



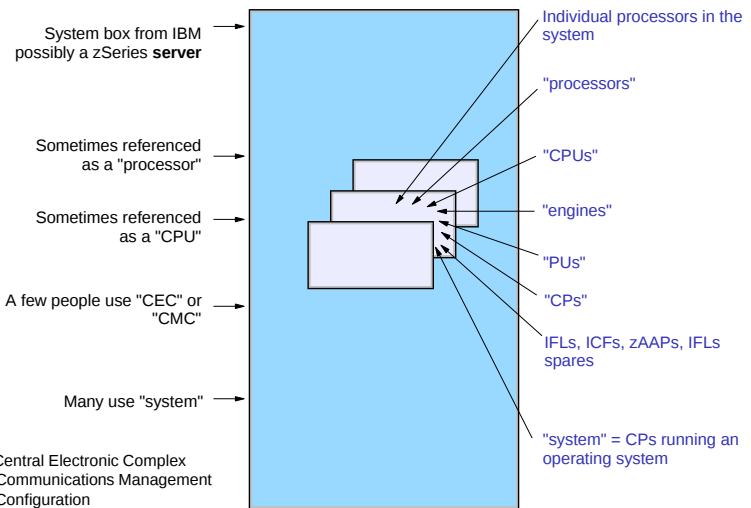
Arquitectura mainframes familia Z

Roberto Gómez Cárdenas
rogomez@itesm.mx

Lámina 1 Roberto Gómez C.



Terminología confusa



The diagram illustrates the Central Processor Complex (CPC) as a large blue rectangle. Inside, there are several smaller, overlapping light blue rectangles representing individual processors. Arrows point from various labels to these components:

- System box from IBM possibly a zSeries **server** (points to the entire CPC box)
- Sometimes referenced as a "processor" (points to one of the inner rectangles)
- Sometimes referenced as a "CPU" (points to one of the inner rectangles)
- A few people use "CEC" or "CMC" (points to one of the inner rectangles)
- Many use "system" (points to the entire CPC box)
- Individual processors in the system (points to one of the inner rectangles)
- "processors" (points to one of the inner rectangles)
- "CPUs" (points to one of the inner rectangles)
- "engines" (points to one of the inner rectangles)
- "PUs" (points to one of the inner rectangles)
- "CPs" (points to one of the inner rectangles)
- IFLs, ICFs, zAAPs, IFLs spares (points to one of the inner rectangles)
- "system" = CPs running an operating system (points to the entire CPC box)

CEC: Central Electronic Complex
CMC: Communications Management Configuration

CPC: Central Processor Complex

Lámina 2 Roberto Gómez C.

