International Technical Support Organization

z9 PU Resources

www.ibm.com/redbooks | System z Hw Update

Redbooks Workshop
IBM ITSO - International Technical Support Organization

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.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurement may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

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.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

Note to U.S. Government Users Restricted Rights -- Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Trademarks

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

IBM has two registered trademarks for the branding of ITSO publications. These registered marks are for the text word "IBM Redbooks" and the Redbooks logo. In a nutshell, the term Redbooks must always be used in the plural form (for both text and logo) since IBM only owns the registered mark for the plural form. Usage must follow the guidelines below:

Using the term Redbooks in written text

Redbooks are only to be referred to in the plural form, NEVER in the singular.
For the initial reference (first occurrence), you must use "IBM Redbooks®" and include "IBM" as well as the ®. For instances thereafter you may use "Redbooks" without "IBM" preceding the word or ® following it.

Correct usage for written text :

In this IBM Redbooks® publication we will explore…..(® symbol required for 1st usage)
This Redbooks publication will show you…..(2nd usage or later - no ® or "IBM" needed)

Using the logo:

Redbooks (logo)

OTHER ITSO PUBLICATIONS - Marks not yet registered

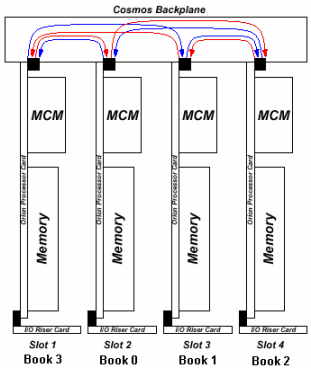
Trademark registration is a lengthy process and until we are officially registered, we cannot use the ® symbol. For those terms/logos in process, we will be using the ™ symbol. In contrast to the ® symbol (placed in the lower right hand corner), the ™ symbol is placed in the upper right hand corner. Please see examples below:

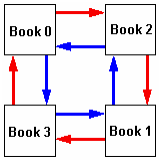
Redpaper ™
Redpapers ™
Redwiki ™
Redwikis ™

TM

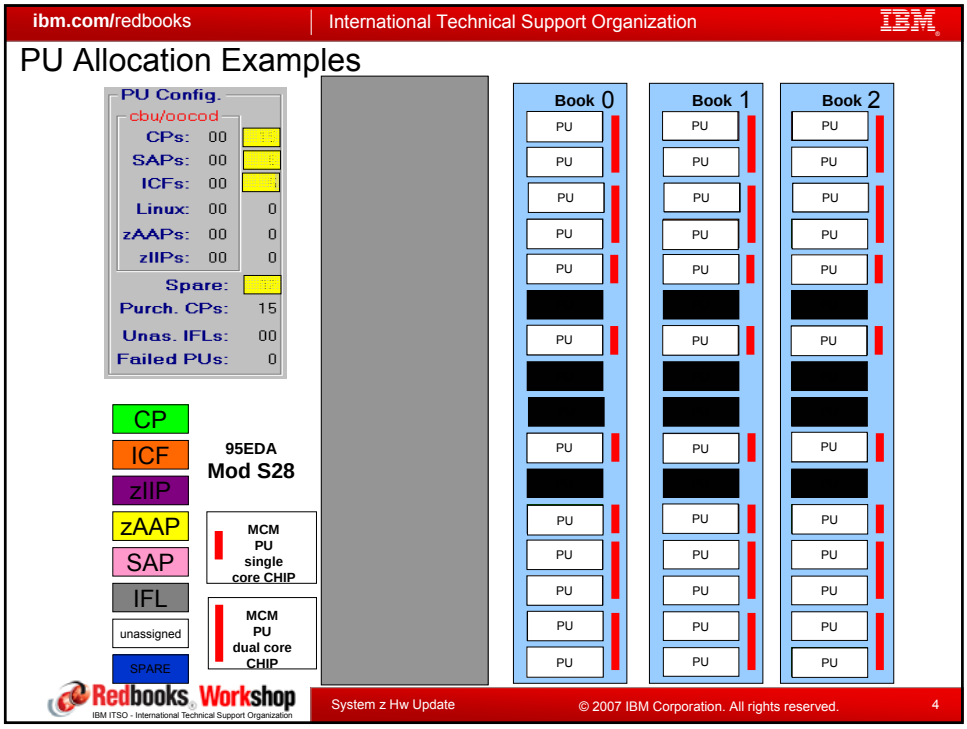
The following terms are trademarks of other companies:
Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.
Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, 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 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.

