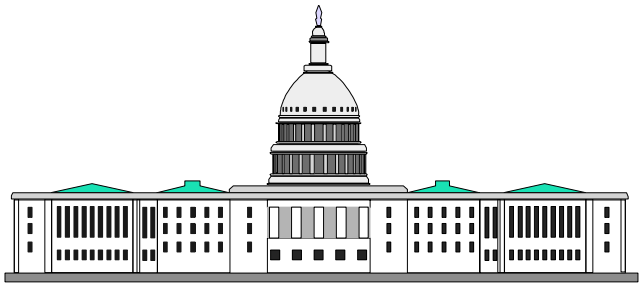
LPAR Advanced Topics

IBM z/OS and OS/390 Expo
October 7 - 11, 2002
Session T09



Harv Emery
emeryh@us.ibm.com
Washington Systems Center

Trademarks

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

APPN*
DB2*
e-business logo*
Enterprise Storage Systems
ESCON*
FICON
GDPS
Geographically Dispersed Parallel Sysplex
HiperSockets
IBM*
IBM logo*
IMS
Magstar*
MVS
Netfinity*
OS/390*
Parallel Sysplex*
PR/SM
S/390*
S/390 Parallel Enterprise Server
Virtual Image Facility
VM/ESA*
VSE/ESA
VTAM*
WebSphere
z/Architecture
z/OS
z/OS.e
z/VM
zSeries

* Registered trademarks of IBM Corporation

The following are trademarks or registered trademarks of other companies.

Lotus, Notes, and Domino are trademarks or registered trademarks of Lotus Development Corporation
LINUX is a registered trademark of Linus Torvalds
Penguin (Tux) complements of Larry Ewing
Tivoli is a trademark of Tivoli Systems Inc.
Java and all Java-related trademarks and logos are trademarks or registered trademarks of Sun Microsystems, Inc., in the United States and other countries
UNIX is a registered trademark of The Open Group in the United States and other countries.
Microsoft, Windows and Windows NT are registered trademarks of Microsoft Corporation.
SET and Secure Electronic Transaction are trademarks owned by SET Secure Electronic Transaction LLC.

* 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.

IBM considers a product “Year 2000 ready” if the product, when used in accordance with its associated documentation, is capable of correctly processing, providing and/or receiving date data within and between the 20th and 21st centuries, provided that all products (for example, hardware, software and firmware) used with the product properly exchange accurate date data with it. Any statements concerning the Year 2000 readiness of any IBM products contained in this presentation are Year 2000 Readiness Disclosures, subject to the Year 2000 Information and Readiness Disclosure Act of 1998.

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.

LPAR Advanced Topics

Introduction to z800 and Running z/OS.e

zSeries PR/SM Workload Pricing Support

zSeries PR/SM IRD Support

- CPU Management for z/OS
- CPU Management for z/VM and Linux
- CSS I/O Priority Queuing
- Dynamic Channel Path Management

HiperSockets (z/OS V1.2 & up)

Fibre Channel Protocol (FCP)

zSeries Coupling Facility Support

Capacity Upgrade on Demand (G5/6 and zSeries)

- Concurrent Memory Upgrade (z900 Dr 3C)
- Nondisruptive CBU CP Downgrade (z900 Dr 3C and z800)

Memory Configuration and Reconfiguration

- Support for CUoD - CIU, CBU and MES Upgrade
- MVS Storage Reconfiguration

Ref: zSeries PR/SM Planning, SB10-7033-03a (July 2002)

IBM zSeries 800

Complete z/Architecture (64 bit)

- OS/390 V2.8 & up, z/OS all, z/OS.e
- VM/ESA V2.4 & up, z/VM all
- VSE/ESA V2.4 & up, TPF 4.1
- Linux Kernel 2.2 & 2.4 (31 and 64 bit)

Flexible Model Structure

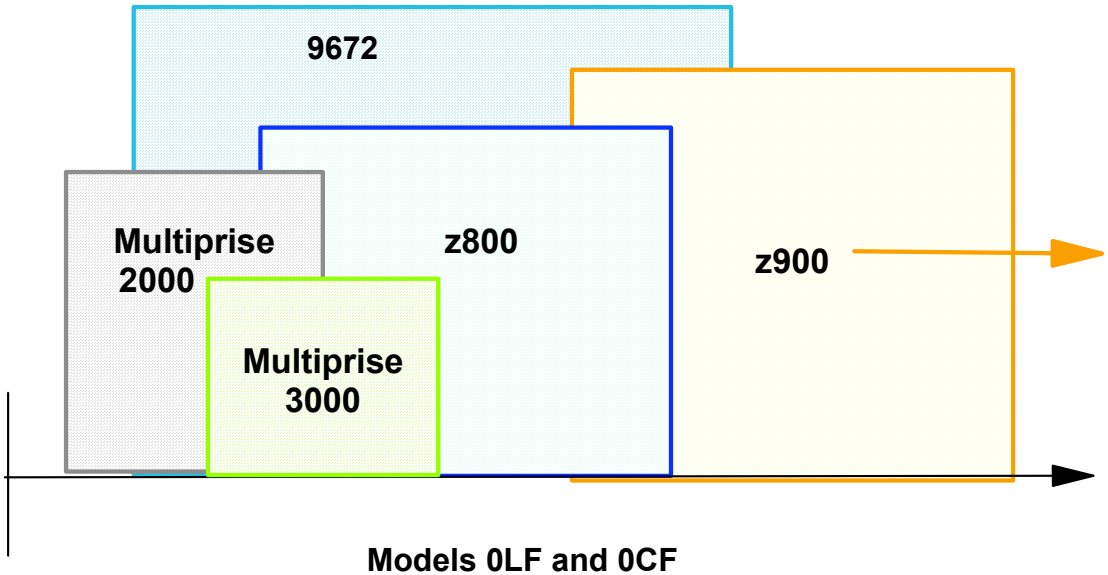
- 1 to 4-way
- z800-001 ITRR close to G6-X17
- 3 sub-uni, 1 sub-dyadic
- CUoD and CBU
- Linux Model 0LF, CF Model 0CF
- z800-004 upgrades to z900-104

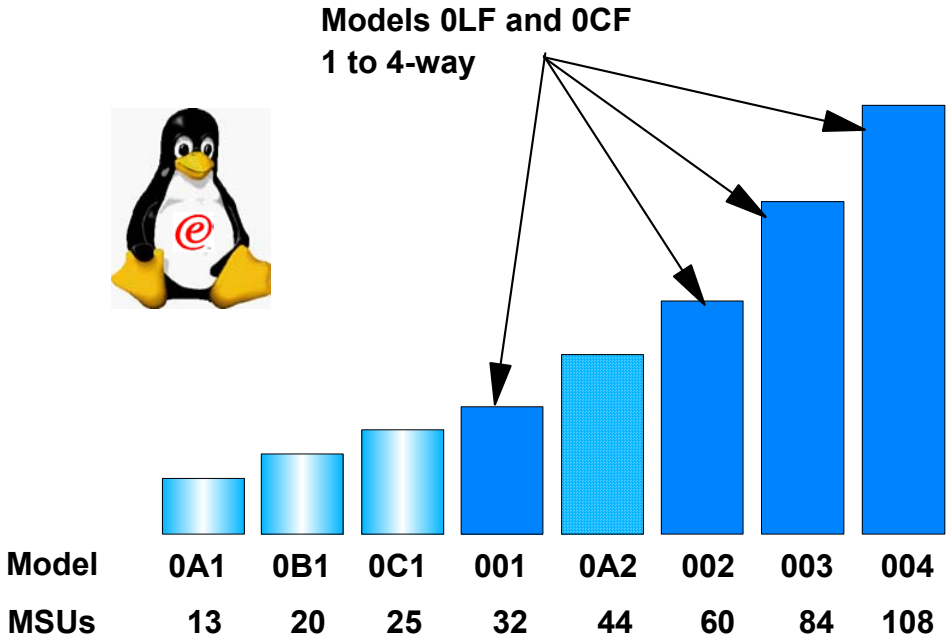
8, 16, 24 or 32 GB memory

- No concurrent upgrade

zSeries I/O Subsystem supports

- All zSeries I/O cards (16 max)
  - Up to 240 Escon
    - No Parallel, OSA-2 FDDI, or ICB-2
- SOD: Linux FCP support





z800 LPAR Mode Exclusive: z/OS.e

z/OS.e Machine Support

- z/Architecture required - 64-bit only
- z800 - YES
- z900 - No! Will detect and fail.

z/OS.e Operating Mode

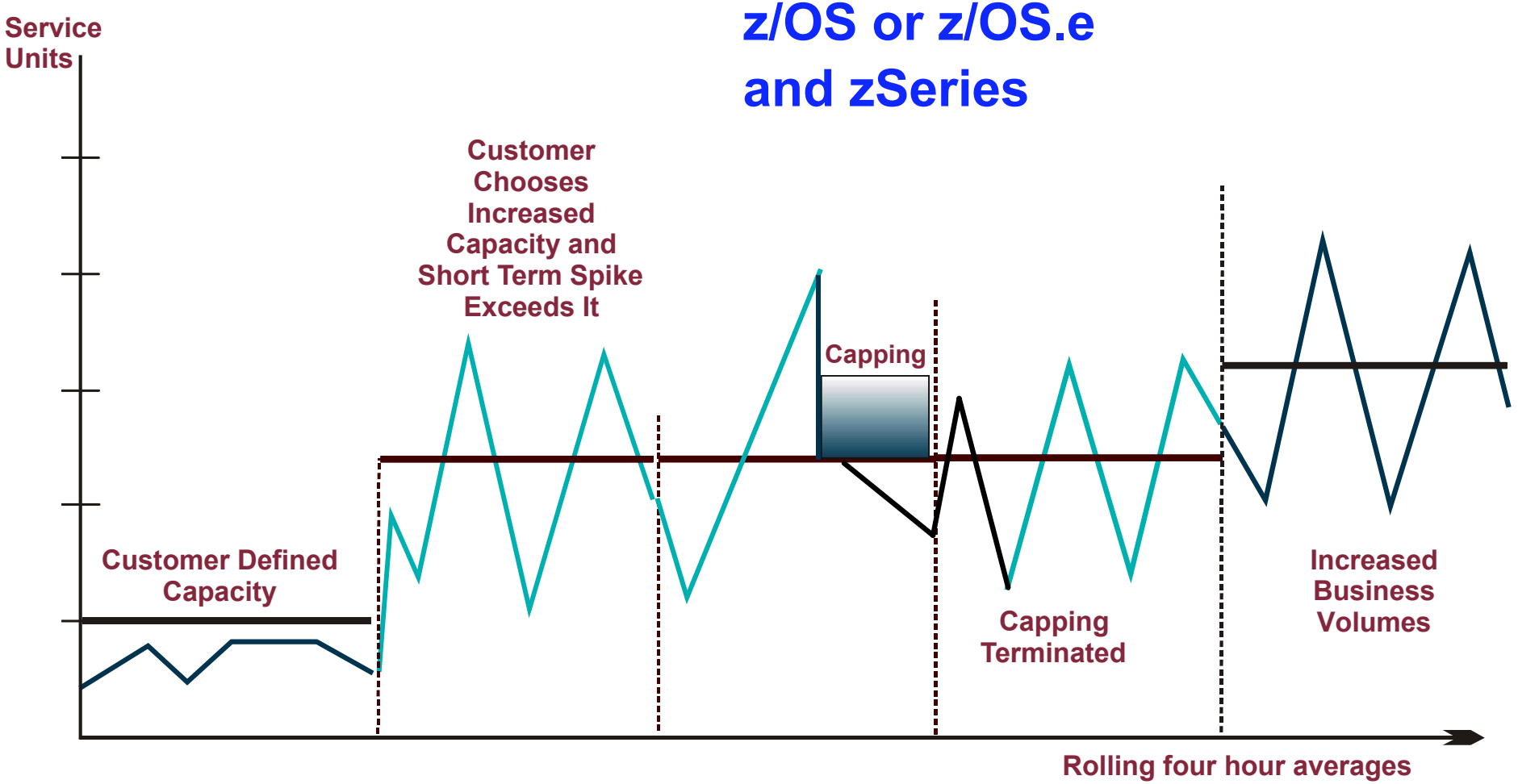
- LPAR Mode - YES
- Basic Mode - No! Will detect and fail.