 **Viendo el sistema**

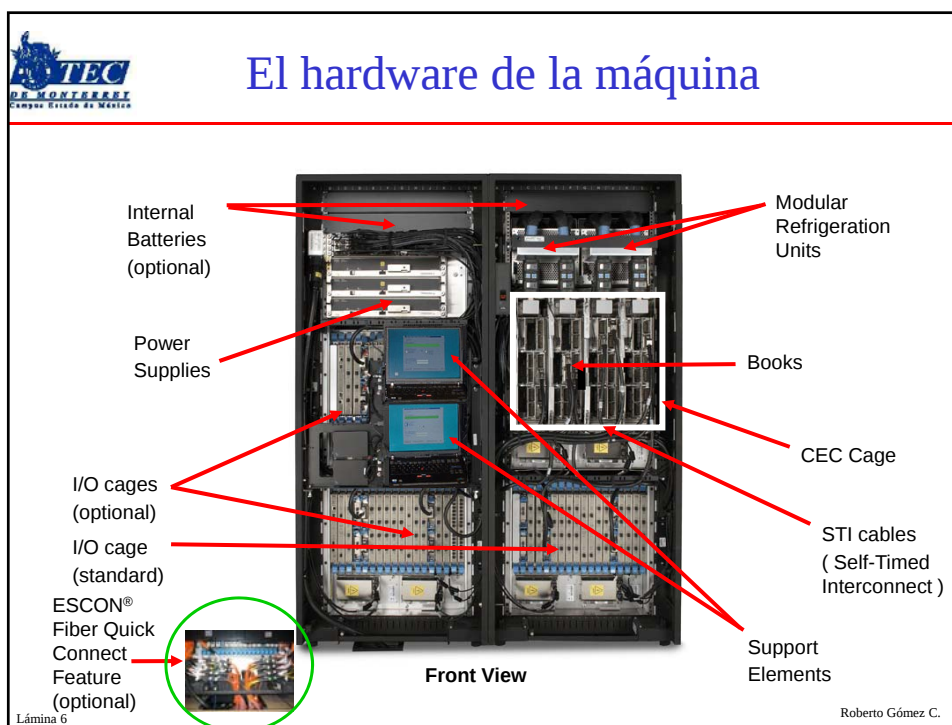


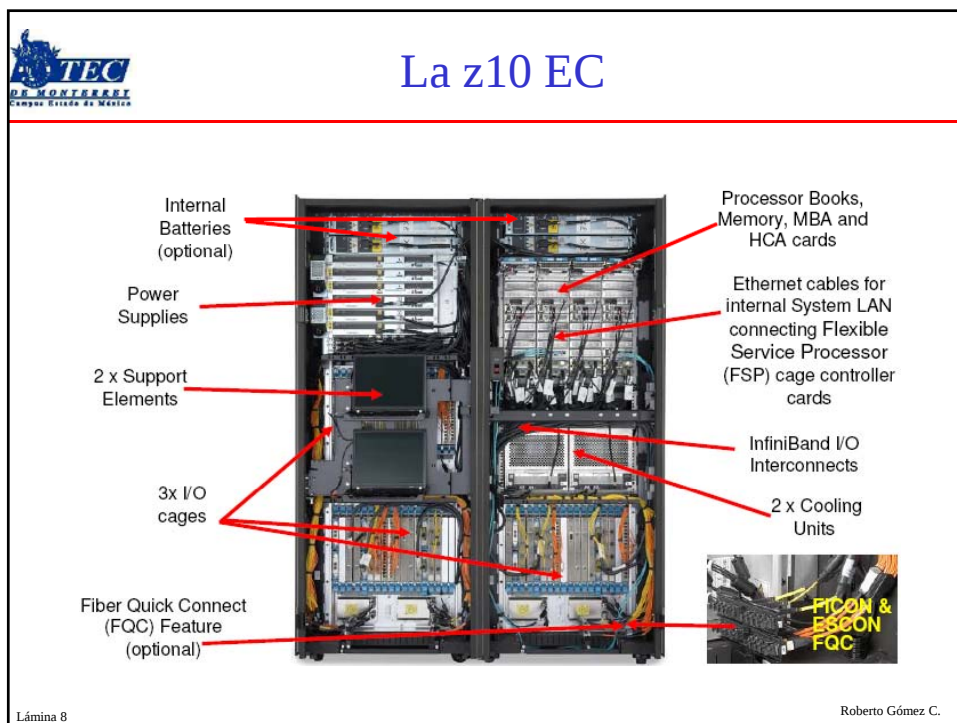
Lámina 3 Roberto Gómez C.

 **Veamos primero el hardware**

componentes zSeries

Lámina 4 Roberto Gómez C.





z9-109 Books and CEC Cage

- La “jaula” acepta entre uno y cuatro “libros de procesadores
- Standard Models
 - Model S08 - Book 0
 - Model S18 - Books 0 and 1
 - Model S28 - Books 0, 1, and 2
 - Model S38 - Books 0, 1, 2, and 3
- Enhanced capacity model
 - Model S54 - Books 0, 1, 2, and 3

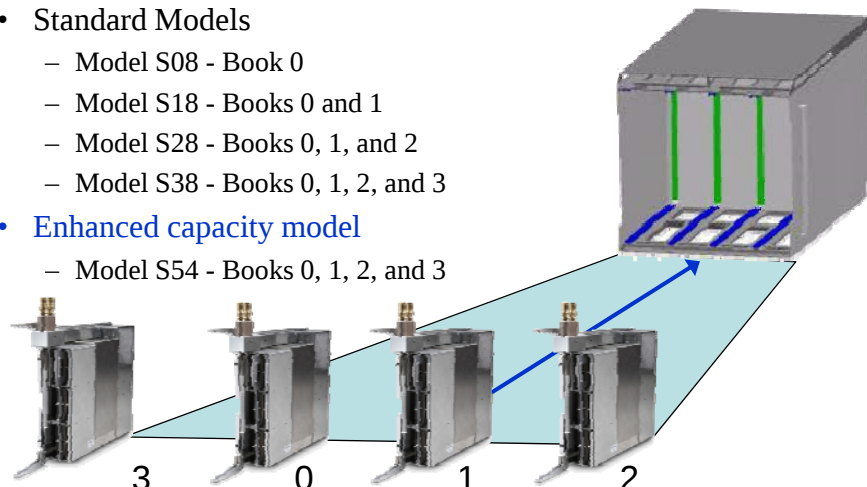


Lámina 9

Roberto Gómez C.

z9-109 Processor Book Layout

Up to 8 Hot pluggable MBA/STI fanout cards

MCM

Memory Cards Up to 128 GB

Front View

Side View

Note: 1. Concept Illustration only - not to scale
2. 4 or 8 pluggable Memory Cards
3. Each MBA fanout card is hot-pluggable and has 2 STIs