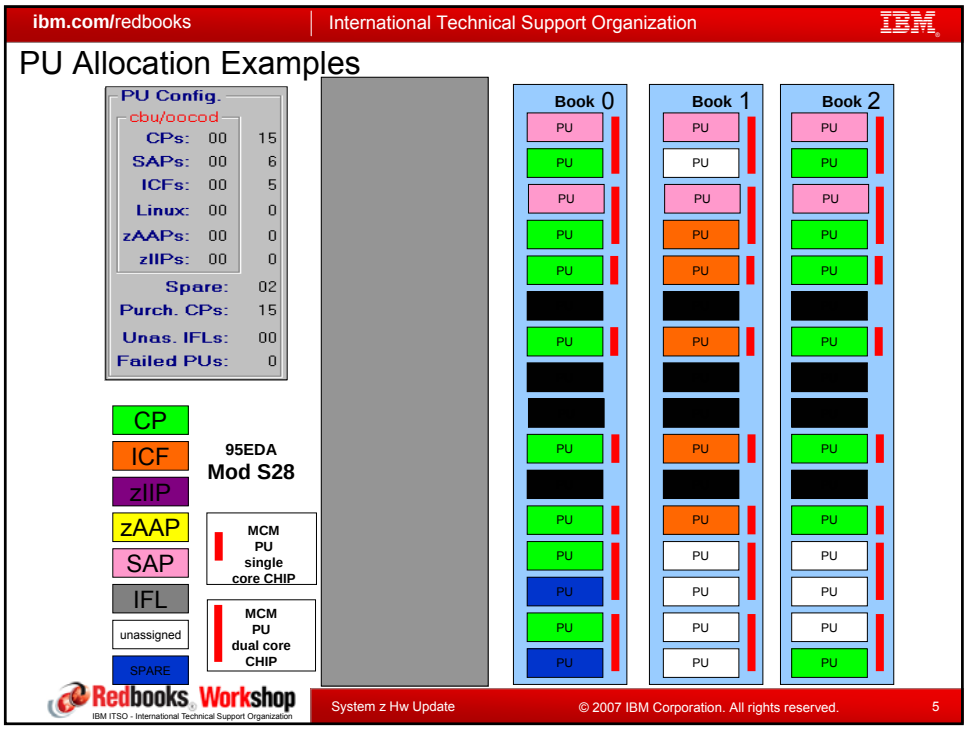
z9 Resources Allocation

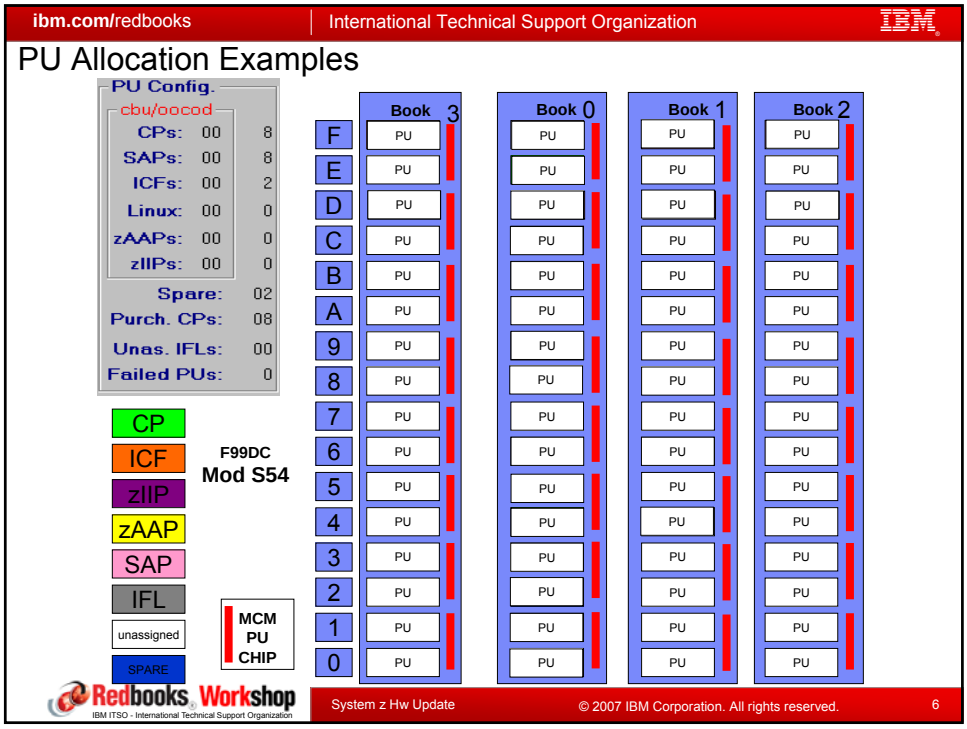


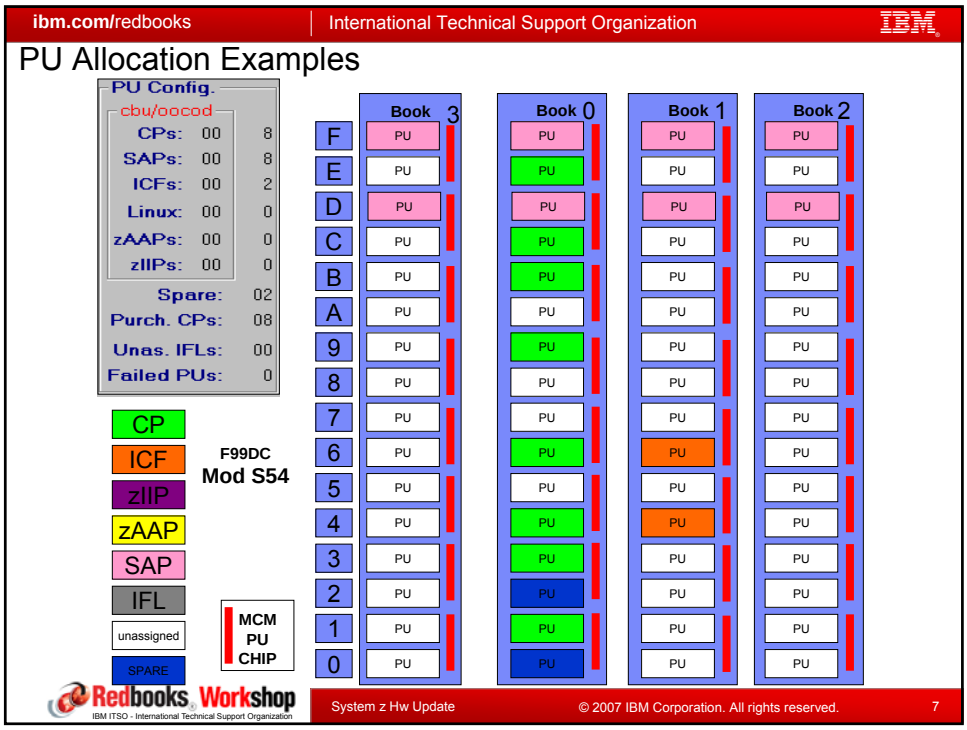


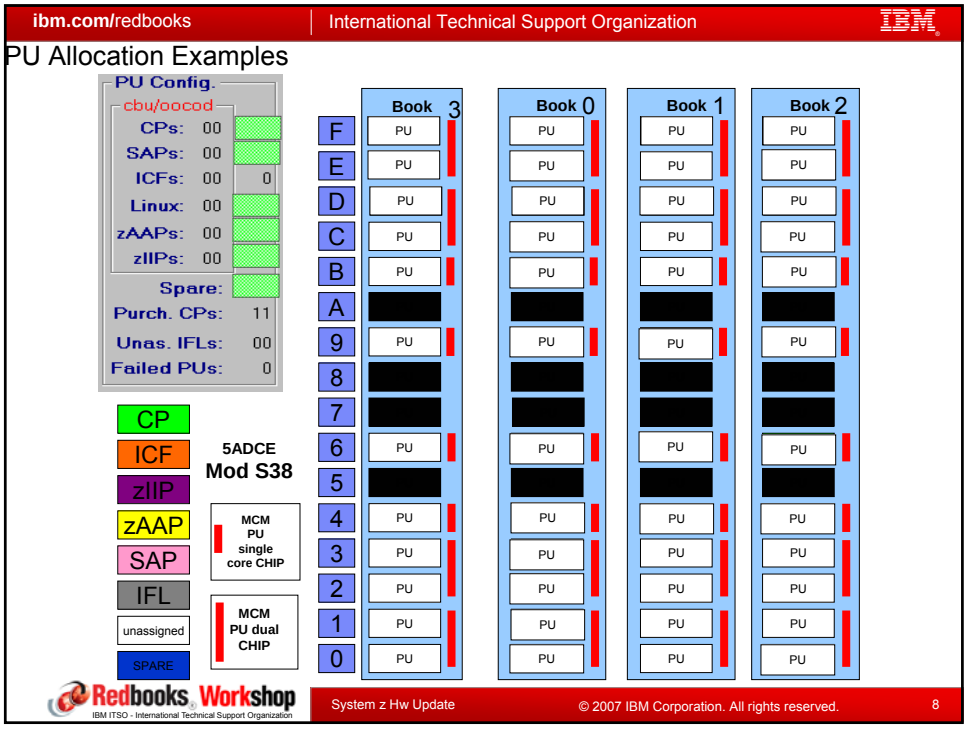
- Physical PU assignments → Performed during IML (POR)
  - balance the SAPs across the existing books
  - cluster the CPs on a single book or adjacent books
  - cluster the ICFs/IFLs together on the opposite books
  - and cluster zIIPs and zAAPs.
- SAPs → two SAPs/book assigned
  - additional SAPs are spread across the available books, one at a time, in book order (Book0, Book1, Book2, Book3)
- CPs → place maximum CPs on Book 0
  - when that is not possible then to an adjacent book on the ring. So the distribution will first fill Book0, then Book2, then Book3 and finally Book1
- ICFs → place maximum ICFs beginning with the Book1
  - then Book3, then Book2 and finally Book0
- zAAPs and zIIPs → zAAPs are assigned next in the same order as **CPs**
- IFLs → IFLs are assigned last in the same order as **ICFs**
- Intra-Book PU assignment algorithm
  - This is done to keep CPUs and zAAPs clustered together away from ICFs and IFLs
  - The algorithm will be very sensitive to the fact that the two PUs in a chip will checkstop together
  - Assigning the two PUs of a chip will be avoided until that is no longer possible
  - As the assignment starts with the SAPs, then the SAPs will be assigned to PU 'F' and PU 'D' (dual core chip
  - The CPs will be assigned one by one starting with PU '4' and only one CP assigned per chip
  - If there are more than 4 CPs, the CPs, will be assigned to share chip with the SAPs first. (performance reason)