z/OS.e LPAR Requirements

- Name in IOCDS must start with "ZOSE"
- "ZOSE" named LPAR on z800:
  - z/OS.e and z/OS.e under z/VM will run
  - z/OS and OS/390 will NOT run, even under z/VM
- "ZOSE" named LPAR on S/390 or z900:
  - Name has NO effect, z/OS or OS/390 will run
- Shared CP - OK, Requires WLM "Defined Capacity"
- Dedicated CP - OK

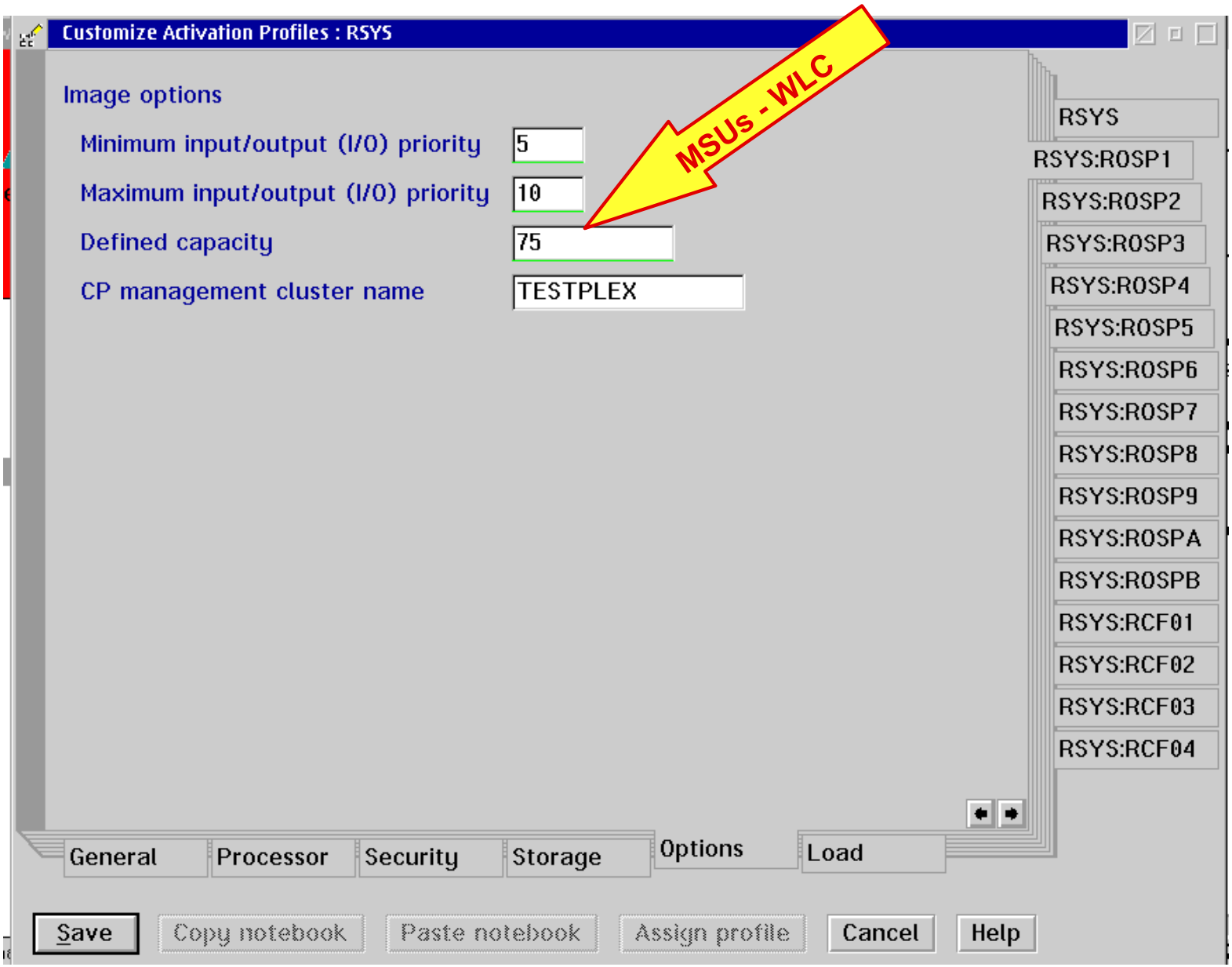
Variable Workload License Charge
PR/SM Management to Defined Capacity

z/OS or z/OS.e and zSeries

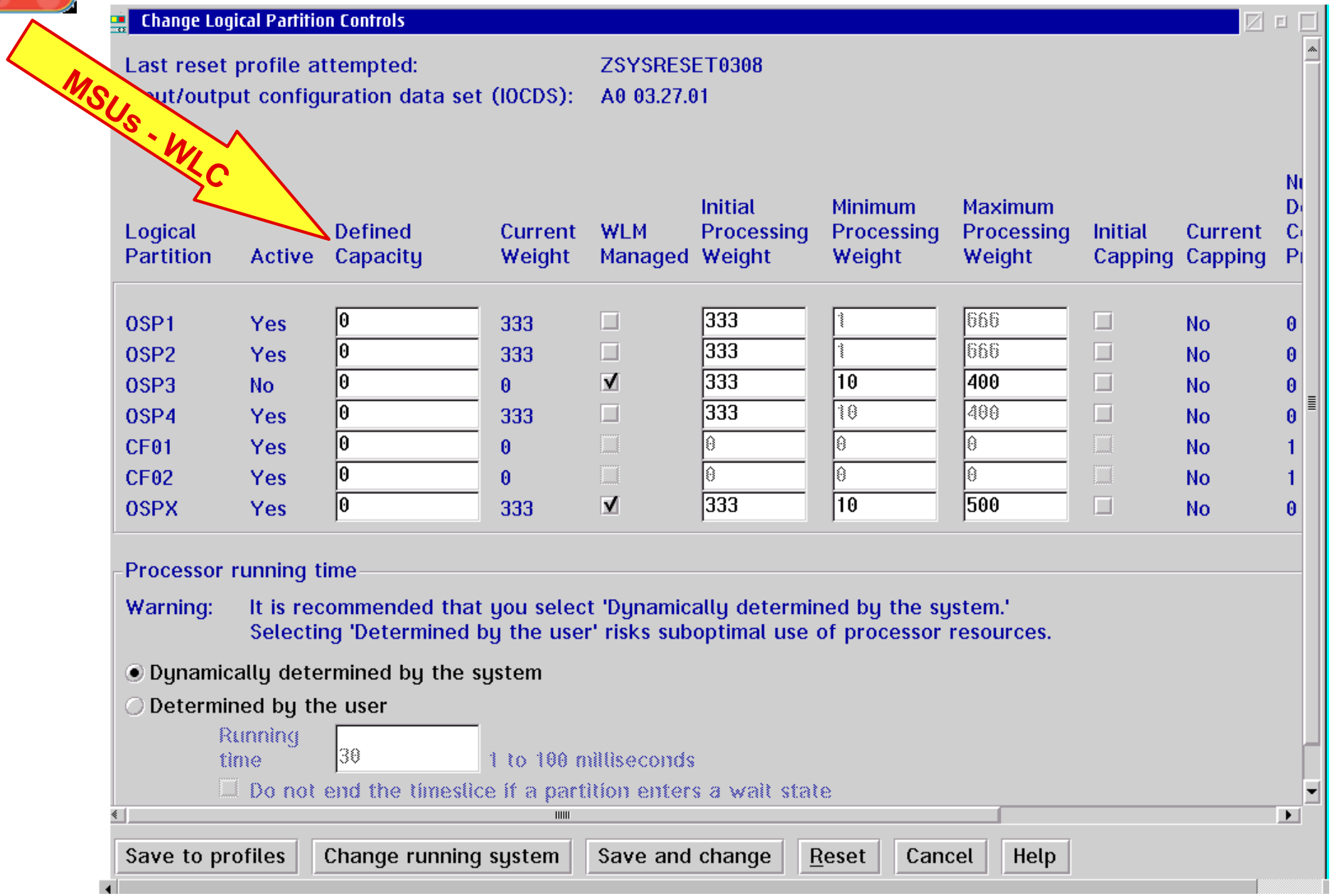


Pay for what you define!

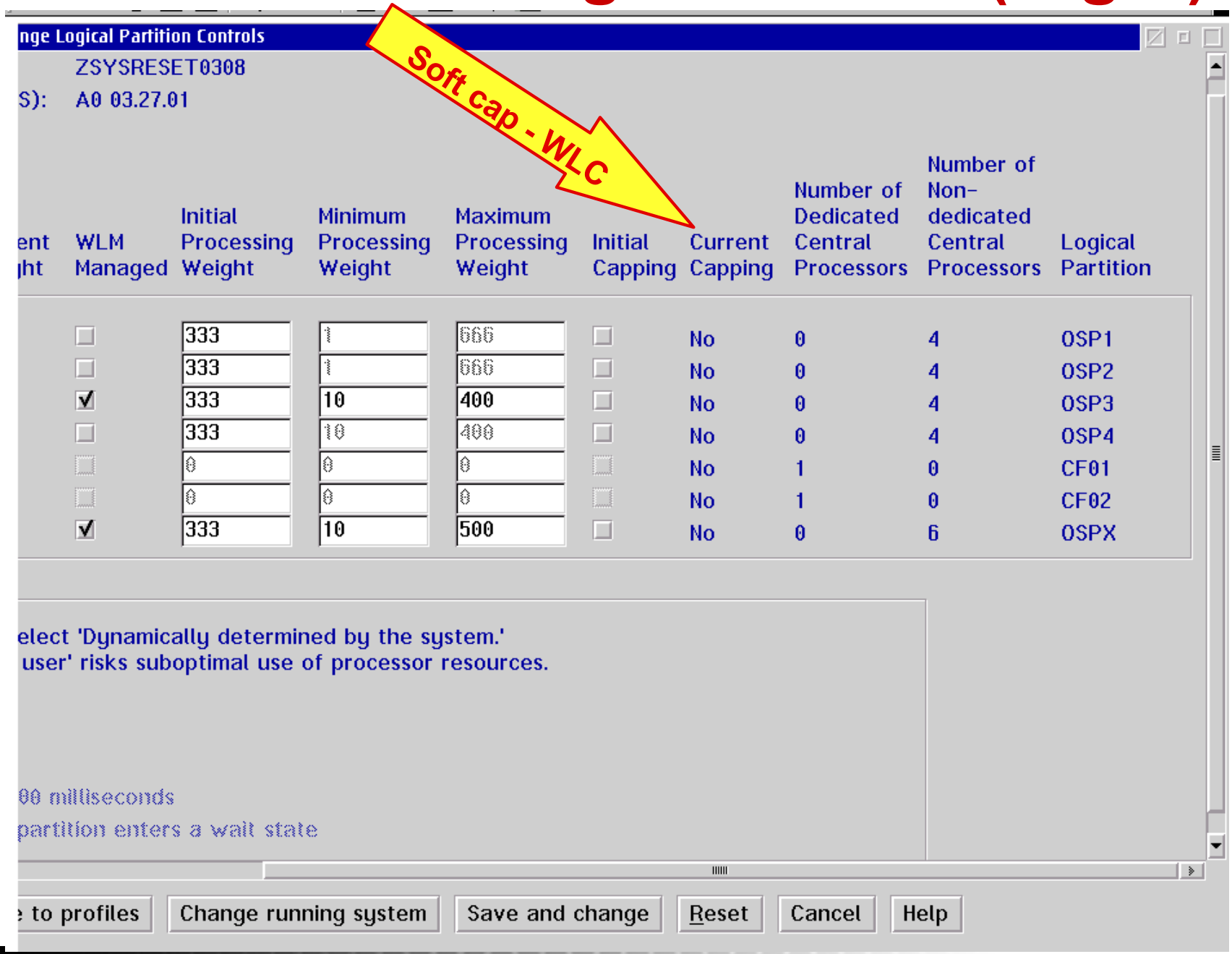
Image Profile Options Tab



HMC/SE Change Controls (Left)



HMC/SE Change Controls (Right)



