice 3000 (7060)
- equivalent emulators (Flex-ES)

Supported VSE Releases

VSE Release	Available	End of Marketing	End of Service
z/VSE 4.2	10/17/2008		
z/VSE 4.1	03/16/2007	10/17/2008	04/30/2010
z/VSE 3.1	03/04/2005	05/31/2008	07/31/2009 (out of service)
VSE/ESA 2.7	03/14/2003	09/30/2005	02/28/2007 (out of service)
VSE/ESA 2.6	12/14/2001	03/14/2003	03/31/2006 (out of service)
VSE/ESA 2.5	09/29/2000	12/14/2001	12/31/2003 (out of service)
VSE/ESA 2.4	06/25/1999	09/29/2000	06/30/2002 (out of service)
VSE/ESA 2.3	07/12/1997	06/30/2000	12/31/2001 (out of service)

Running z/VSE V4 under z/VM

- § z/VM V5.2 (or later) is a prerequisite for running z/VSE V4.1 or V4.2 under z/VM
 - If you IPL z/VSE V4.1 or V4.2 in a guest system of z/VM version 4 or z/VM 5.1, you may experience severe performance problems
 - Because of that the following message is issued during IPL:
 - 0J86I WARNING: VM RELEASE NOT SUPPORTED BY VSE 4.1 - Z/VM 5.2 OR LATER REQUIRED
 - If you receive this message, you must urgently upgrade your VM system to z/VM 5.2 or a later release.
- § Note: It is not required to run z/VSE under z/VM, you can also run z/VSE in an LPAR

Overhead Deltas for VSE Releases

VSE/ESA 2.3
Standard
Dispatcher
CICS/VSE

Remember that you get a lot of new functions that in most cases helps you to increase VSE system performance and throughput:

Partition Balancing, PRTY SHARE (Turbo Dispatcher), FlashCopy, Buffer Hashing, Shared data Tables (CICS TS), NOPDS with larger VSIZE

+5-10%

VSE/ESA 2.3
**Turbo
Dispatcher**
CICS/VSE

8%

VSE/ESA 2.7
Turbo
Dispatcher
CICS/VSE

3%

z/VSE 3.1
Turbo
Dispatcher
CICS/VSE

3%

z/VSE 4.1
Turbo
Dispatcher
CICS/VSE

5%

z/VSE 4.2
Turbo
Dispatcher
CICS/VSE

+5%

+5%

+5%

+5%

These numbers are for
**a specific average I/O
intensive workload**
(PACEX16)

More I/O intensive
workloads can cause
more overhead

VSE/ESA 2.7
Turbo
Dispatcher
CICS TS

3%

z/VSE 3.1
Turbo
Dispatcher
CICS TS

3%

z/VSE 4.1
Turbo
Dispatcher
CICS TS

5%

z/VSE 4.2
Turbo
Dispatcher
CICS TS

z/VSE V4.2 Performance Considerations

§ More tasks

- z/VSE V4.2 now supports up to 512 tasks

§ PAV Performance Considerations

- PAV = Parallel Access Volumes

§ CPU Balancing

- Automatically start and stops CPUs according to the current workload

z/VSE V4.2 – More tasks

§ z/VSE 4.2 supports up to 512 concurrent tasks

- the previous limit was 255 concurrent tasks).
- Additional task IDs will be in the range X'0100' to X'0200'.
- A QUERY SYSTEM command has been provided to display information on the current task usage

§ Please note:

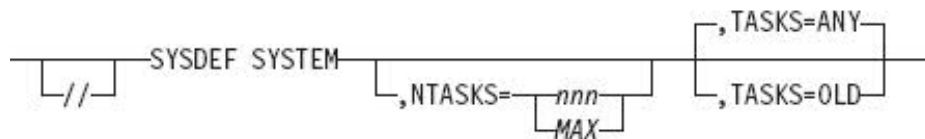
- The maximum **number of tasks per partition remains 32**.
- The default maximum remains 255. If you wish to increase the maximum number of tasks, you must use the SYSDEF command to increase it to any number up to 512.