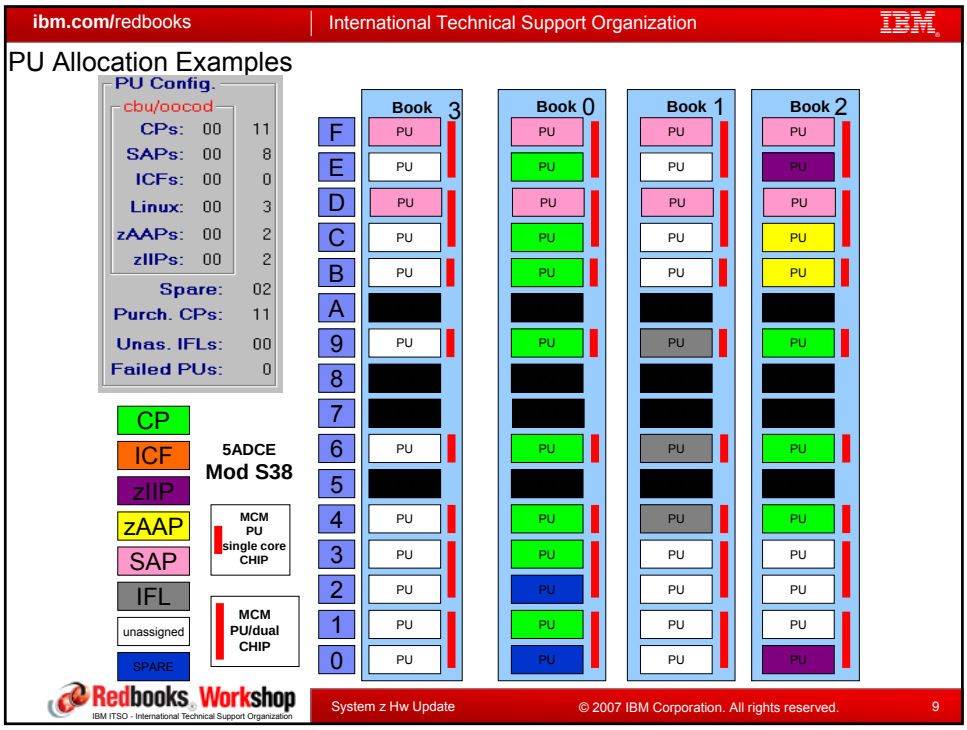


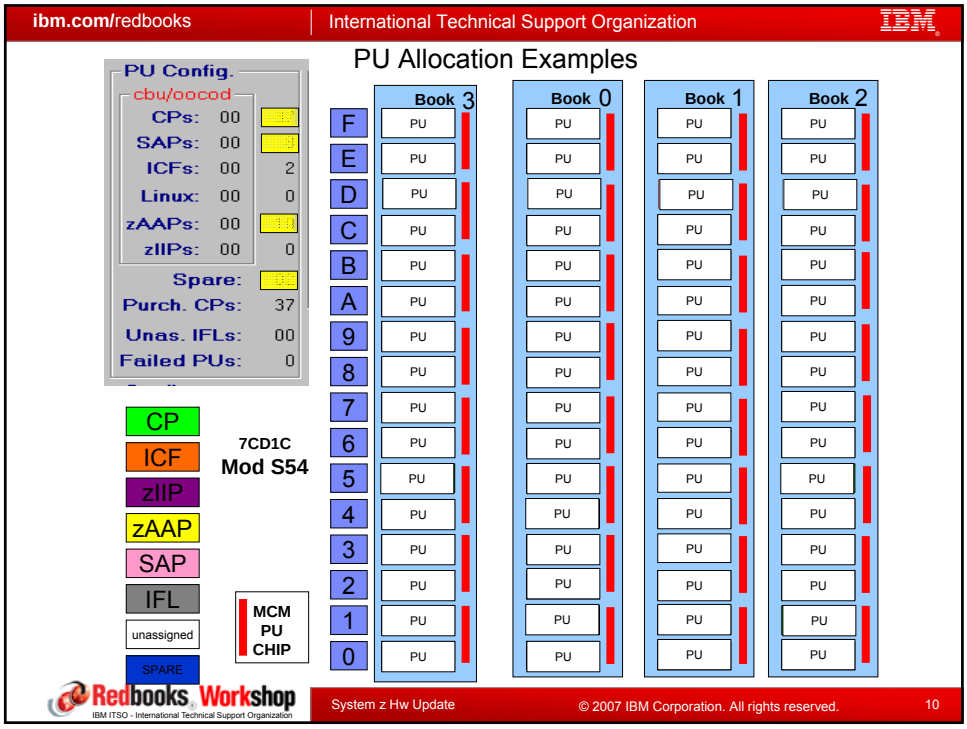


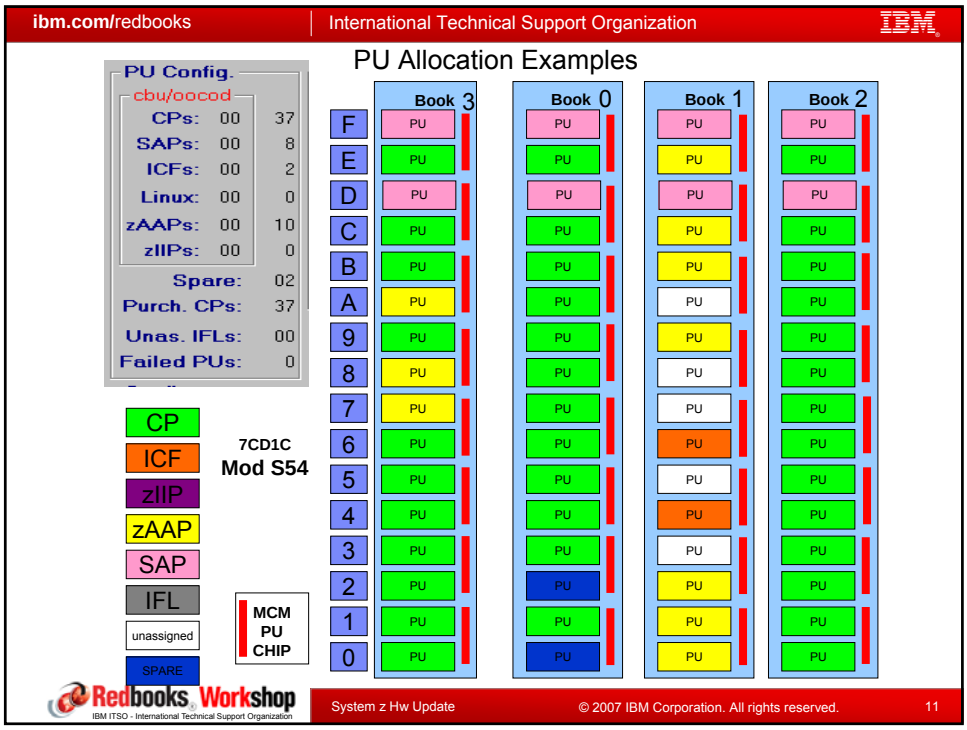




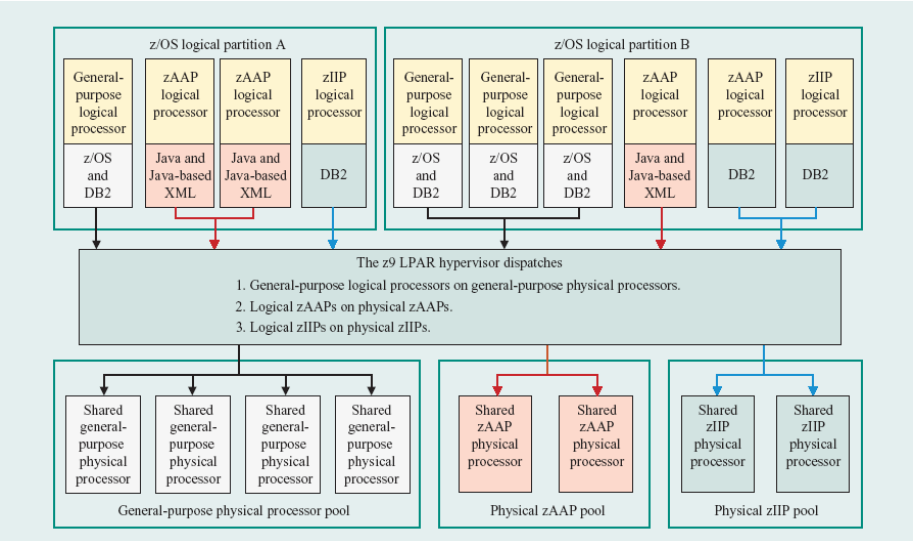