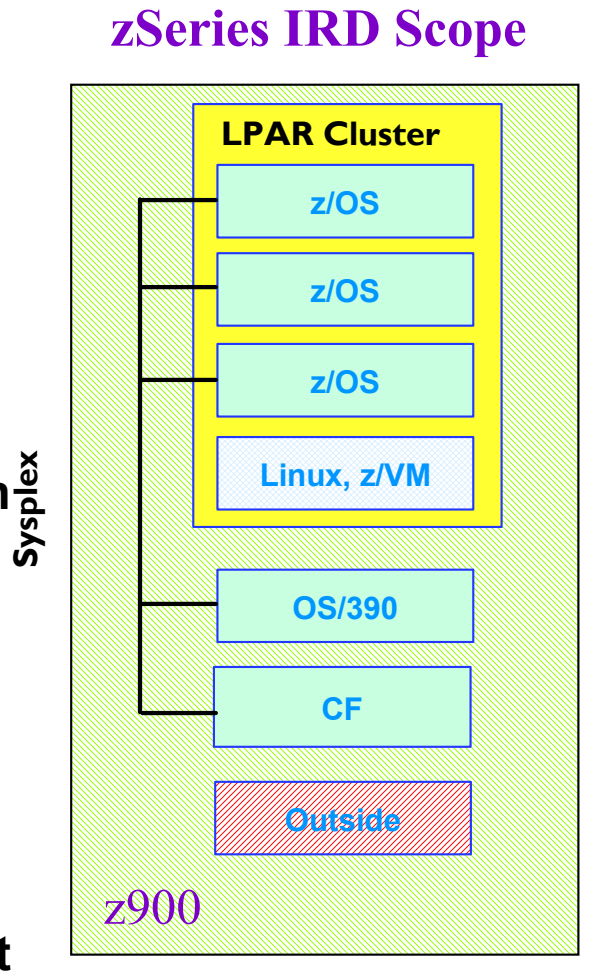
Intelligent Resource Director

Leverage platform strengths through integration

- Workload Manager
- Parallel Sysplex
- PR/SM
- Channel Subsystem

View a cluster of LPs on a zSeries as single pool of computing resource

- Move physical resource to priority workloads in an LPAR cluster
- Extend goal oriented resource management across logical partitions transparently to application subsystems
- Initial resources managed: CPU and I/O
- Requires Parallel Sysplex, WLM Goal Mode, WLM Structure and Level 9 Coupling Facility
- z/OS V1.2 adds z/VM and Linux for zSeries support for LPAR weight management



Intelligent Resource Director

LPAR CPU Management

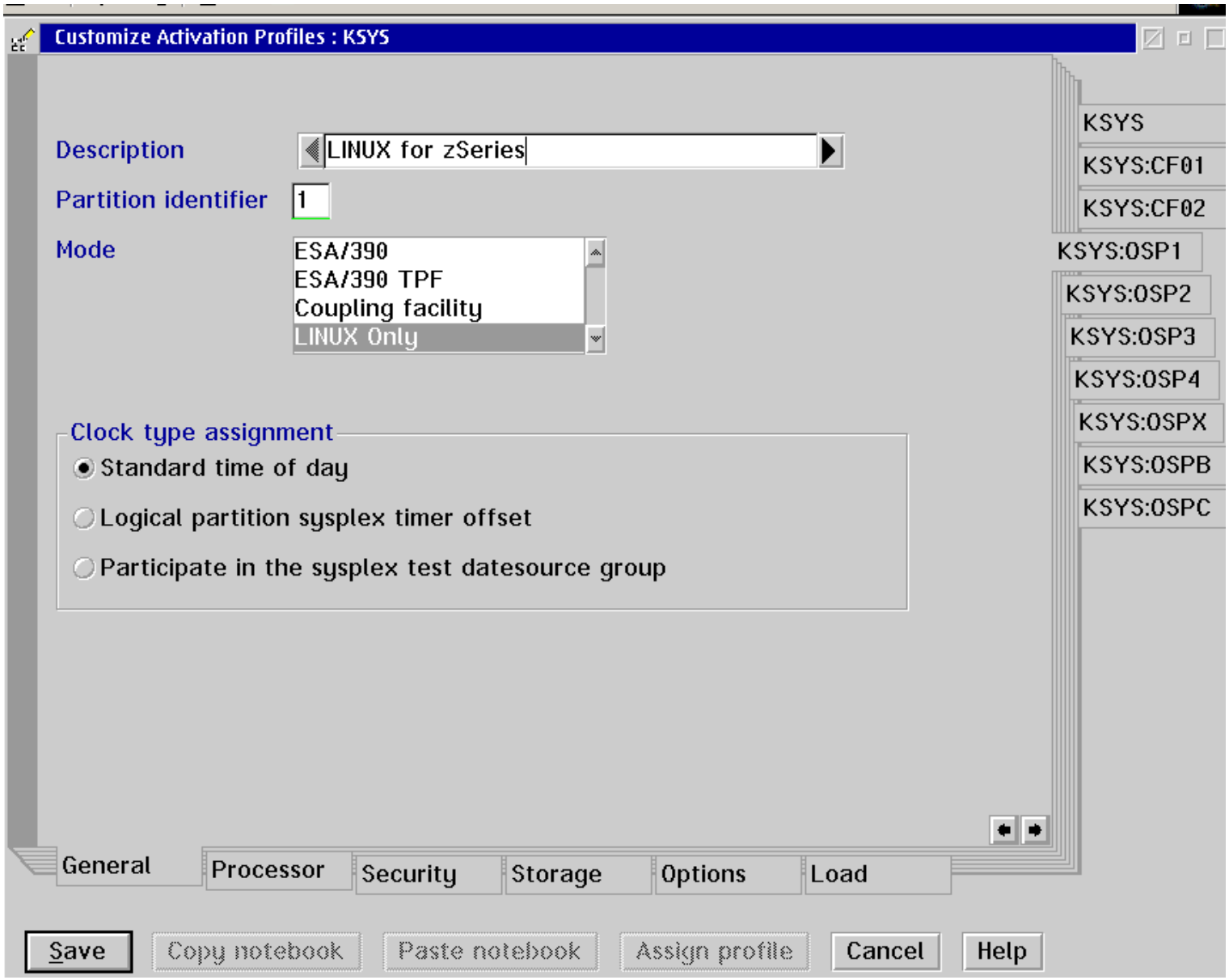
Description

- **LPAR Weight Management - z/OS, z/VM and Linux for zSeries**
  - Dynamically manages a partition's CPU access based on workload demands and goals (z/VM and Linux on shared CPs only)
- **Vary Logical CPU Management - z/OS Only**
  - Optimizes number of logical CPs based on partition's current weight and CPU consumption

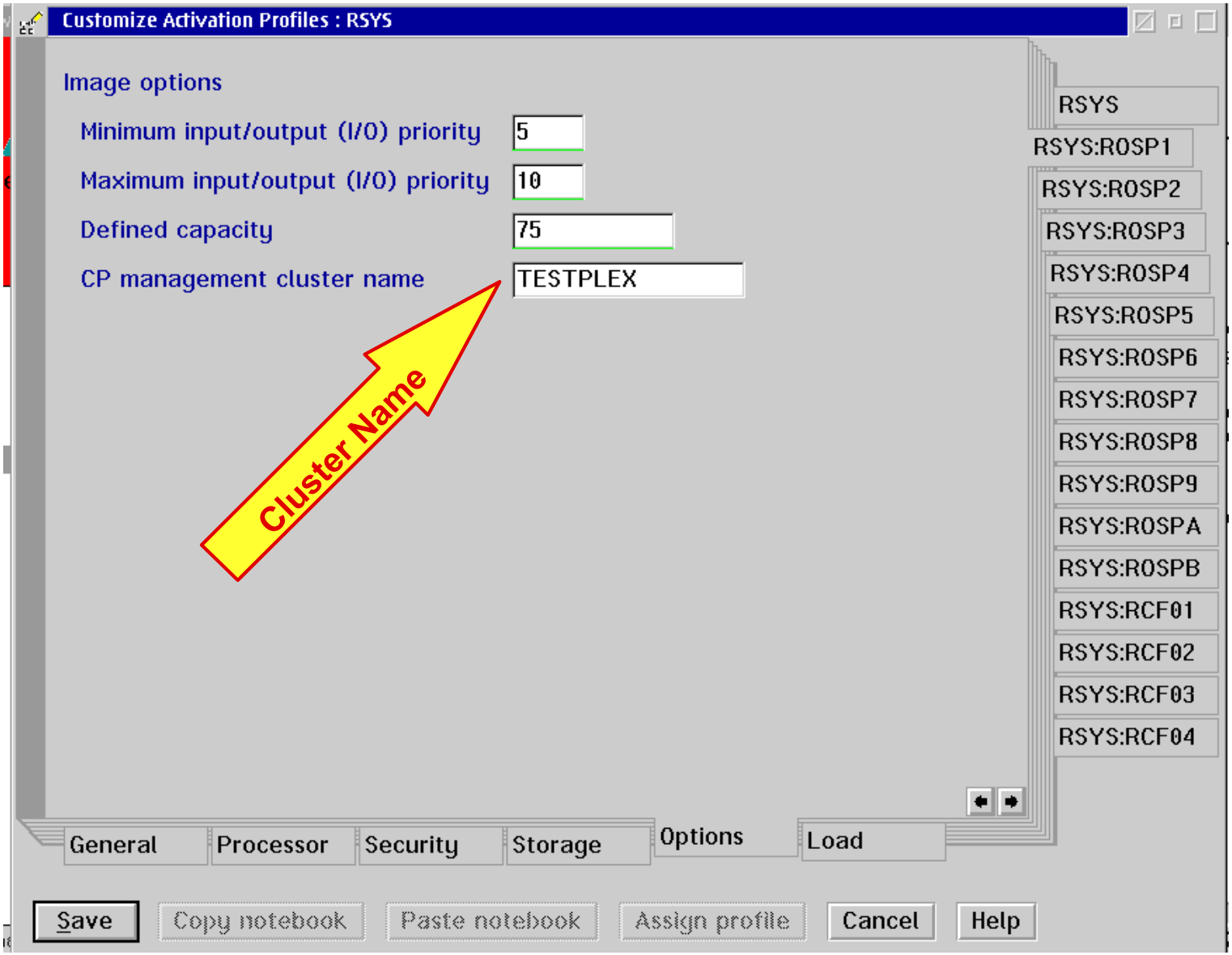
Benefits

- **Provides flexibility in managing** CPU **resources across logical partitions in accordance with workload goals.**
  - Dynamic change of LPAR weights
  - Manage tradeoffs between meeting service goals for work and making efficient use of a system's resource
  - Prevent or mitigate possible problems
  - Provides the fastest Uniprocessor speed for single tasking workloads
  - Reduces LPAR overhead