Concurrent upgrade repair for I/O including MBA fan out cards and memory

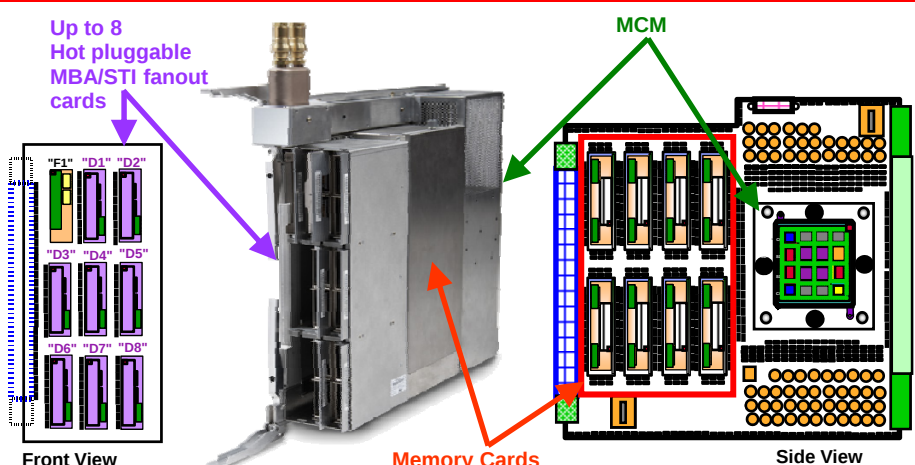
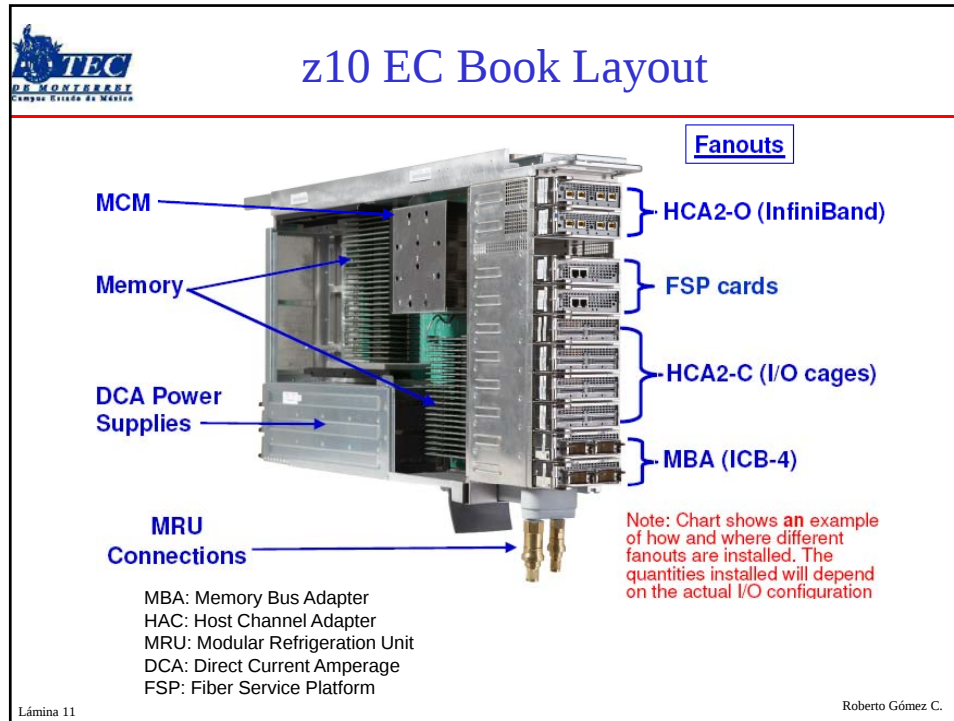


Lámina 10

Roberto Gómez C.





Tecnología CMOS 10K chip

- Chips PU, SC, SD y MSC
- Interconexiones de cobre, 10 niveles de cobre
- 8 chips PU/MCM
- 4 chips System Data (SD) cache
- 1 chip Storage Control (SC)
- 2 chips Memory Storage Control (MSC)
- 1 chip de reloj

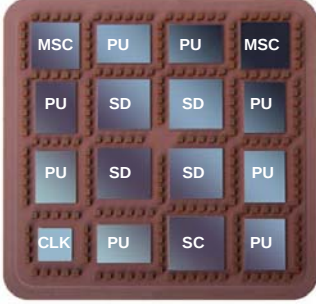


Lámina 13
Roberto Gómez C.



Unidades de procesamiento

- Procesador ofrece varios tipos de procesadores,
- La mayor parte de tipos adicionales no cuentan como procesadores completos para software.
 - Central Processor (CP)
 - procesador z/OS completo
 - System Assistance Processor (SAP)
 - usado para el subsistema de I/O
 - cada máquina cuenta con al menos uno
 - Integrated Facility for Linux (IFL)
 - procesador especial para Linux - opcional
 - zAAP –
 - usado para código Java – opcional
 - zIIP
 - usado para procesamiento DB2 – opcional
 - Integrated Coupling Facility (ICF)
 - usado para coupling facilities
 - Spares
 - PU no caracterizado funciona como de repuesto

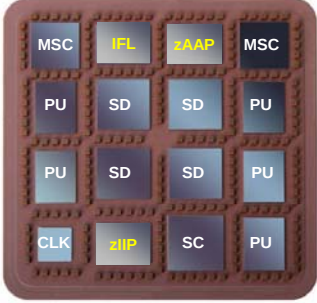


Lámina 14
Roberto Gómez C.

System z – Processor RAS Design

- Processor
 - Doble ejecución con comparación
 - Enhanced application preservation
 - Transparent processor sparing
- Unique Cache Design
 - L1 cache store through
 - Shared L2 cache allows most hard failures to be recovered

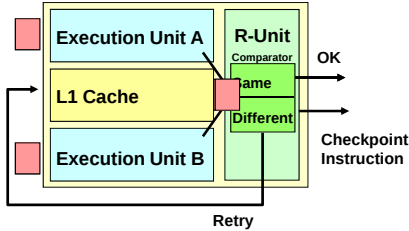
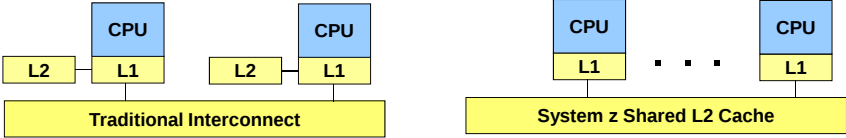



Lámina 15

Roberto Gómez C.