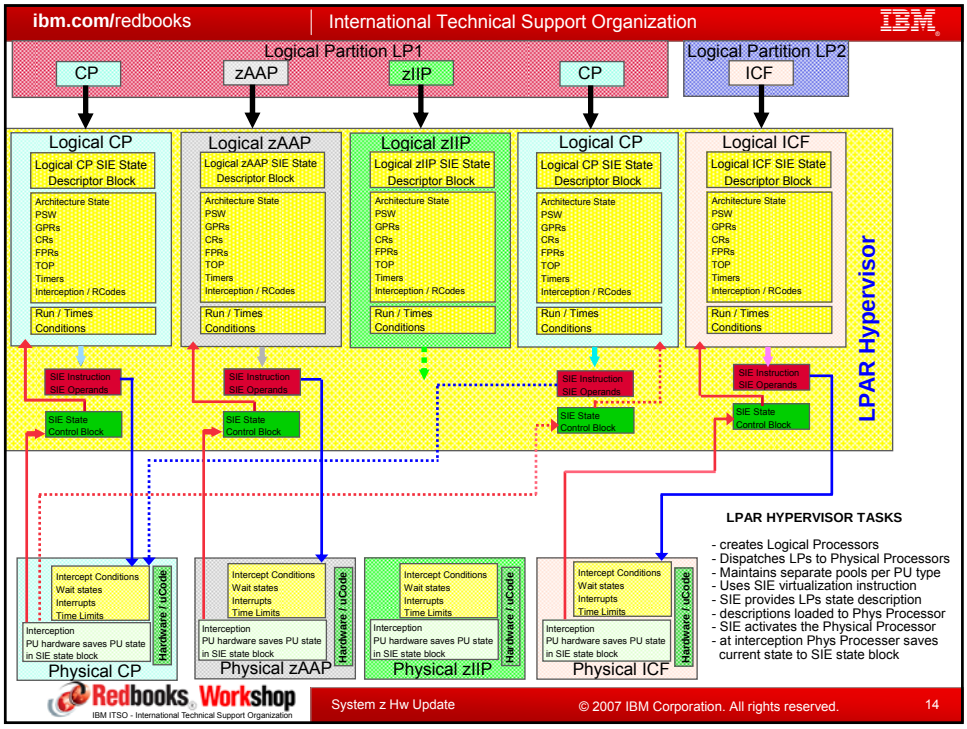


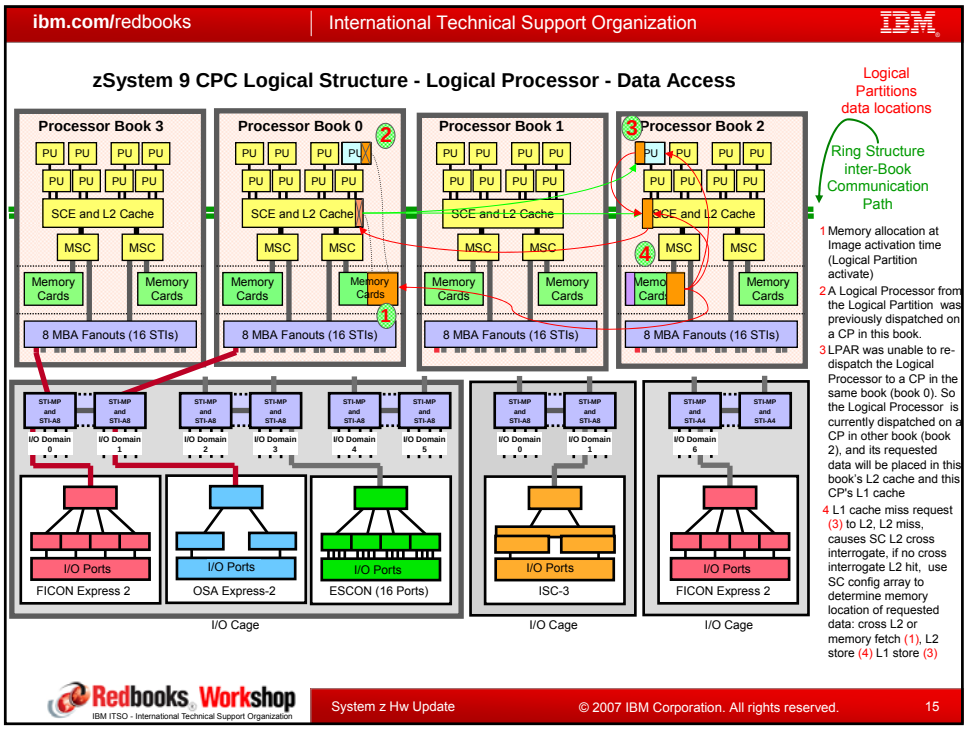
Logical / Physical Processors and LPAR hypervisor