§ Benefits:

- The workload resulting from CICS and batch programs can now be more efficiently distributed
- Previously, some customers might have been reluctant to perform migration from CICS/VSE to the CICS Transaction Server for VSE/ESA because of its more intensive usage of VSE tasks (compared to CICS/VSE). This inhibitor has now been removed.
- Using more tasks have **no measurable performance impact** (neither negative nor positive)

z/VSE V4.2 – More tasks

- § The SYSDEF SYSTEM command is used to activate the new tasks support system-wide.



§ NTASKS=nnn|MAX

- nnn specifies the total number of tasks (old and new) that can be allocated to the system and application programs. This is possible only during BG ASI processing. The value of nnn can be between 255 and 512.

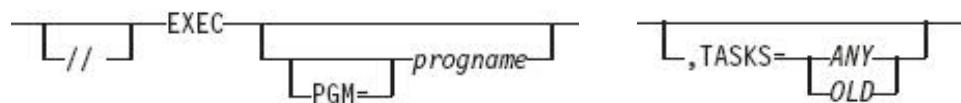
§ TASKS=ANY|OLD

- This is the system-wide default mode for subtask attaching, it can be run at any time after IPL:
 - If TASKS=ANY is specified, new or old tasks can be attached as subtasks. However, primarily new tasks will be attached.
 - OLD specifies that only old tasks will be attached as subtasks.

§ Use QUERY SYSTEM to display what has been specified with SYSDEF SYSTEM

z/VSE V4.2 – More tasks

§ The TASKS specification in the SYSDEF SYSTEM command **can be overwritten** with the TASKS parameter in the JCL EXEC statement:



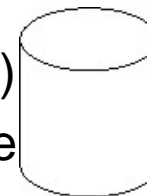
- When the SYSDEF SYSTEM command is used with TASKS=OLD, application programs which want to make use of the new tasks, must be started with parameter TASKS=ANY in the EXEC statement.
- When the SYSDEF SYSTEM command is used with TASKS=ANY, but an application program needs to be restricted to old tasks, the parameter TASKS=OLD has to be specified in the EXEC statement.

z/VSE V4.2 – Parallel Access Volumes (PAV)

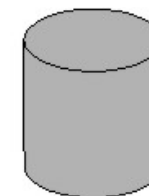
§ Parallel Access Volume (PAV) is an **optional licensed feature** on the IBM System Storage DS6000/DS8000 series

§ Parallel access volumes:

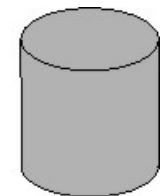
- Are managed by creating **multiple addresses (aliases)** for a single **logical device**.
- Allow your z/VSE system to **access volumes in parallel**.
- Enable **more than one I/O operation** to be processed for a single logical device.
- Depending upon workload, might represent a significant performance improvement by the storage unit over traditional I/O processing.
- Can significantly **reduce device queue delays** to busy devices (all I/O operations are normally processed from a single queue)
- Allow your system to access a single volume from a single z/VSE host with **multiple concurrent requests**.



Base



Alias X



Alias Y

z/VSE V4.2 – Parallel Access Volumes (PAV)

- § To gain a performance benefit from PAV, it is required that:
 - the workload can **utilize multiple paths** (sub channels) to the same volume
 - A single job that **starts an I/O and waits** until its completion before it starts another I/O **does not benefit** from PAV by itself
 - However, **running multiple such jobs concurrently** in multiple partitions can utilize parallel access to the same volume.

- § Besides the simple ‘Start I/O and wait’ behavior, more sophisticated programs (like CICS) may be capable to **start multiple I/Os and wait for its completion asynchronously**
 - This behavior also is eligible for benefiting from PAV, if the I/Os are for the same volume

- § Jobs that use multiple subtasks that perform I/O operations may benefit from PAV

- § Running **multiple jobs concurrently** in separate partitions and access the same volume are also good candidates for PAV

z/VSE V4.2 – Parallel Access Volumes (PAV)

§ Candidates for PAV volumes

- In general, all volumes that contain files, catalogs of data in general that are **used from multiple jobs** can benefit from PAV
 - Without PAV, I/Os might get queued and therefore delayed in case of parallel access to the same volume
 - PAV allows parallel access to these files from different jobs without disturbing each other.
- VSAM Catalogs
- Shared VSAM Clusters
- VSE Libraries