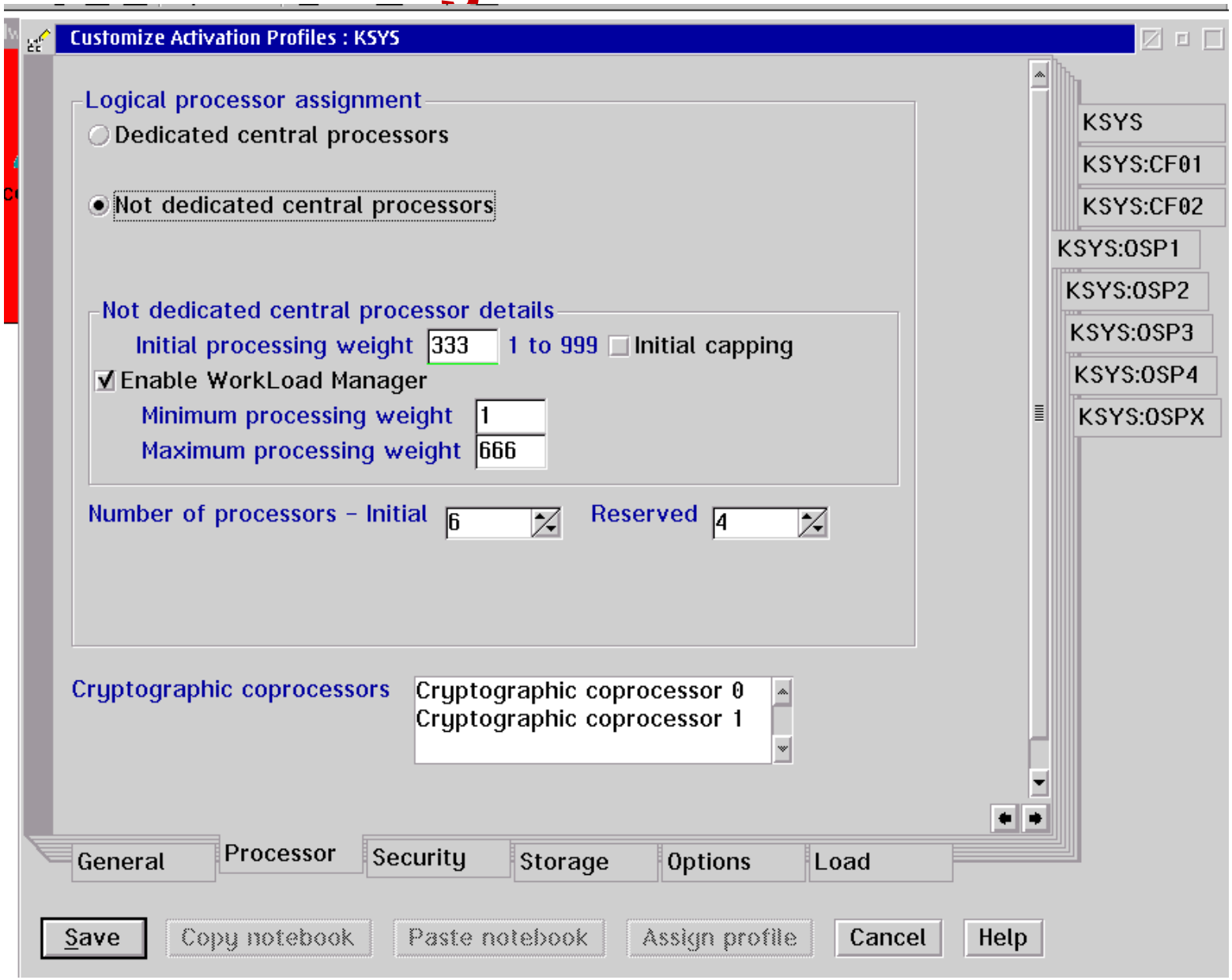
zSeries Image Profile Control Partition Mode Selection