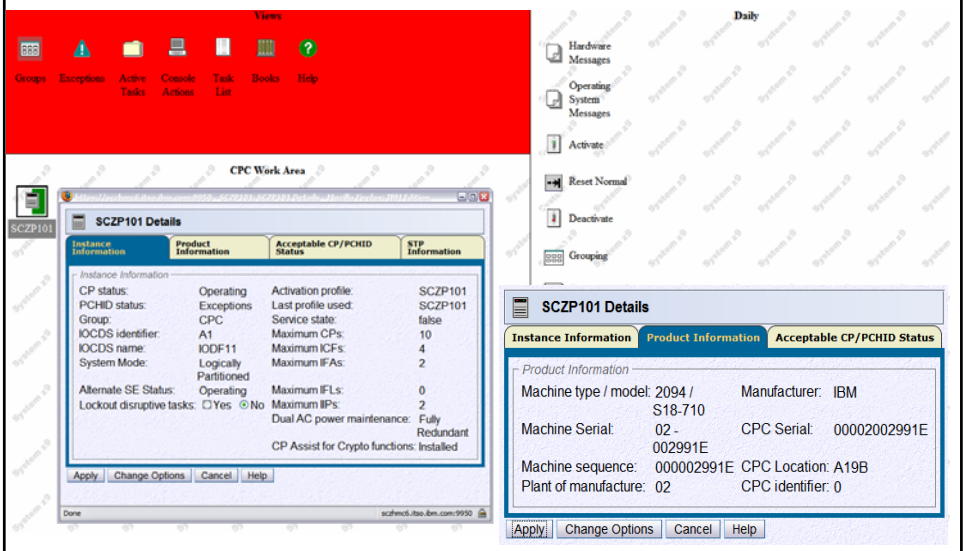
z9 LPAR Hypervisor & Logical Processors

- LPAR Hypervisor Characteristics
  - Creates logical processors
    - A Logical Processor comprises
      - Hardware, millicode, LPAR firmware controls
        - These resources collectively represent the Physical Processor
      - Contains a complete set of Physical Processor controls and operating states
        - Necessary to execute the LP on the hypervisor selected Physical Processor
      - Each partition a separate set of LPs configured to it
  - Dispatches the logical processors to physical processors
  - Creates/maintains separate physical processor pools for each processor type
    - CPs, zIIPs, zAAPs, IFLs and ICFs
  - Uses the proprietary zArchitecture SIE → Start Interpretive Execution
    - SIE is a virtualization technology to
      - Define, initiate and control the execution of the LPs on their corresponding Physical Processors
    - SIE provides LP state description
      - loaded into the Physical Processor when SIE is executed by the hypervisor
    - SIE activates the Physical Processor on behalf of its associated LPAR
  - When activated, dispatched Log Processors remain active for the corresponding LPAR until:
    - Physical Processor is placed in a wait state
      - For example, no programs to be executed
    - A preemptive interrupt condition occurs
      - Time-slice end interruption
      - I/O interruption associated with a different LPAR is recognized by the hypervisor
        - (In both cases, hypervisor typically dispatches the Physical Processor to another sharing LPAR)
    - z/OS executes an instruction that hypervisor must take a special action to assist in the execution
      - Rare cases – very low frequency privileged instructions which SIE controls do not execute
    - Physical Processor encounters an error condition
      - Hypervisor assistance is required to do error recovery
    - Operator action which requires hypervisor assistance
      - LP is varied offline / operator deactivates LPAR