Ahora si, la arquitectura

componentes

Lámina 16

Roberto Gómez C.



Características del sistema 360

- La caja de procesamiento central contenía los procesadores, la memoria, los circuitos de control y los canales con las interfaces.
- Canales proporcionan un path de datos y control independiente entre dispositivos E/S y la memoria.
 - Cada sistema podía contener hasta 16 canales.
 - Sistemas modernos pueden tener hasta 1024.
- Canales conectados a las unidades de control.
- Unidades de control se conectan a dispositivos como drivers de disco, drivers de cinta e interfaces de comunicaciones.

Lámina 17

Roberto Gómez C.



Esquema sistema S/360

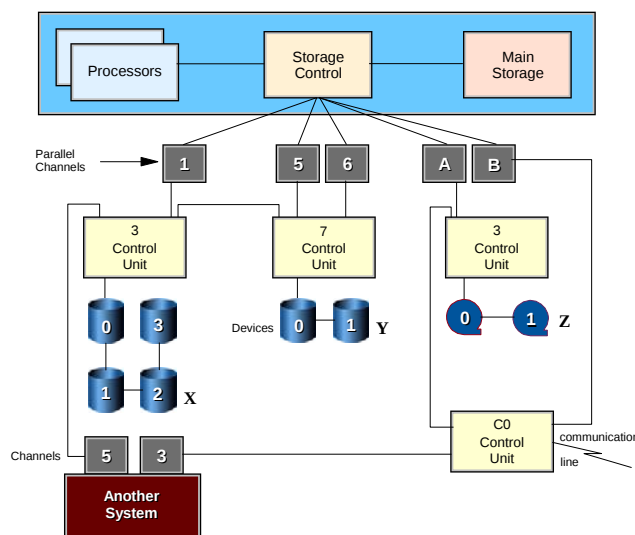


Lámina 18

Sistemas actuales no se encuentran conectados de esta forma.

Roberto Gómez C.



Características canales

- Un canal paralelo se puede conectar aun máximo de 8 unidades.
- La mayor parte de unidades de control se pueden conectar a varios dispositivos.
 - El máximo depende de la UC.
 - 16 es un número típico.

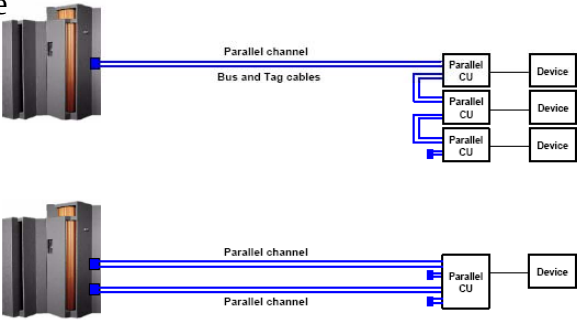


Lámina 19



Los canales y la direcciones de dispositivos

- Números hexadecimales.
- En los diseño anteriores las direcciones de los dispositivos estaban físicamente relacionados con el hardware de la arquitectura.
- Conocidos como canales paralelos
 - canales de diámetro grueso y cobre pesado