zSeries Image Profile Control CP Management Cluster

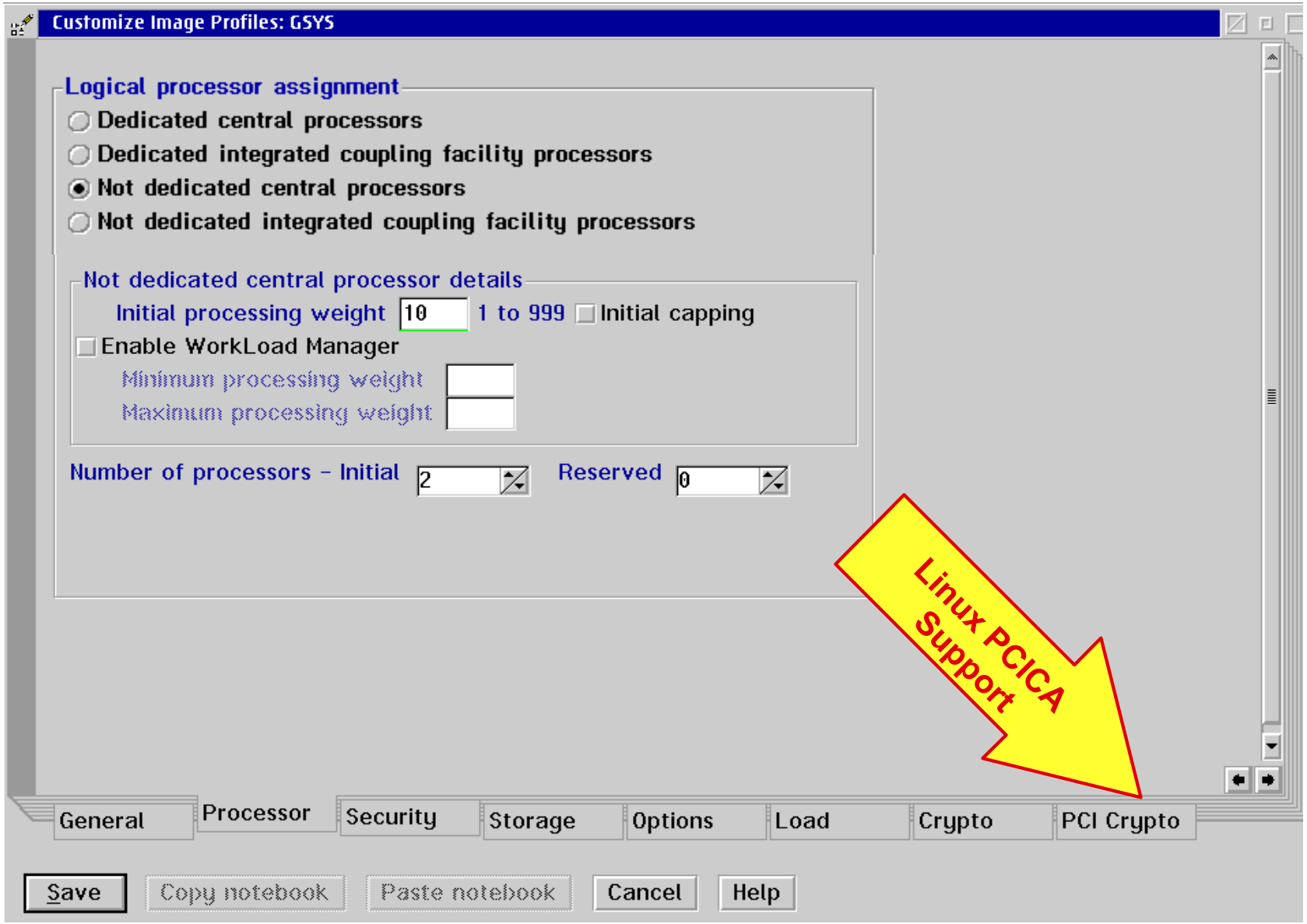


zSeries Image Profile Control
LP Assignment - ESA/390 Mode

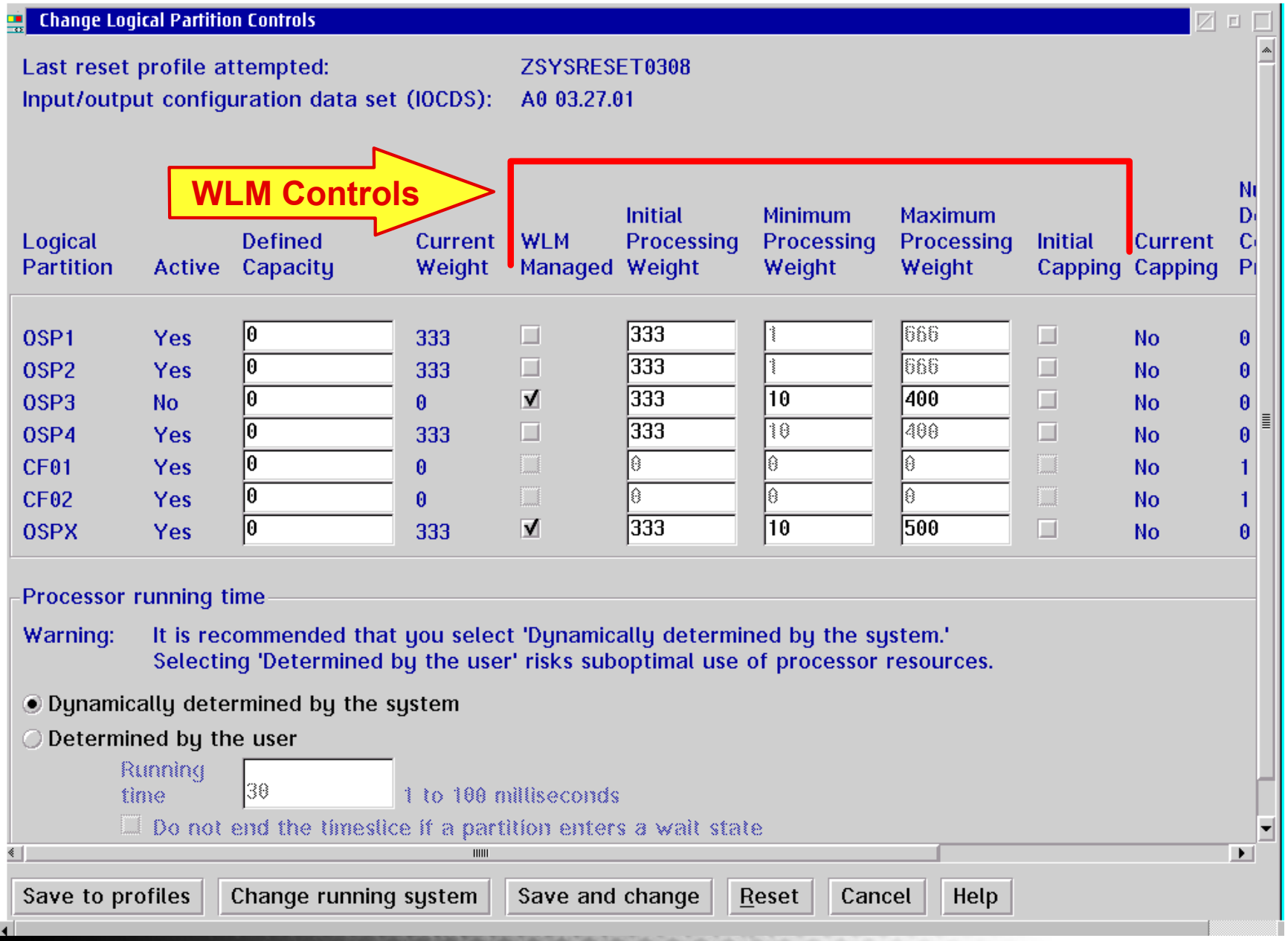


zSeries Image Profile Control LP Assignment - Linux Mode

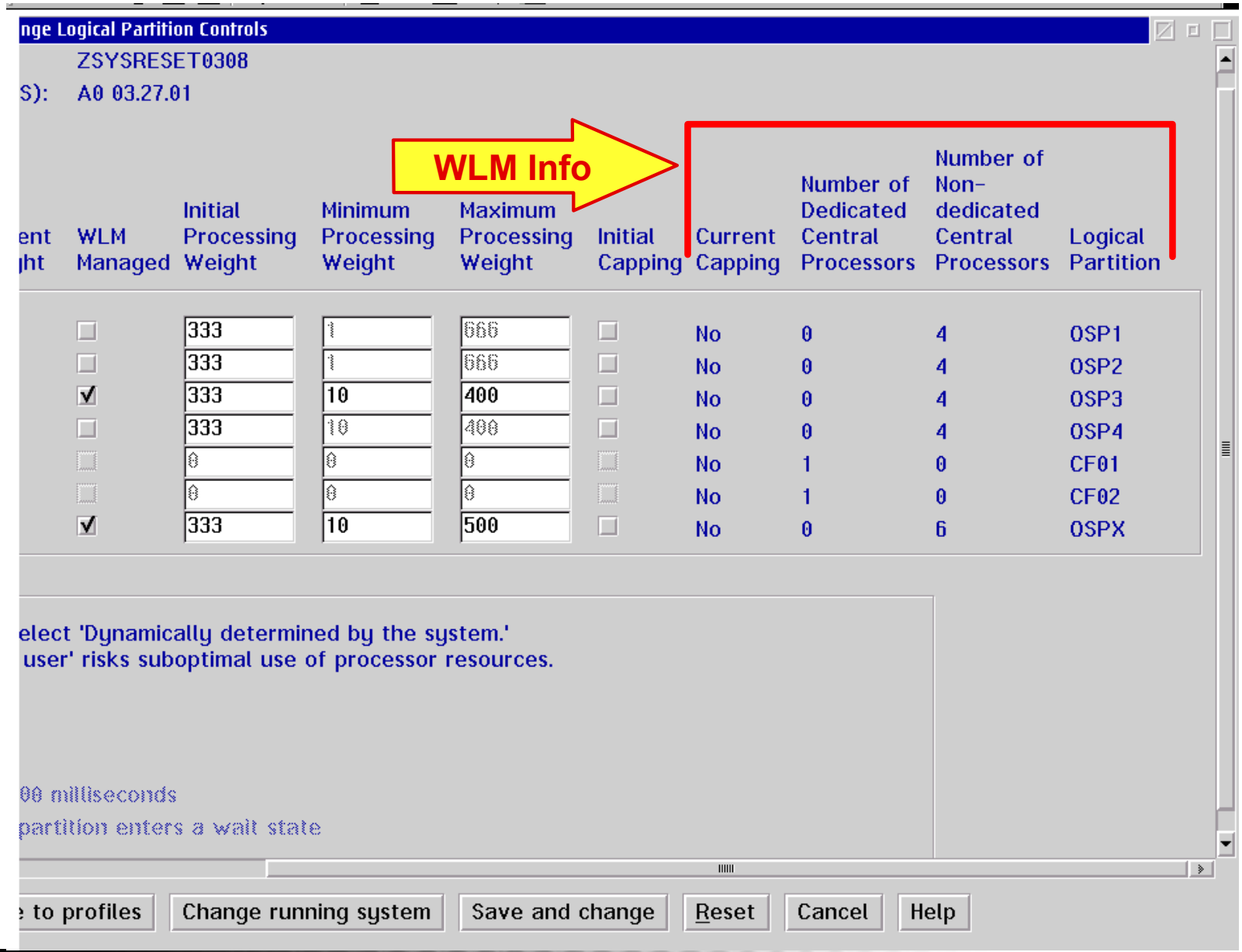
Note: To PR/SM IFL = ICF



zSeries Change Logical Partition Controls CPU Management



zSeries Change Logical Partition Controls CPU Management



Intelligent Resource Director

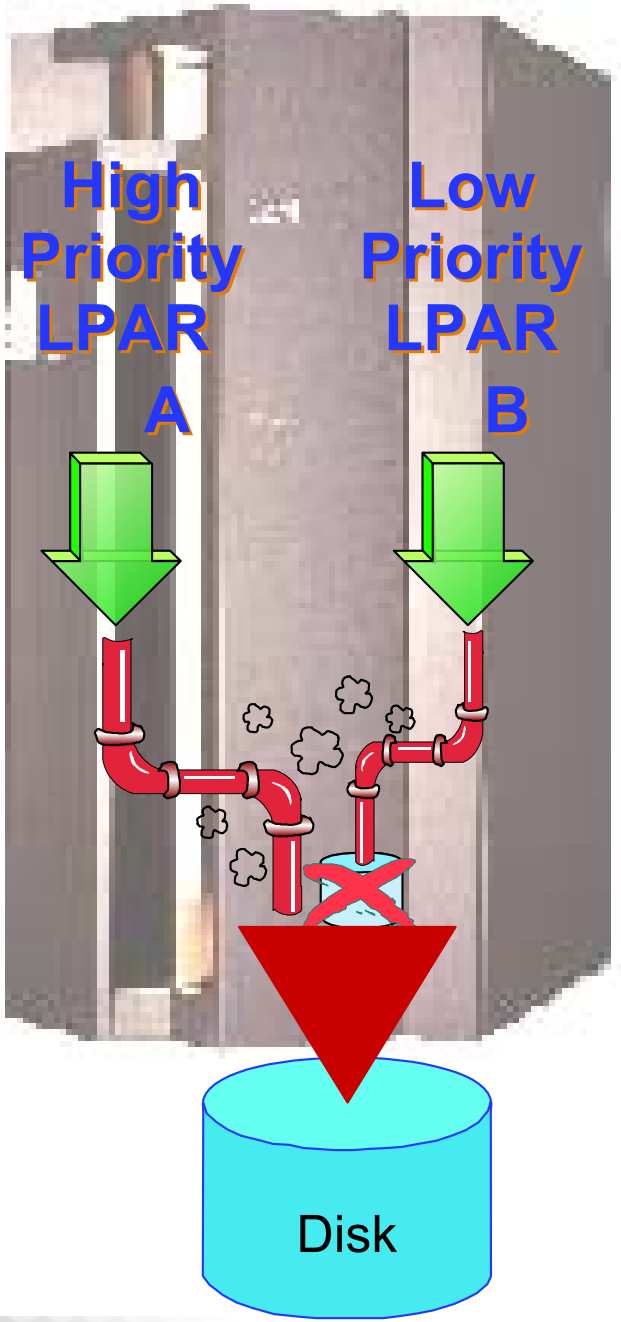
Channel Subsystem Priority Queuing

Description

- I/O Priority Queuing prioritizes I/O within an LPAR across workloads
  - Available since OS/390 V1.3
- Channel Subsystem Priority Queuing prioritizes I/O within an LPAR cluster
  - LPAR priorities based on workload goals
  - Exclusive to zSeries
- z/VM and Linux for zSeries - static only

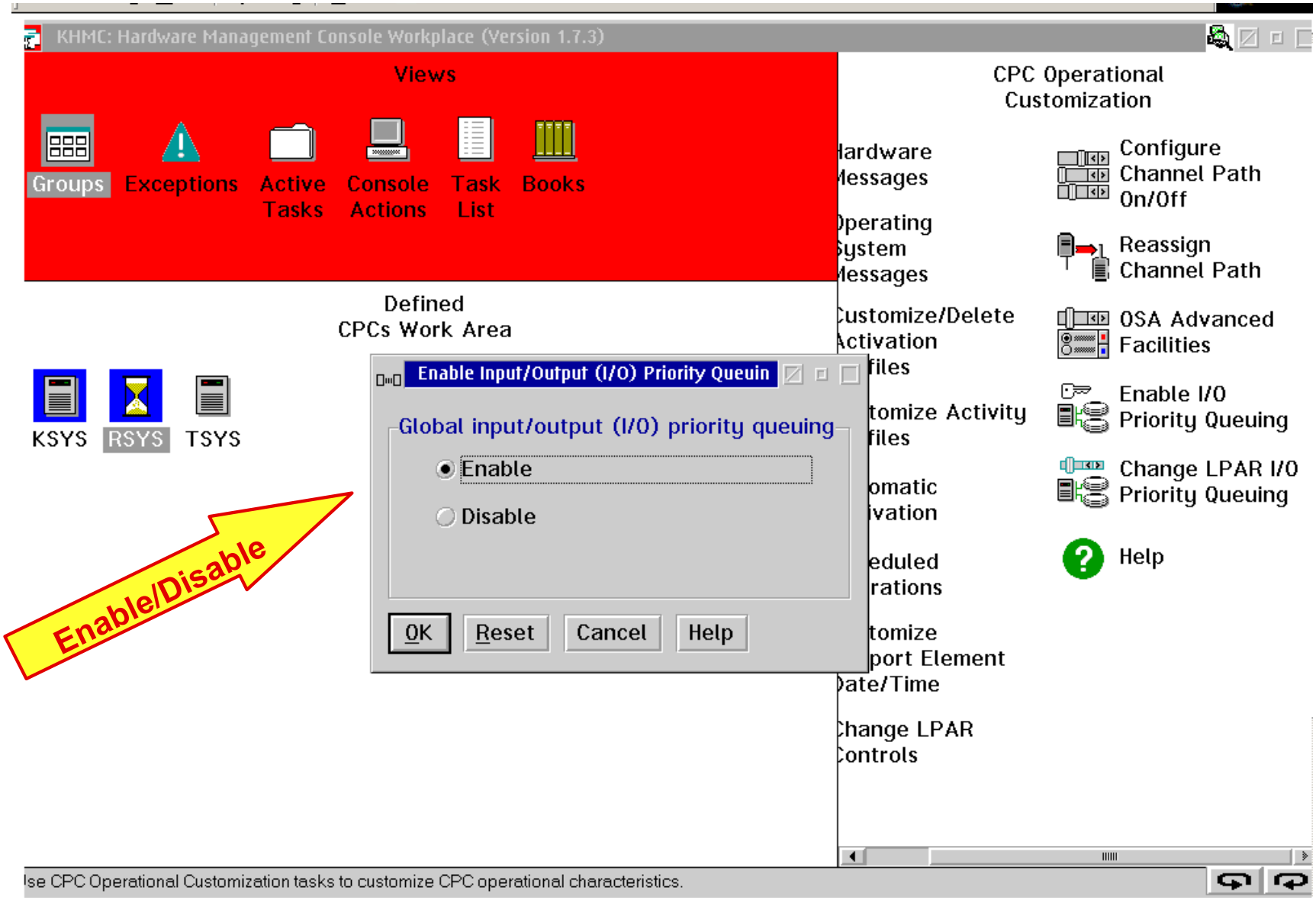
Benefits

- Allows better channel resource management with EMIF
  - Low priority work will not preempt high priority work from other LPARs