§ Besides volumes that contain files that are shared between multiple jobs, also volumes that contain **heavily used system files** can benefit from PAV

- Although there might not be any parallel access to those system files as such, there may be **parallel access to other files residing on the same volume**
 - Without PAV, I/Os might get queued and therefore delayed in case of parallel access to the same volume
 - PAV allows parallel access to these volumes without being disturbed by access to the system files.
- POWER Spool files
- Workfiles
- Log files

z/VSE V4.2 – Parallel Access Volumes (PAV)

§ Besides the general PAV setup, **no changes** in application programs or JCL are required

- PAV is enabled in z/VSE via the SYSDEF SYSTEM command:

```
SYSDDEF SYSTEM // ,PAV= START STOP
```

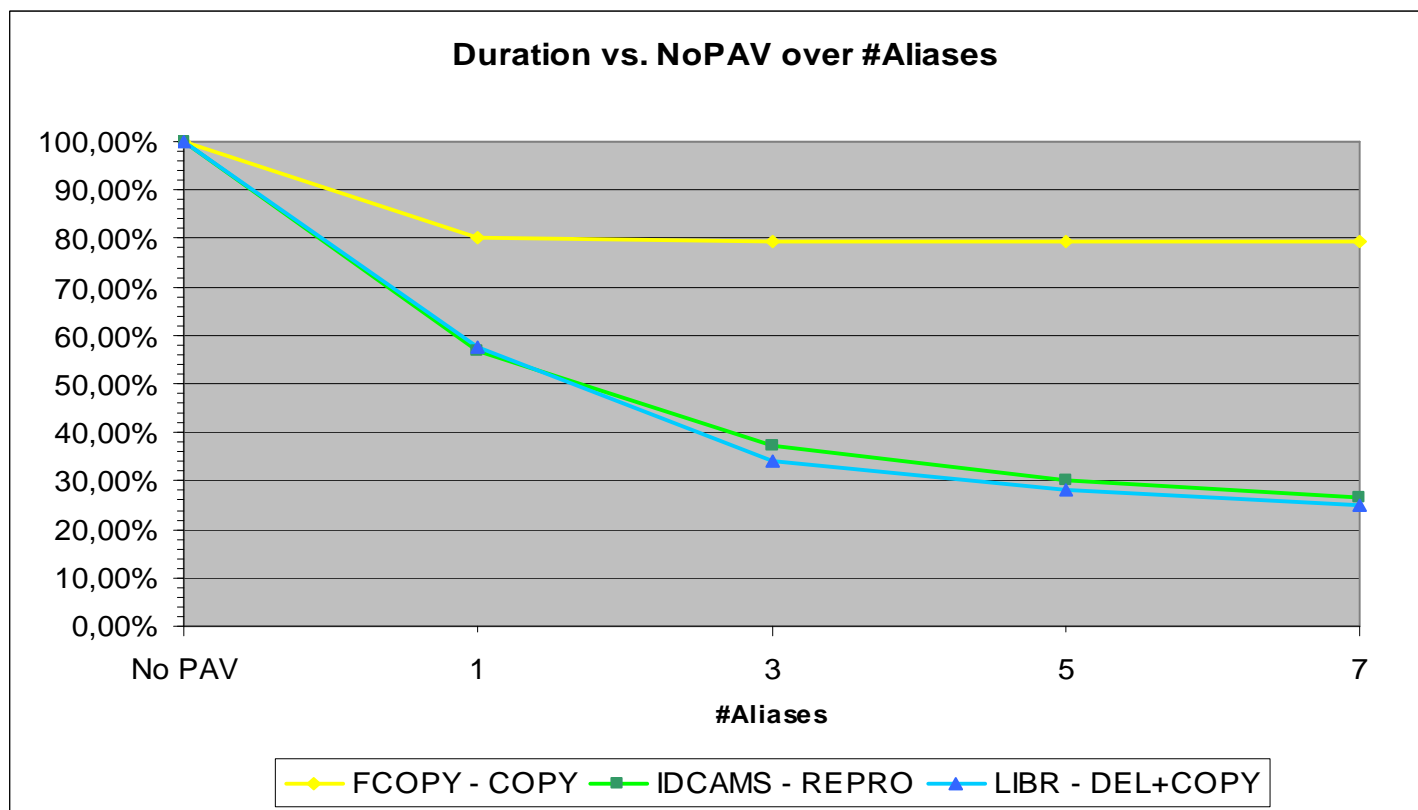
- Define the PAV base devices to z/VSE via ADD statement
- You must NOT define your PAV-alias volumes to z/VSE!
 - An ADD statement for an alias device will be ignored