address: 1 3 2

channel number

↙

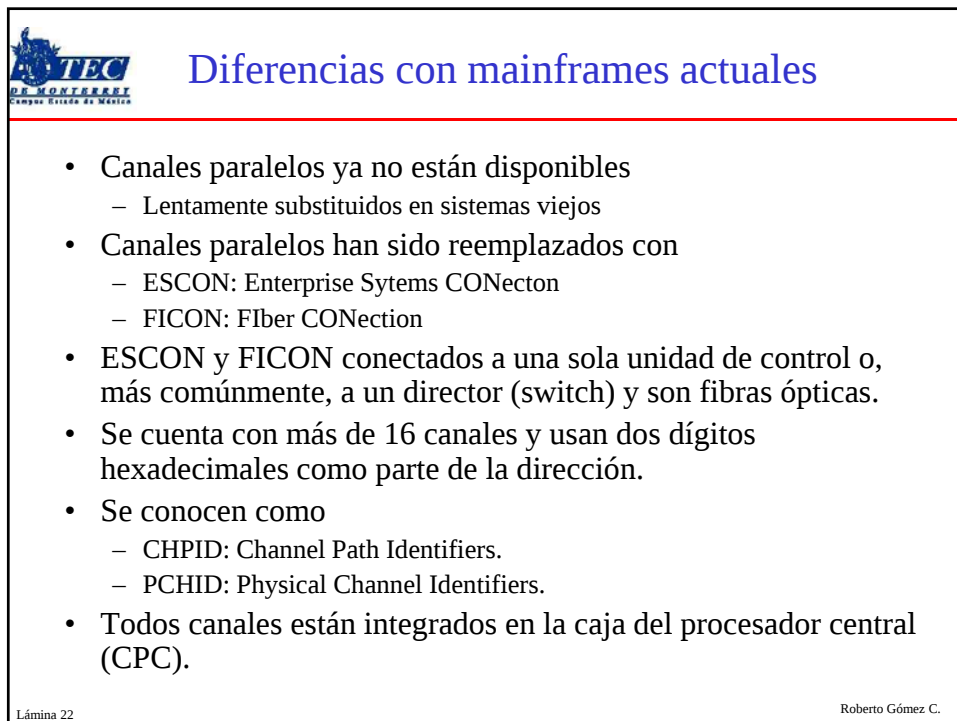
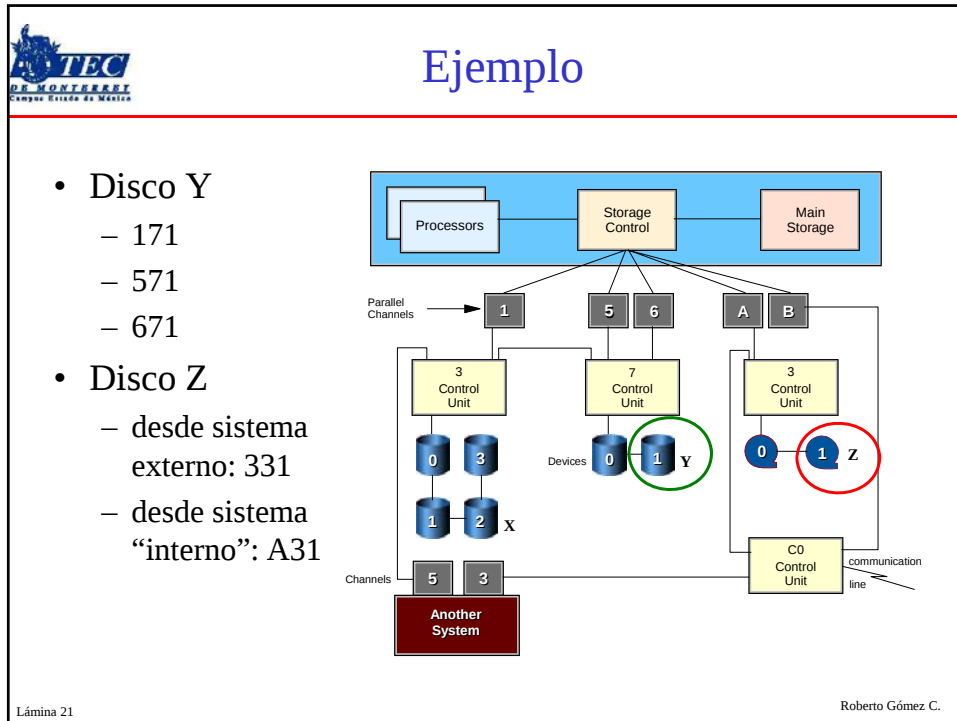
control unit number

↑

device number

↘

Lámina 20
Roberto Gómez C.



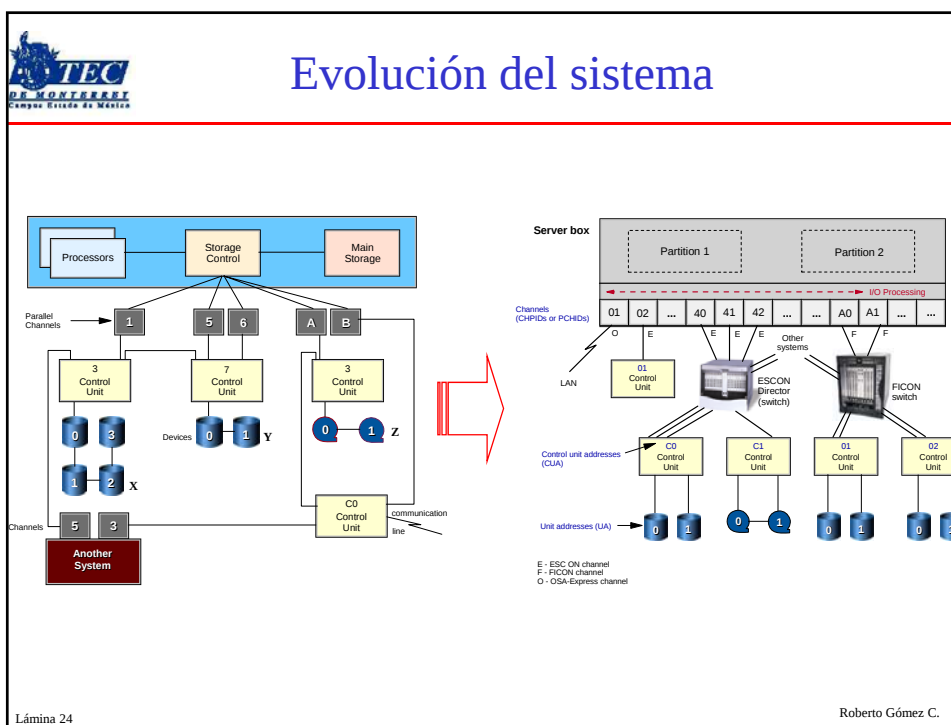


Diseños actuales

- Diseños CPC actuales más complejos que los antiguos diseños.
- Esta complejidad incluye:
 - Conectividad E/S y configuración.
 - Operación E/S .
 - Particionamiento del sistema.

Lámina 23

Roberto Gómez C.





ESCON FICON directors



Lámina 25

Roberto Gómez C.



Conectividad E/S (I/O)

- Basado en canales ESCON y FICON.
- ESCON y FICON se conectan a un solo puerto o un puerto de un switch.
- Mainframes modernos usan switches entre canales y unidades de control.
- Direcciones CHPID cuentan con dos dígitos.
 - Direcciones: 00 a FF (256 direcciones).
- Varias particiones pueden compartir CHPIDs.
 - Depende naturaleza unidades de control .
 - En general, CHPIDs de discos se comparten.
- Nivel de subsistema de I/O existe entre sistemas operativos en particiones.