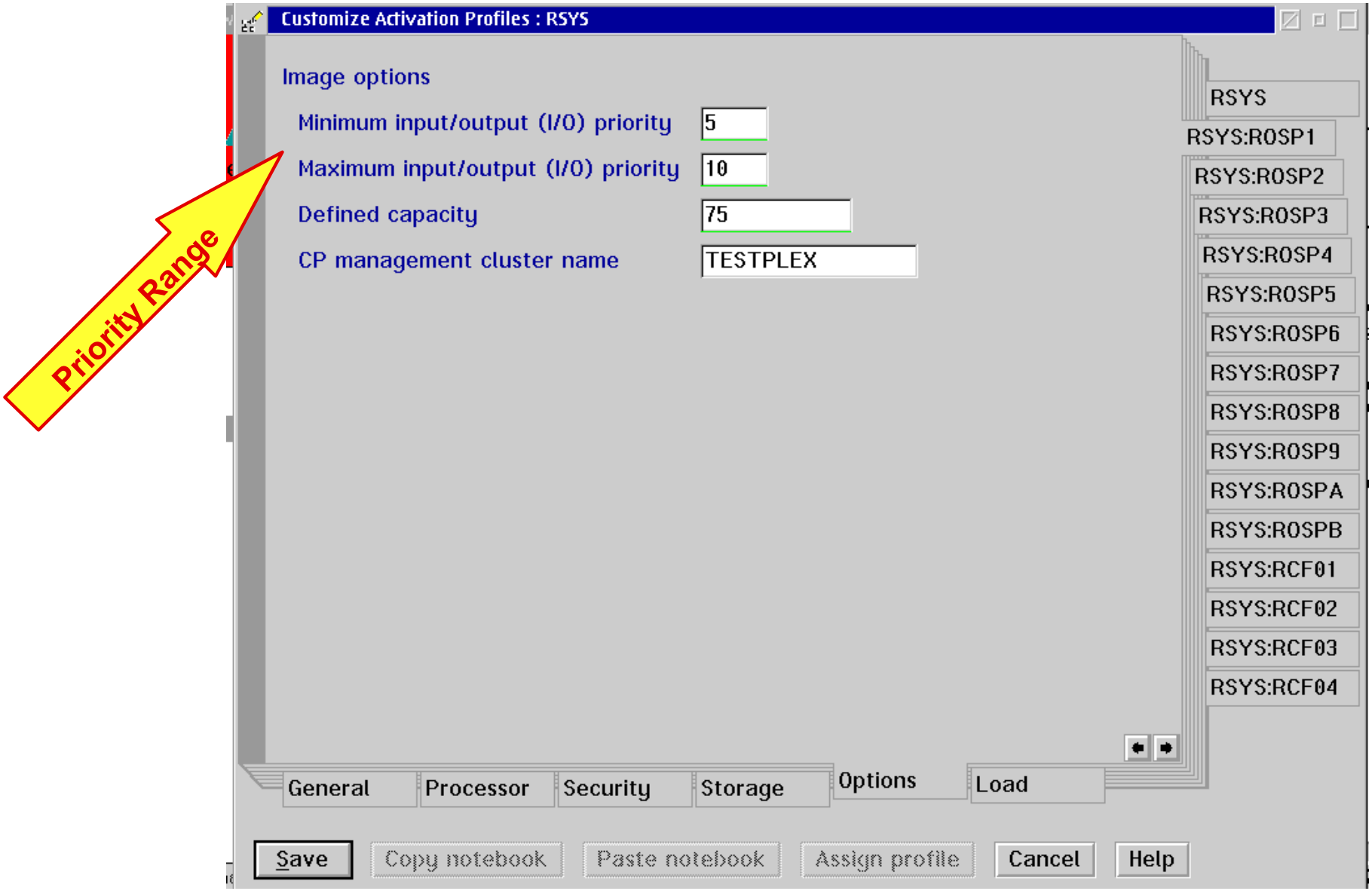


zSeries Reset Profile Control
CSS I/O Priority Queuing

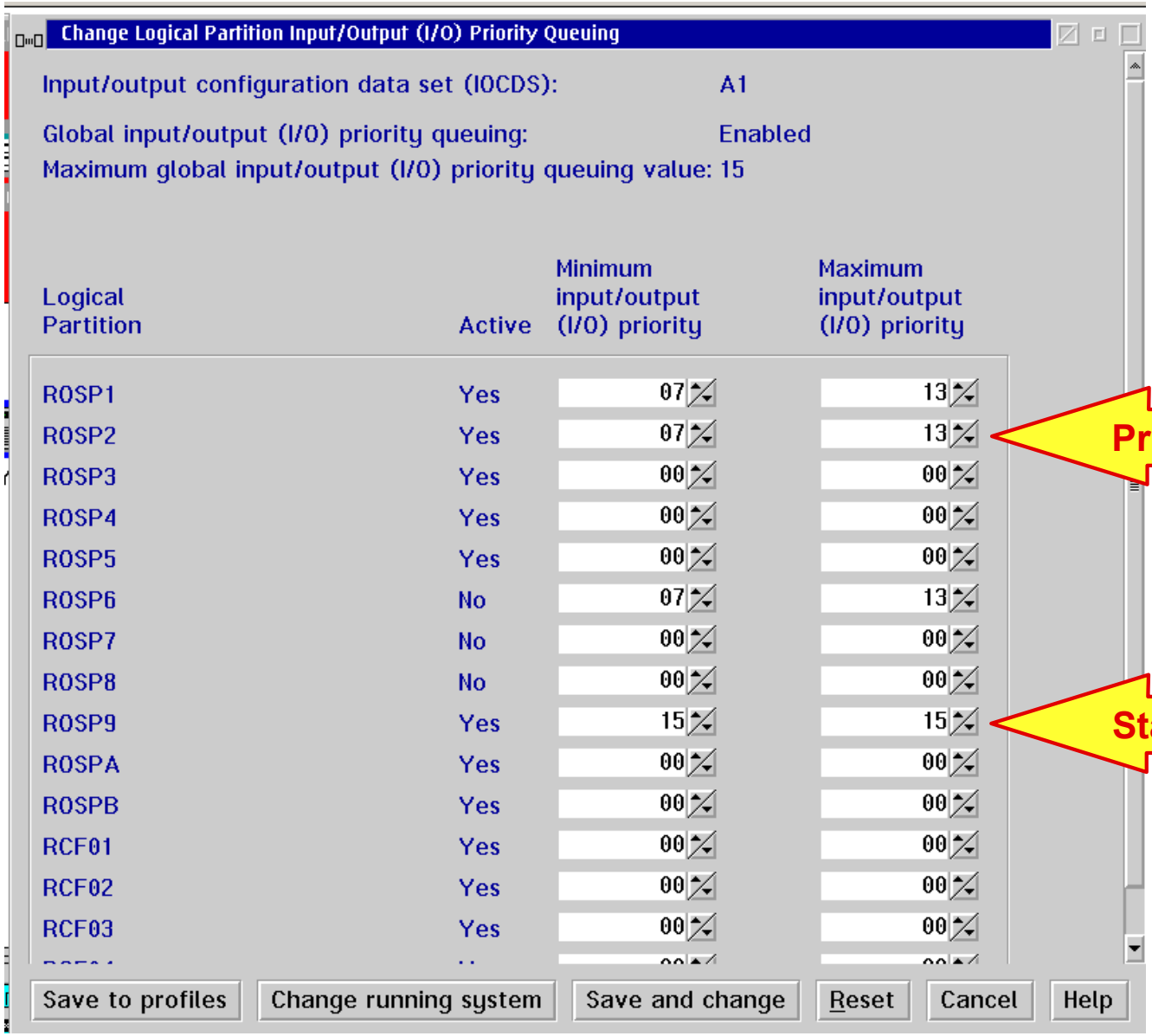
zSeries System-wide Control CSS I/O Priority Queuing



zSeries Image Profile Control CSS I/O Priority Queuing



zSeries Change Logical Partition CSS I/O Priority Queuing



Priority Range

Static Priority

Intelligent Resource Director

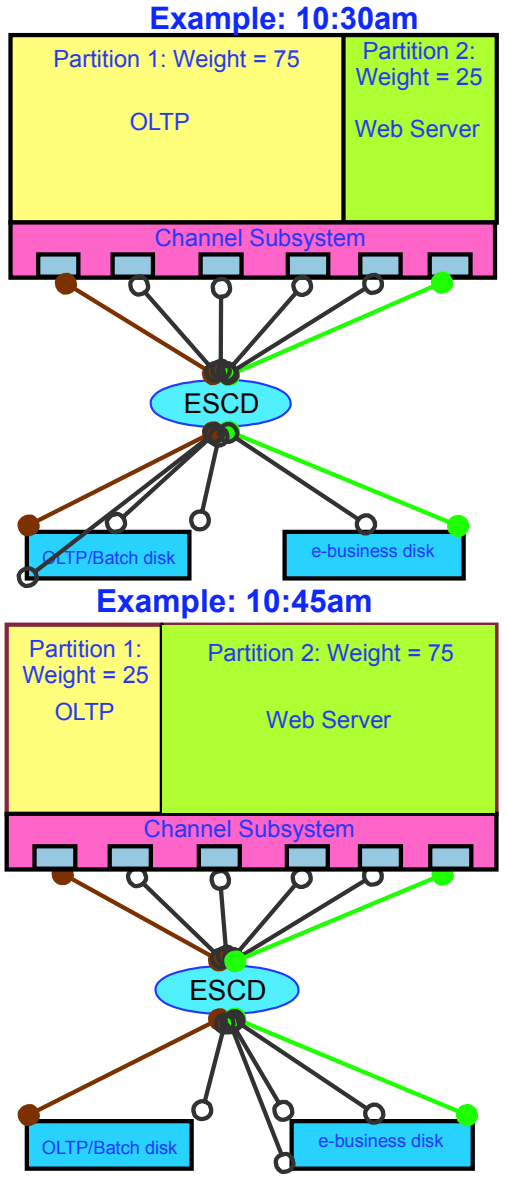
Dynamic Channel Path Management

Description

- Allows the system to dynamically manage channel paths in response to changing workload demands
- Moves channel capacity (bandwidth) to the disk subsystem(s) experiencing need based on workload requirements
- Optimized with Channel Subsystem Priority Queuing
- Exclusive to zSeries and z/OS

Benefits

- More efficient use of hardware resource
- Reduces channel requirements
- Simplifies I/O configuration planning and definition
- Dynamically balances I/O connectivity based on workload demand



I/O Definition for DCM

Rules for managed channels:

- zSeries CEC only
- DCM supports ESCON (CNC) and FICON Conversion (FCV)
- ESCON Director attachment is required
- Specify the I/O cluster name to "own and operate"
- Do NOT specify a candidate list, candidates are IOCLUSTER members
- Do NOT specify as a PATH for any Control unit

Rules for CUs to accept managed paths:

- Specify some shared, "static" paths, these will be the only paths at IPL
- Specify placeholder asterisks for managed paths

Examples (IOCP output from HCD)

- Managed ESCON channel 40 attached to ESCON director 1B:

```
CHAN40   CHPID     PATH=40,TYPE=CNC,SWITCH=1B,
         OS=01,IOCLUSTER=TESTPLEX
```

- Control unit 1234 with two static, two managed paths:

```
CU1234   CNTLUNIT CUNUMBR=1234,UNIT=2105,.......
         PATH=(20,30,**,**),LINK=(C0,C0,**,**)
```

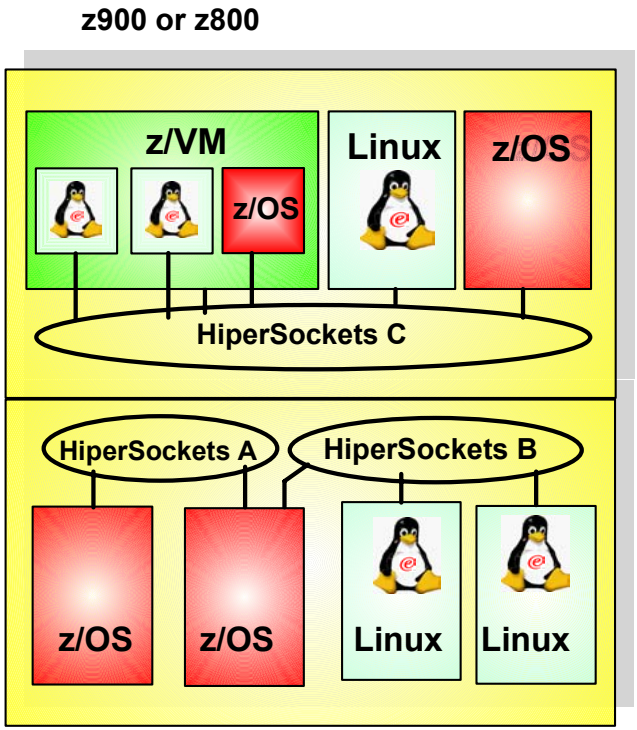
HiperSockets - The Network in the Box

IP networking among virtual servers in a zSeries

- Up to four HiperSockets IQD channels - "internal LANs" accessible by all partitions
- LAN frame size configurable - 16K, 24K, 40K, 64K
- Up to 1,024 TCP/IP stacks across all four HiperSockets
- iQDIO data transfer (similar to cross-address-space memory move) using memory bus does not use CP cache = minimal performance impact
- Each partition configures its own usage as it would another type of IP network.

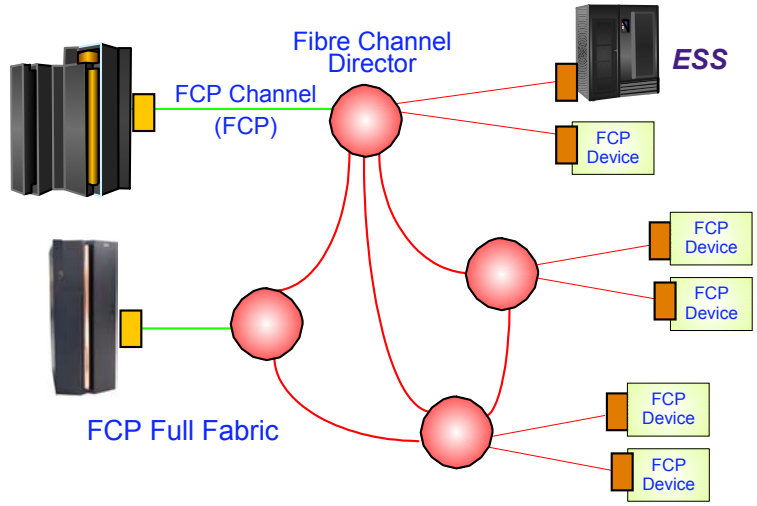
Prerequisites

- z/OS 1.2 & up or z/OS.e (z800 only)
- z/VM 4.2
- Linux kernel 2.4 (64- and 31-bit)
- z900 (Driver 3C or 3G) or z800