§ z/VM support for PAV (since z/VM 5.2)

- Dedicated PAV devices
- PAV Mindisks.
- For more information see here:
<http://www.vm.ibm.com/storman/pav/pav2.html>

z/VSE V4.2 – Parallel Access Volumes (PAV)

PAV performance measurements:



Conclusion: Dependent on the workload, PAV can dramatically increase the I/O throughput and reduce the job duration (e.g. 1/3 of the duration when running without PAV), if it is not limited by other factors (like LOCKs).

z/VSE V4.2 – CPU Balancing

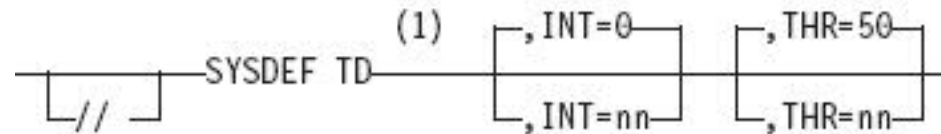
- § Using **CPU Balancing can reduce the multiprocessing overhead**
 - When CPU Balancing is active, the Turbo Dispatcher only selects CPUs that are required for the current workload
 - CPU Balancing might also reduce the overhead or performance-degradation for I/O-intensive workloads that are running in one partition only (the remaining partitions are idle) when multiple CPUs are active

- § With CPU Balancing active, Turbo Dispatcher inspects the CPU utilization periodically
 - When the the utilization falls below a certain threshold, CPUs are dynamically stopped
 - When the the utilization increases over the threshold, CPUs are dynamically started

- § This results in only as many CPUs being active that are utilized by the current workload

z/VSE V4.2 – CPU Balancing

§ CPU Balancing is activated using the SYSDEF TD command



§ INT=nn - nn defines the interval in seconds, after which the CPUs utilization is to be inspected.

- When 0 is specified, CPU balancing will be deactivated
- The default value is 0.
- The value must be in the range of: nn=0..99

§ THR=nn - nn defines the threshold value in percent

- An additional CPU is activated, when the CPU utilization is larger than nn
- The default value is 50
- The value must be in the range of: nn=10..99

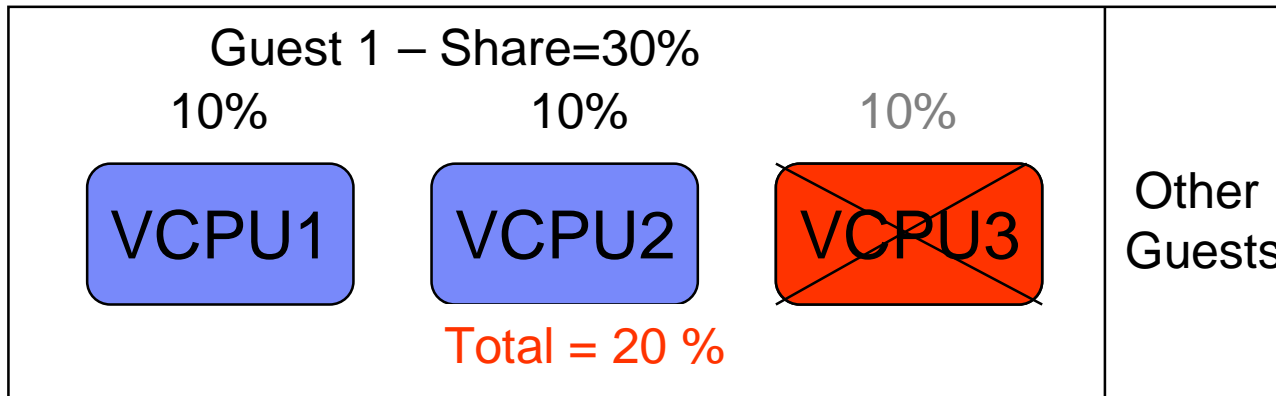
§ You can use the QUERY TD command to obtain information about the current status of CPU balancing