Lámina 26

Roberto Gómez C.



ESCON - FICON

- Director ESCON y FICON switch.
- Dispositivo sofisticado que puede manejar altos rangos a través de varias conexiones.
 - Un director grande puede contar con 200 conexiones.
- Director o switch da seguimiento de que CHPID inicio, así como la operación de I/O que comenzó, de tal forma que los datos y status sean regresados al lugar correcto.

Lámina 27

Roberto Gómez C.



Conectividad ESCON (antes)

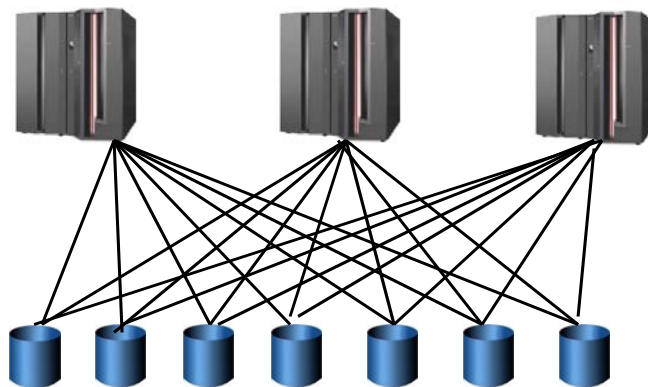
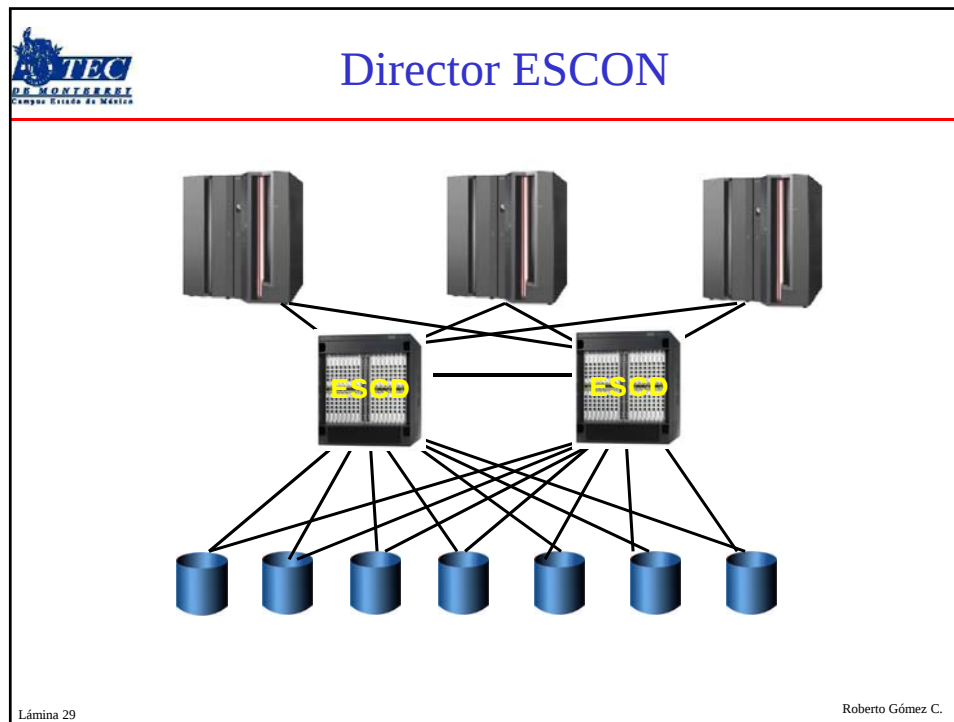


Lámina 28

Roberto Gómez C.