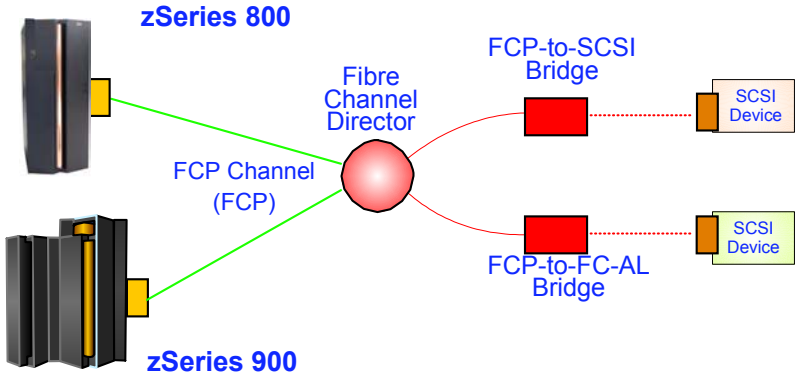


zSeries FCP for Linux Limited Availability

Initially, FCP support with limited availability



- **FCP Full Fabric Connectivity**
  - Homogeneous, single vendor fabric
  - Fibre channel directors
    - INRANGE FC/9000 (IBM 2042)
    - McDATA ED-6064/ 5000 (IBM 2032-064/ 2032-001)
    - IBM 2109-F16/F08/S16/S08 (Brocade 3800/3200/2800/2400)
  - Devices
    - Enterprise Storage Server

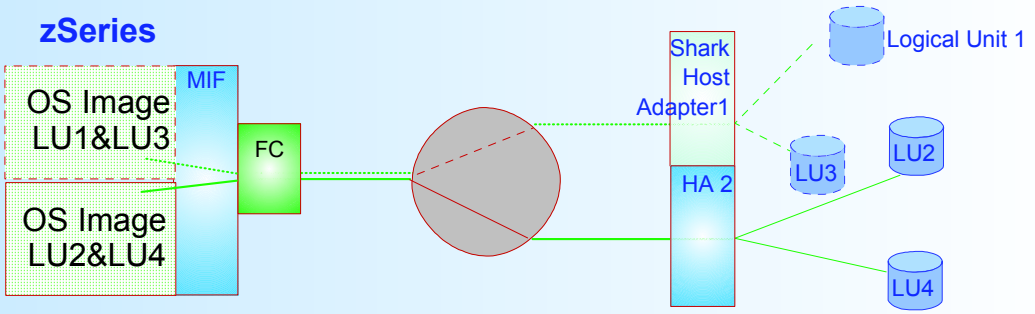


- **FCP to SCSI Bridge**
  - FCP-to-SCSI Bridges (via switch)
    - IBM 2108-G07 SAN Data Gateway
  - FCP-to-FC-AL bridge (via switch)
    - McDATA ES-1000 Loop Switch (IBM ES-2031-L00)

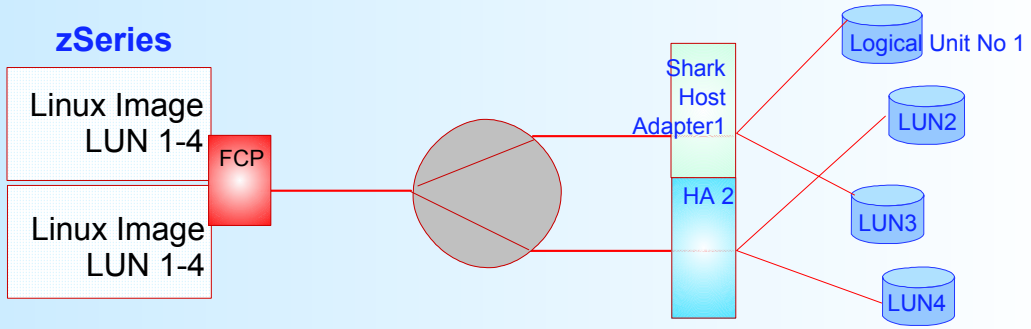
Initially, no direct attachment to device or bridge

FCP and MIF Channel Sharing

- MIF can ensure data integrity and security when FICON or ESCON channels are shared if device candidate lists (DCLs) are used.

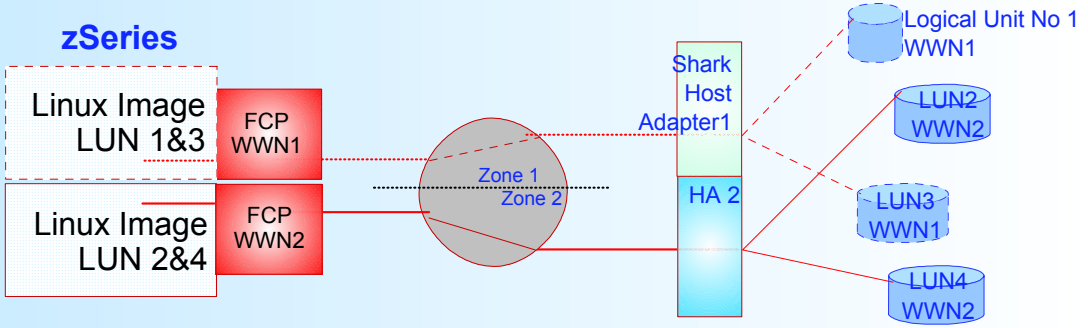


- MIF cannot ensure data integrity and security when FCP channels are shared - FCP/SCSI does not allow.
- All Linux images can define and access all LUNs
- One image at a time, first-come, first-served.



- Data integrity for Linux and FCP requires the use of separate FCP channels, switch zoning, and LUN masking

WWN1 = World Wide Port Name 1 = LUN1 and LUN3
WWN2 = World Wide Port Name 2 = LUN2 and LUN4



IBM Coupling Facility Control Code

Dynamic ICF Expansion into shared ICFs
Structure rebuild

Level 8 (6/99)
G3/4 Last Level

Shared ICF
IBM Concurrent Conditioning for S/390

Level 7 (3/99)

CF Resource Mgmt
- IC Channel
- ICB
- TPF support

Level 6

- DB2 CACHE Duplexing
- Dynamic ICF Expansion

DB2 Availability

Level 5 - G2 Last Level

MQSeries V2.2
Multi-system Enclaves
IRD Support

Level 9 (6/00)

z900 GA
G5/6 Dr 26

Initial CF HW and Functions

Level 0

Level 1

- Dynamic Alter
- CICS TSQ
- System Logger

Continuous Operations

- Dynamic CF Dispatch
- ICF Support

Level 4 - G1 Last Level

Dynamic I/O Peer and Receiver Structure Changes Sender in CF

Level 10 (10/01)

z900 Dr 3C

List KEY and List cursor Enhancements

Level 1

IMS Shared Message Queue Base

Level 3

Level 2

64-bit
48 internal tasks
CF Duplexing

Level 12 (3/02)

z800 GA
z900 Dr 3G

Register Name List and Batch Unlock

Level 2
DB2 Performance

Structure Enhancements (255/Structure & 1023 Structure/CF

CF Duplexing

Level 11 (4Q/02)

G5/6

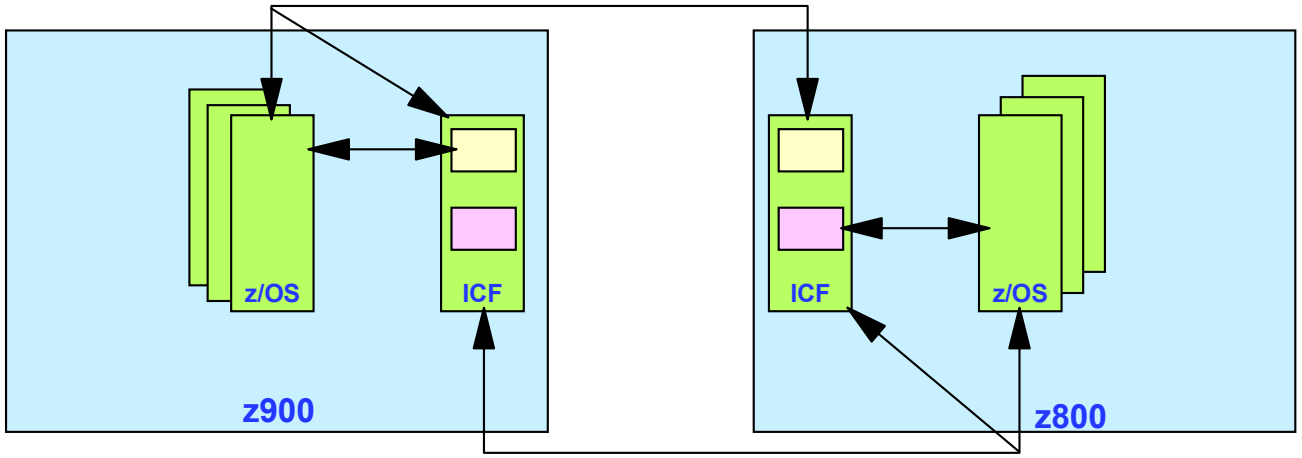
Cache Vector index Comparison Enhancement

VSAM RLS
Level 2

Systems Management

World's Leading and Unrivaled Coupling Technology

System Managed CF Structure Duplexing



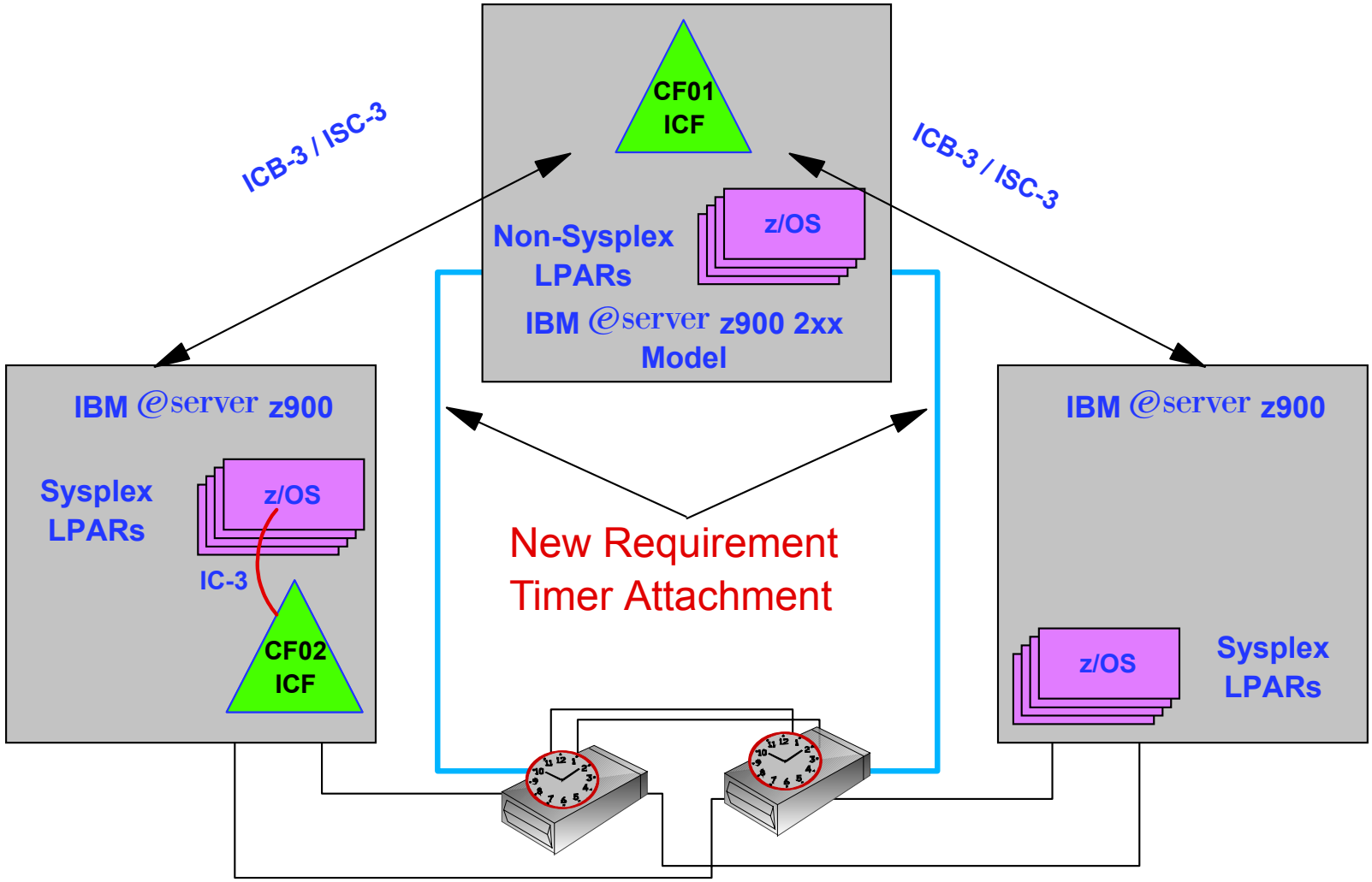
Robust, standard recovery capability
- Ease of use for middleware and ISVs
- Eliminates rebuild delay
- Reduces need for standalone CF