z/VM 5.4 Considerations

- § z/VM V5.4 removes one of the consequences of stopped virtual processors for virtual processor prioritization
- § Prior to z/VM V5.4, a guest's CPU share was distributed equally among its virtual processors
 - by dividing its share value by the number of non-dedicated virtual processors in the configuration
 - regardless of whether the virtual processors were in a stopped or started state
- § z/VM V5.4 performs share redistribution whenever a virtual processor is started or stopped
 - and no longer includes stopped virtual processors in the calculation of how much share to distribute to each virtual processor
- § The share redistribution also continues to occur in z/VM V5.4, as in previous releases
 - at the time a SET SHARE command is issued
 - or when the number of defined virtual processors is altered.

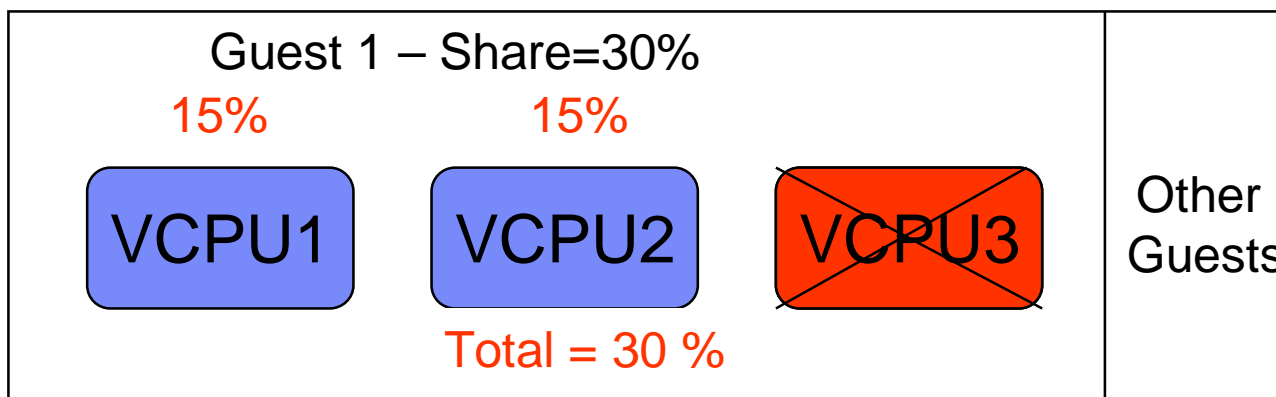
z/VM 5.4 Considerations

<= z/VM 5.3



A guest's CPU share is distributed equally among its virtual processors by dividing its share value by the number of processors, **regardless of whether the virtual processors were in a stopped or started state.**

z/VM 5.4



z/VM V5.4 performs share redistribution whenever a virtual processor is started or stopped and **no longer includes stopped virtual processors** in the calculation of how much share to distribute to each virtual processor.

Sizing a system for z/VSE

- § Sizing a system for z/VSE is different from sizing a system for z/OS
 - Although z/VSE supports multiprocessing, z/VSE does not scale as good as z/OS does
 - Do not use more than 3 active processors per z/VSE LPAR or z/VM Guest

- § In general, a faster single CPU is better than multiple smaller CPUs
 - One partition can only exploit the power of one CPU
 - The largest partition (e.g. CICS) must fit into one single CPU
 - Dependent on nonparallel share (NPS) value

- § Additional CPUs can be useful when multiple LPARs or z/VM Guests are used
 - Define only up to 3 CPUs per LPAR or z/VM Guest, even if more than 3 CPUs are available on the CEC

- § Do not use MIPS tables for capacity planning purposes
 - Use zPCR Tool (see page 24) instead with the CB-L workload

Sizing a system for z/VSE

**The fastest
uni-processor
is (almost always *)
the best processor**

(*) from a single VSE-image point o view

IBM Processor Capacity Reference for zSeries (zPCR)

- § The zPCR tool was released for customer use on October 25, 2005
 - <http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS1381>
 - 'As is', no official support, e-mail to zpcr@us.ibm.com