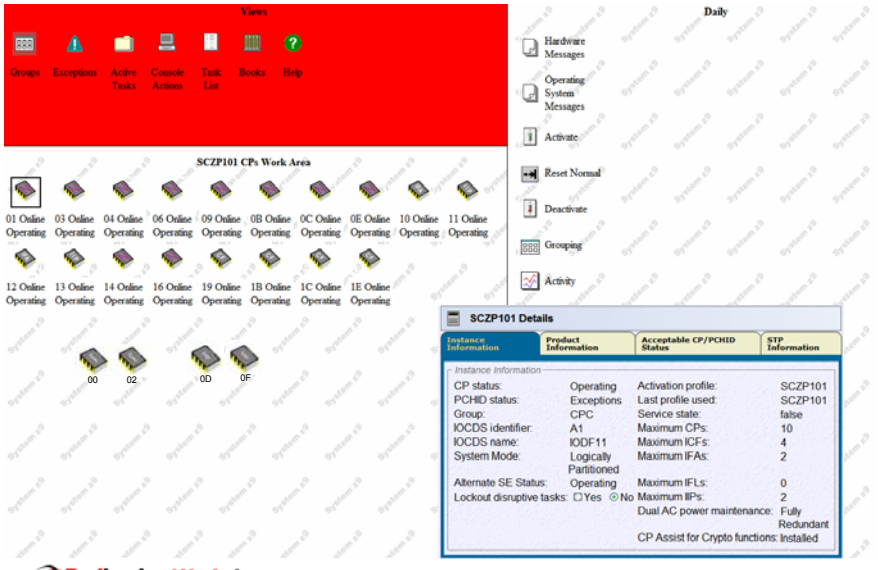

ibm.com/redbooks
International Technical Support Organization
Logical Partition LP1
CP
zAAP
zIIP
CP
Logical Partition LP2
ICF
Logical CP
Logical CP SIE State Descriptor Block
Logical zAAP
Logical zAAP SIE State Descriptor Block
Logical zIIP
Logical zIIP SIE State Descriptor Block
Logical CP
Logical CP SIE State Descriptor Block
Logical ICF
Logical ICF SIE State Descriptor Block
Architecture State
PSW
GPRs
CRs
FPRs
TOP
Timers
Interception / RCodes
Run / Times Conditions
LPAR Hypervisor
SIE Instruction SIE Operands
SIE State Control Block
Intercept Conditions
Wait states
Interrupts
Time Limits
Interception
PU hardware saves PU state in SIE state block
Hardware / uCode
Physical CP
Physical zAAP
Physical zIIP
Physical ICF
LPAR HYPERVISOR TASKS
- creates Logical Processors
- Dispatches LPs to Physical Processors
- Maintains separate pools per PU type
- Uses SIE virtualization instruction
- SIE provides LPs state description
- descriptions loaded to Phys Processor
- SIE activates the Physical Processor
- at interception Phys Processer saves current state to SIE state block
Redbooks Workshop
IBM ITSO - International Technical Support Organization
System z Hw Update
© 2007 IBM Corporation. All rights reserved.
14


ibm.com/redbooks
International Technical Support Organization
zSystem 9 CPC Logical Structure - Logical Processor - Data Access
Processor Book 3
Processor Book 0
Processor Book 1
Processor Book 2
PU
SCE and L2 Cache
MSC
Memory Cards
8 MBA Fanouts (16 STIs)
Logical Partitions data locations
Ring Structure inter-Book Communication Path
1 Memory allocation at Image activation time (Logical Partition activate)
2 A Logical Processor from the Logical Partition was previously dispatched on a CP in this book.
3 LPAR was unable to re-dispatch the Logical Processor to a CP in the same book (book 0). So the Logical Processor is currently dispatched on a CP in other book (book 2), and its requested data will be placed in this book's L2 cache and this CP's L1 cache
4 L1 cache miss request (3) to L2, L2 miss, causes SC L2 cross interrogate, if no cross interrogate L2 hit, use SC config array to determine memory location of requested data: cross L2 or memory fetch (1), L2 store (4) L1 store (3)
STI-MP and STI-A8
STI-MP and STI-A4
I/O Domain
I/O Ports
FICON Express 2
OSA Express-2
ESCON (16 Ports)
ISC-3
I/O Cage
Redbooks Workshop
IBM ITSO - International Technical Support Organization
System z Hw Update
© 2007 IBM Corporation. All rights reserved.
15

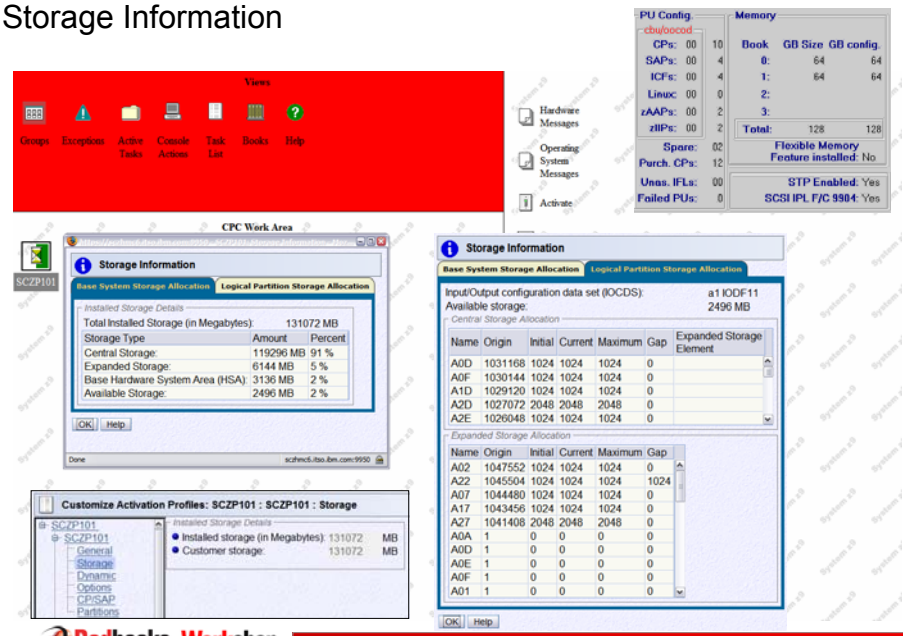
PU Allocation –SE



PU Allocation – SE (cont)



Storage Information



Memory size representation in the SE panels

2^{10}	1,024	1,024 MB	1 GB
2^{11}	2,048	2,048 MB	2 GB
2^{12}	4,096	4,096 MB	4 GB
2^{13}	8,192	8,192 MB	8 GB
2^{14}	16,384	16,384 MB	16 GB
2^{15}	32,768	32,768 MB	32 GB
2^{16}	65,536	65,536 MB	64 GB
2^{17}	131,072	131,072 MB*	128 GB
2^{18}	262,144	262,144 MB	256 GB
2^{19}	524,288	524,288 MB	512 GB
2^{20}	1,048,576	1,048,576 MB	1 TB