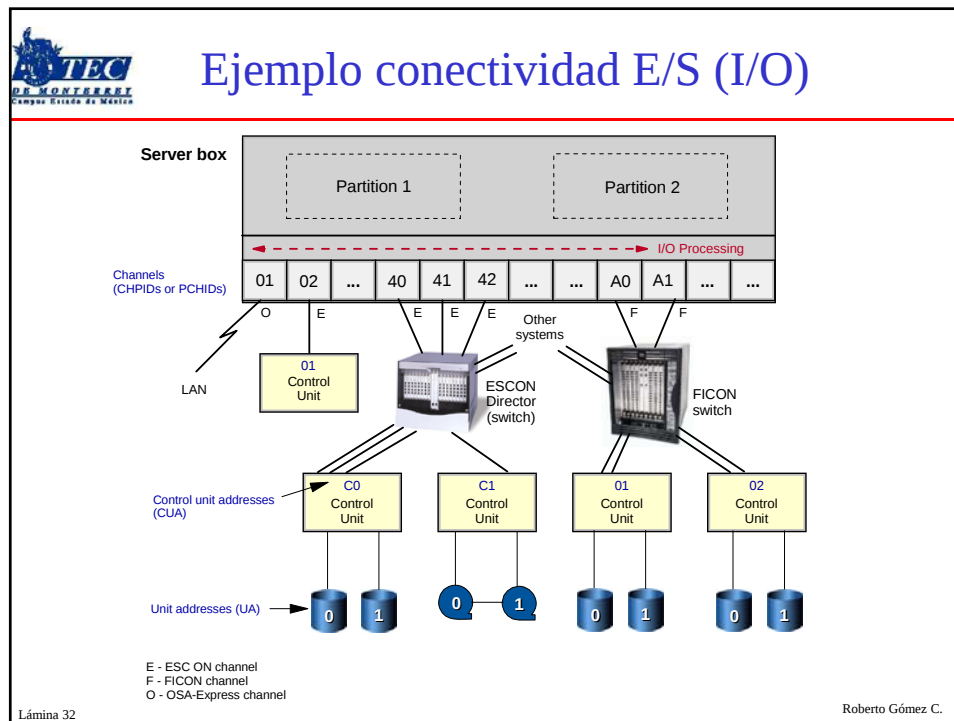
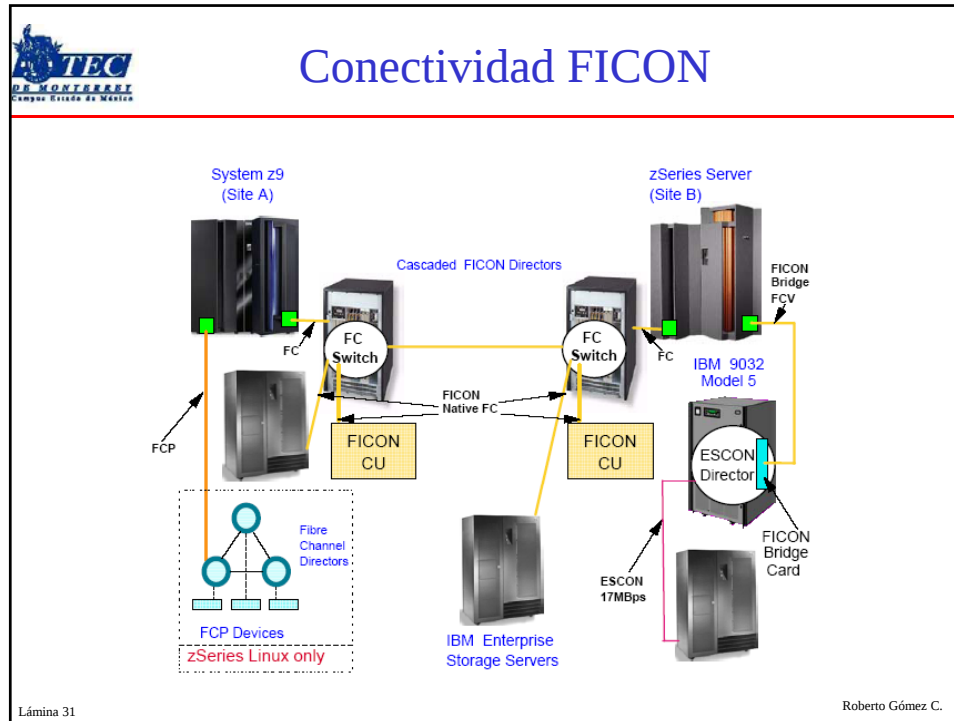


ESCON vs FICON

- ESCON
 - 20 Mbytes/segundo.
 - Lotes de “tiempo muerto”.
 - Una petición actividad .
 - Una unidad de control.
- FICON
 - 400 Mbytes/segundo.
 - Usa estándar FCP
 - Fiber Channel Protocol
 - Proporciona conexión entre dispositivos SCSI e imágenes del S.O.
 - Cable fibra óptica (menos espacio bajo suelo).
 - Actualmente, hasta 64 paquetes de E/S simultáneos en un tiempo con 64 unidades de control diferentes.
 - Soporta switches en cascada.

Lámina 30

Roberto Gómez C.





Archivo IOCDS

- IOCDS = I/O Control Data Set.
- Archivo de control usado para las operaciones de E/S.
- Traduce direcciones físicas E/S (números CHPID, números puertos switch, direcciones unidad de control y direcciones de unidades) en números de dispositivos usados por el software del sistema operativo para acceder a los dispositivos.
- Esto se carga en el HSA en el encendido y puede ser modificado dinámicamente.
 - HSA: Hardware Save Area
 - Subsistema de canales lógicos y conjuntos de subcanales
 - La memoria del HSA esta aislada de la memoria del resto

Lámina 33

Roberto Gómez C.