- § PC-based productivity tool under Windows

- § It is designed to provide capacity planning insight for IBM System z9/z10 and eServer zSeries processors running various workload environments

- § Capacity results are based on IBM's LSPR data supporting all IBM System z9/z10 and eServer zSeries processors
 - Large System Performance Reference:
<http://www.ibm.com/servers/eserver/zseries/lspr/>

- § For VSE use z/VSE workloads Batch, Online or Mixed

VSE CPU Monitor Tool

- § Intended to help customers to **measure the CPU utilization** of their VSE system **over a period of time**.
- § When you plan for a processor upgrade it is very important to know the CPU utilization of your VSE system over a day or a week.
 - Helps you to estimate the size of the new processor.
- § The VSE CPU Monitor Tool is not intended to replace any existing monitoring product provided by partners.
- § It provides only very **basic monitoring** capabilities on **an overall VSE system level**.
- § No details about CPU usage of certain applications are provided
- § **Download**
 - <http://www.ibm.com/servers/eserver/zseries/zvse/downloads/tools.html>
 - 'As is', no official support, e-mail to zvse@de.ibm.com

VSE CPU Monitor Tool

- § CPUMON **periodically** issues a TDSERV FUNC=TDINFO macro to get performance relevant data.
- § The data provided by the macro is the same as command **QUERY TD** shows.
- § The data from each measurement interval is printed to SYSLST in a comma separated format.
- § Later on this data can be imported into a spreadsheet (EXCEL)
- § CPUMON runs in a VSE partition (dynamic or static).
- § CPUMON is started using:

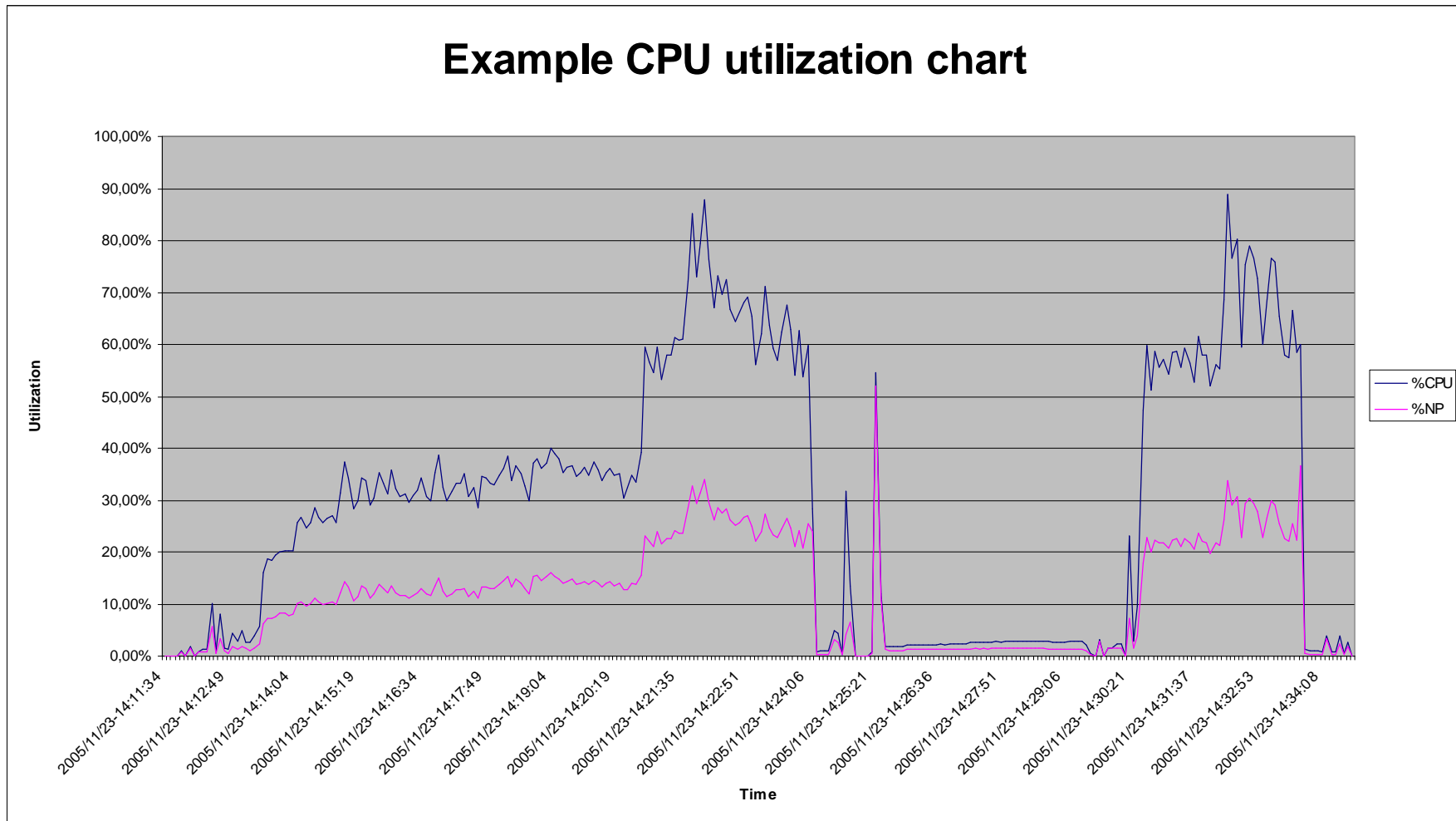
```
// EXEC DTRIATTN, PARM='SYSDEF TD, RESETCNT`  
/*  
// EXEC CPUMON, PARM='nn'   nn = interval in seconds  
/*
```