* Note: 128 GB is really 2^{37} =137,438,953,472 Bytes and not 131,072,000,000 as presented in the SE

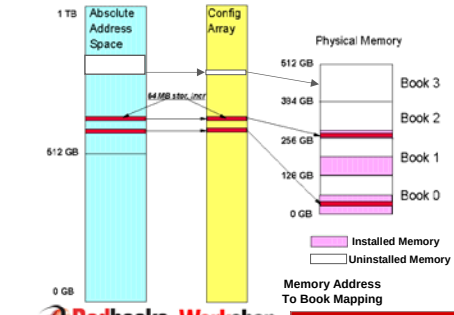
Quantities of bytes

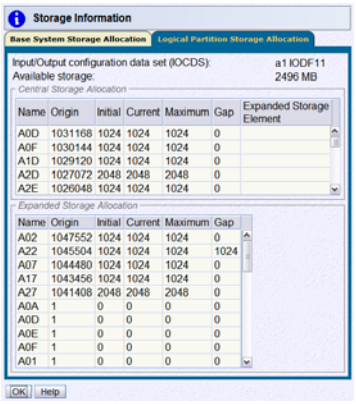
SI prefixes		Historical use		Binary prefixes	
Symbol (name)	Value	Symbol	Value	Symbol (name)	Value
kB (kilobyte)	$1000^1 = 10^3$	KB	$1024^1 = 2^{10}$	KiB (kibibyte)	2^{10}
MB (megabyte)	$1000^2 = 10^6$	MB	$1024^2 = 2^{20}$	MiB (mebibyte)	2^{20}
GB (**gigabyte**)	$1000^3 = 10^9$	GB	$1024^3 = 2^{30}$	GiB (gibibyte)	2^{30}
TB (terabyte)	$1000^4 = 10^{12}$	TB	$1024^4 = 2^{40}$	TiB (tebibyte)	2^{40}
PB (petabyte)	$1000^5 = 10^{15}$	PB	$1024^5 = 2^{50}$	PiB (pebibyte)	2^{50}
EB (exabyte)	$1000^6 = 10^{18}$	EB	$1024^6 = 2^{60}$	EiB (exbibyte)	2^{60}
ZB (zettabyte)	$1000^7 = 10^{21}$	ZB	$1024^7 = 2^{70}$	ZiB (zebibyte)	2^{70}
YB (yottabyte)	$1000^8 = 10^{24}$	YB	$1024^8 = 2^{80}$	YiB (yobibyte)	2^{80}

Memory size representation in the SE panels (cont..)

2^{10}	1,024	1,024 MB	1 GB
2^{11}	2,048	2,048 MB	2 GB
2^{12}	4,096	4,096 MB	4 GB
2^{13}	8,192	8,192 MB	8 GB
2^{14}	16,384	16,384 MB	16 GB
2^{15}	32,768	32,768 MB	32 GB
2^{16}	65,536	65,536 MB	64 GB
2^{17}	131,072	131,072 MB*	128 GB
2^{18}	262,144	262,144 MB	256 GB
2^{19}	524,288	524,288 MB	512 GB
2^{20}	1,048,576	1,048,576 MB	1 TB

1,048,576 – 1024 = 1,047,552

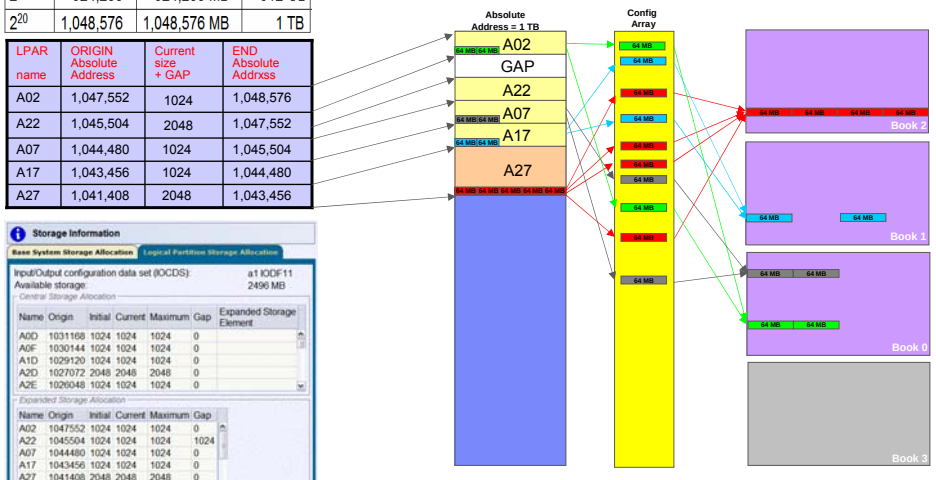




Physical Memory Allocation

2^{19}	524,288	524,288 MB	512 GB
2^{20}	1,048,576	1,048,576 MB	1 TB

LPAR name	ORIGIN Absolute Address	Current size + GAP	END Absolute Addrxss
A02	1,047,552	1024	1,048,576
A22	1,045,504	2048	1,047,552
A07	1,044,480	1024	1,045,504
A17	1,043,456	1024	1,044,480
A27	1,041,408	2048	1,043,456



Questions ?

Thank You !

- Luiz A. Fadel
  - fadel@br.ibm.com
- Ewerson Palacio
  - bird@br.ibm.com