Considerations
- Not all structures supported
- Overhead to duplex
- Storage required to duplex
- See Announcement 102-181 (06/02)

Requirements:
- Sysplex Images -
  - **z/OS v1.2 & up**
  - **APAR OW41617**
- Coupling Facilities -
  - ICF or standalone
  - **zSeries with CFCC Level 12**
  - **S/390 G5/6 with CFCC Level 11**
  - **CF - CF link connectivity**

Message Time Ordering - Fast CF Connection to Sysplex Timer

z900 New 2xx Model with ICF and non-Parallel Sysplex LPARS



- **Required z/OS and OS/390 New Function APAR - OW53831**

LPAR Support for CUoD and CBU zSeries and G5/6

Nondisruptive CIU/CBU/MES Upgrade

- Add physical CPs to shared CP Pool
- Add physical ICFs to shared ICF Pool (Not supported for CBU))
- Added physical engines become available and visible (STSI and event)

LP Profile: Initial and Reserved Processors (G5/6 Dr 26 6/2000)

- Support for ESA/390, LINUX (CP or IFL), and CF (ICF and/or CP)
- Initial - On at Activation
- Reserved - Can be configured on if resource is available
- Initial + Reserved can equal maximum engines possible for the MCM

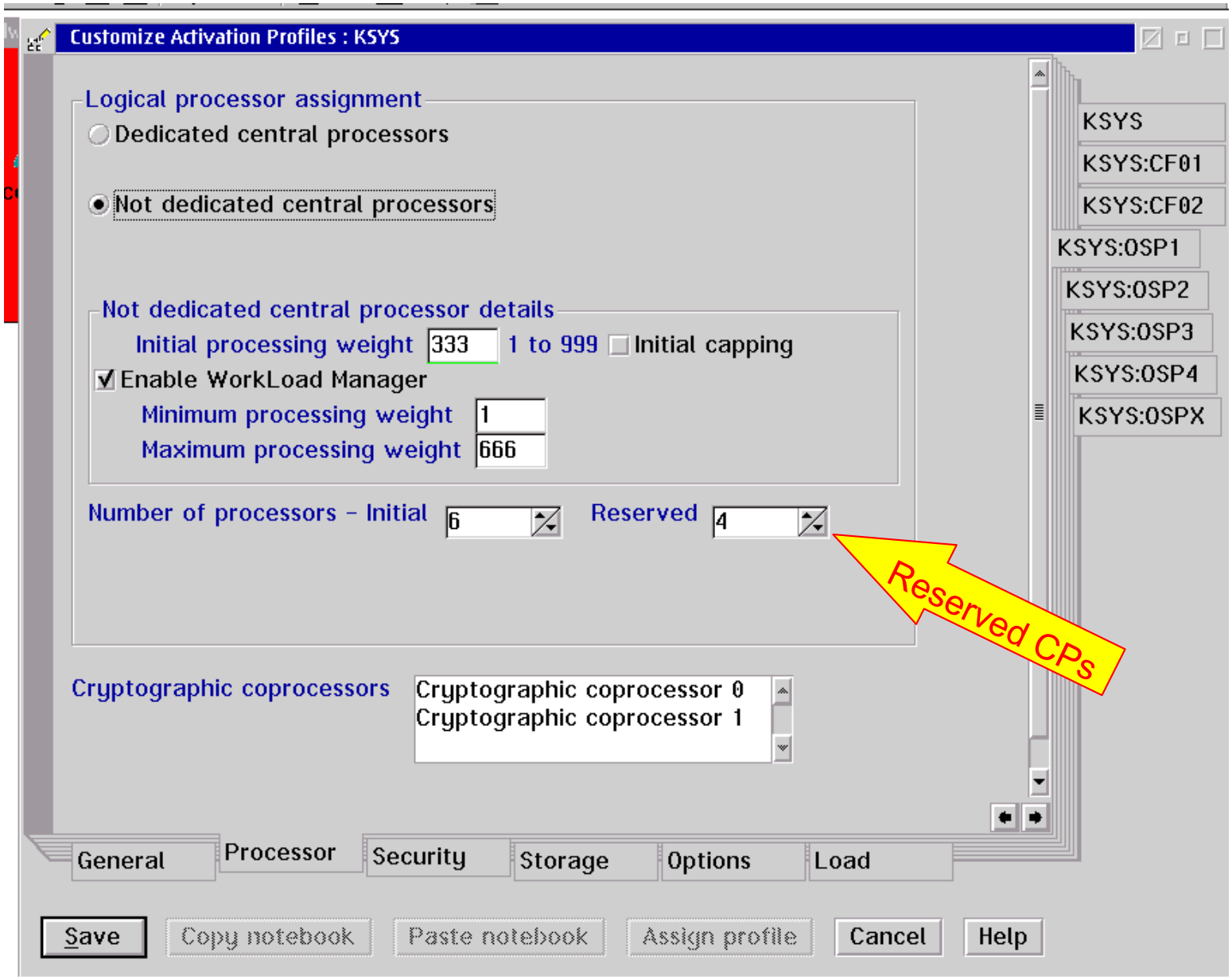
Nondisruptive CBU CP Downgrade (z900 Dr 3C and z800)

LP Activation and Processor Config Rules

- Can't activate LP with more initial shared logicals than shared pool contains.
- Can't activate LP with initial dedicated logicals if physicals taken from shared pool would "starve" an active LP with shared logicals.
- z900 nondisruptive CBU CP downgrade takes CPs ONLY from the shared CP pool, will not take the last CP.

Initial and Reserved CP Support

zSeries at GA
S/390 G5 or G6 (Dr 26w)



zSeries Storage Granularity

zSeries and G5/6 (Dr 22e and later)

- Single Storage Pool - All central storage
- ES configured as needed from CS - No POR needed

Earlier Machines

- POR required to change CS/ES split

Granularity

- Note: LPAR Mode ONLY, 1 MB in BASIC mode for HSA
- Was 1 MB prior to G3 Dr 88 for LPAR, too

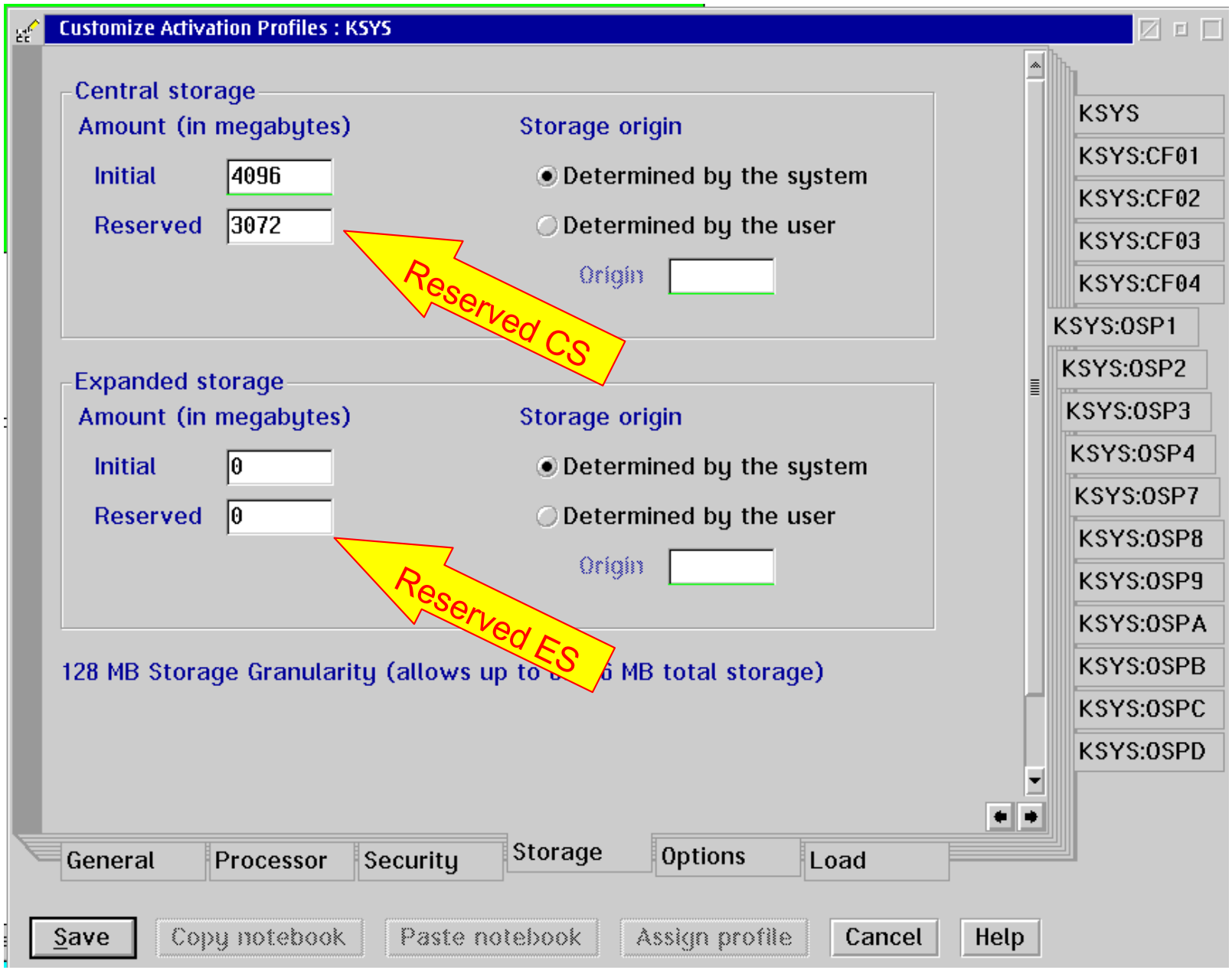
Total Storage G5/6 or zSeries	Granularity CS & ES
5 - 8 GB	16 MB
10 - 16 GB	32 MB
18 - 32 GB	64 MB
40 - 64 GB	128 MB

RSU Increment Size!!!

Initial and Reserved Storage

Expanded Storage

- Supported by z/Architecture
- Not used - zSeries CFCC, 64-bit OS/390, z/OS, or z/OS.e



z900 Concurrent Memory Upgrade

Concurrent Memory Upgrade

- LIC CC memory activation
- From current size to any size supported by cards installed (card change is disruptive)
- Add to partition using DSR/2 (MVS only) - Central or Expanded

Prerequisites

- z900 Server, LPAR mode
- Must have spare memory capacity on installed cards, otherwise disruptive
- Must predefine additional memory to partition as "Reserved Storage"

Note: No CBU for memory

Memory Card Size and Number

Storage	Models 100-109	Models 110-116 210-216	Models 1C1-1C9 2C1-2C9
5 6 7 8	4 GB x 2	Not Offered	Not Offered
10 12 14 16	8 GB x 2	4 GB x 4	4 GB x 4
18 20 24 28 32	16 GB x 2	8 GB x 4	8 GB x 4
40 48 56 64	Not Offered	16 GB x 4	16 GB x 4

MVS RSU Parameter

In IEASYSxx. Specifies the number of central storage increments to be made available for central storage reconfiguration

- MVS attempts to keep this area free of long term fixed pages

$$RSU = \frac{\text{CS amount to be reconfigured}}{\text{storage increment size}}$$

Or: Storage to be kept free = RSU * increment

- If memory is upgraded, check the RSU parameter!

OS/390 V2.10 and z/OS - Better RSU Options

- All OFFLINE storage (Reserved Storage)
- An amount (%, MB or GB) - System calculates increments