- § The tool can be stopped by entering the following command:

```
MSG xx, DATA=EXIT           xx = partition id
```

VSE CPU Monitor Tool

Example CPU utilization chart



Performance Tips

- § A partition can only exploit **1 CPU** at a time
 - 2 CPUs do not have any benefit for a single CICS partition
 - Use as many partitions as required for selected n-way

- § Use/define only as many CPUs as really needed
 - additional CPUs create more overhead, but no benefit

- § Partitions setup
 - Set up more batch and/or (independent) CICS partitions
 - Split CICS production partitions into multiple partitions (AOR, TOR, FOR)

- § Try to exploit Turbo Dispatcher functions
 - Priority settings
 - Partition balancing
 - Partition balancing groups

Performance Tips (2)

- § 1 CPU must be able to handle all non-parallel workload
- § Non-parallel code limits the n-Way exploitation
 - QUERY TD: $NP/TOT = NPS$ (non parallel share)
 - Measure NPS before migration
 - max CPUs = $0.8 - 0.9 / NPS$

NPS	#CPUs	NPS	#CPUs
0.20	4.0-4.5 (4)	0.45	1.8-2.0 (2)
0.25	3.2-3.6 (3)	0.50	1.6-1.8 (2)
0.30	2.7-3.0 (3)	0.55	1.5-1.6 (2)
0.35	2.3-2.6 (2)	0.60	1.3-1.5 (1)
0.40	2.0-2.2 (2)	0.65	1.2-1.4 (1)

Performance Tips (3)

- § Non-parallel code limits the maximum MP exploitation
- § System code (Key 0) increases non-parallel share
 - Vendor code can have significant impact
- § Overhead increases when NP code limits throughput

- § Data In Memory (DIM) reduces non-parallel code
 - less system calls (I/Os)
 - may increase throughput
 - CICS Shared Data Tables
 - Large/many VSAM Buffers (with buffer hashing)
 - Virtual Disks

- § Change VSE/POWER startup to WORKUNIT=PA
- § Switch tracing/DEBUG off for production

Documentation

§ z/VSE homepage:

- <http://www.ibm.com/servers/eserver/zseries/zvse/>

§ VSE Performance:

- <http://www.ibm.com/servers/eserver/zseries/zvse/documentation/performance.html>

§ z/VM homepage:

- <http://www.ibm.com/vm>

§ z/VM Performance:

- <http://www.vm.ibm.com/perf/>

§ z/VM Preferred Guest Migration Considerations

- <http://www.vm.ibm.com/perf/tips/z890.html>

§ IBM System z Software Pricing

- <http://www.ibm.com/servers/eserver/zseries/swprice/>

§ IBM's MSU ratings for IBM System z

- <http://www.ibm.com/servers/eserver/zseries/library/swpriceinfo/hardware.html>

Questions ?