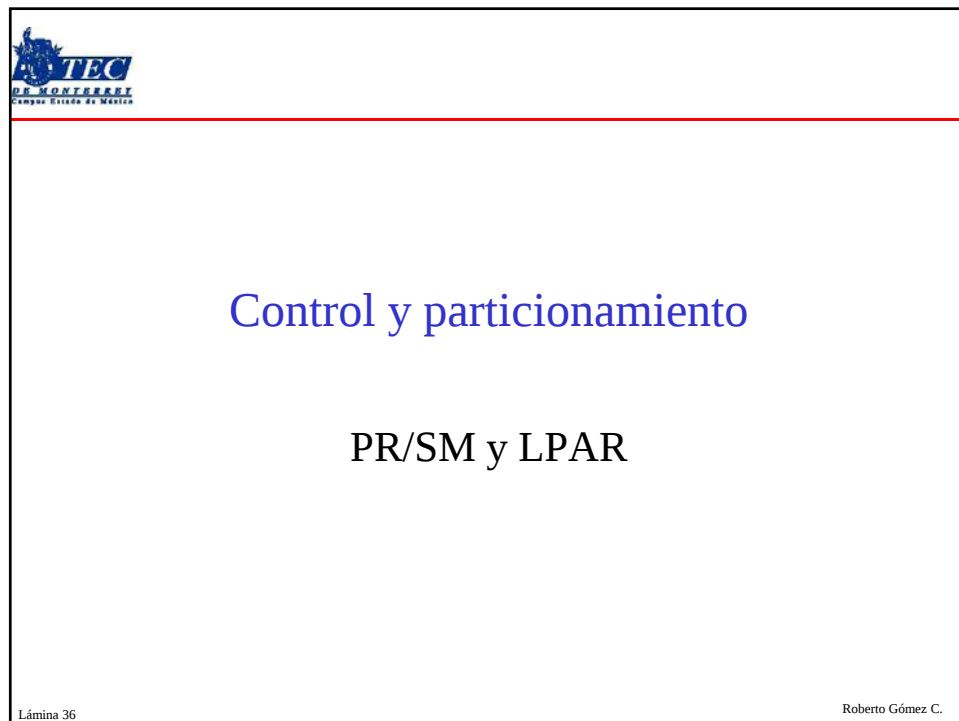
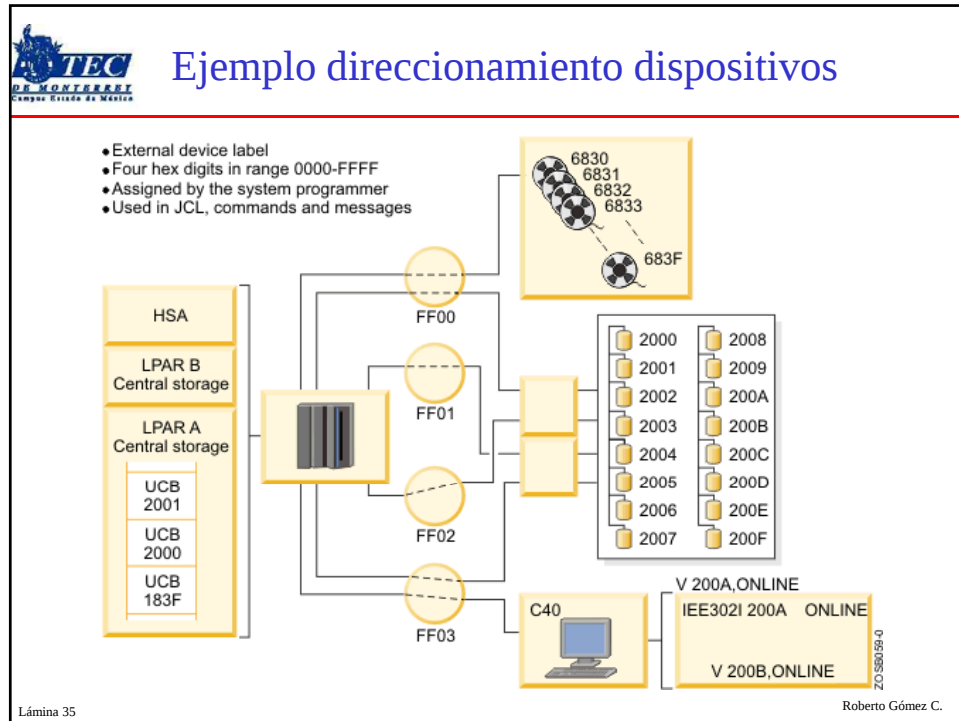


Números dispositivos

- Números dispositivos son asignados por el programador del sistema cuando se crea el IODF y el IOCDS y son arbitrarios.
 - Pero no aleatorios.
 - IODF: Input/Output Definition File.
- En máquinas modernas existen tres o cuatro dígitos hexadecimales.
 - FFFF = pueden definirse 64K dispositivos

Lámina 34

Roberto Gómez C.





Control sistema y particionamiento

- Entre las funciones del control de sistema esta la habilidad de particionar el sistema en particiones lógicas.
- Inicialmente limitado a 15 LPARS.
 - Nuevas máquinas permiten 60 particiones.
- Limitaciones prácticas de tamaño memoria, disponibilidad E/S y disponibilidad de energía eléctrica limitan el numero de LPARs a menos de estos límites.

Lámina 37

Roberto Gómez C.



Las particiones lógicas

- LPAR
 - Subconjunto del hardware del CPC, definido para soportar un sistema operativo.
- Un LPAR cuenta con recursos
 - Procesadores.
 - Memoria.
 - Dispositivos E/S.
- Varios particiones lógicas pueden existir dentro del hardware del CPC.

Lámina 38

Roberto Gómez C.



PR/SM y LPARS

- PR/SM: Processor Resource/Systems Manager
 - Hardware y firmware que proporciona el particionamiento.
- Son la funciones del PR/SM las que son usadas para crear y correr LPARs.
- La diferencia es
 - PR/SM: facilidades implementadas dentro del mainframe.
 - LPAR: el resultado de usar PR/SM.
- Esta diferencia es ignorada frecuentemente y el termino LPAR es usado frecuentemente para referirse a las facilidades y a sus resultados.

Lámina 39

Roberto Gómez C.



Características LPARS

- Un LPAR es el equivalente a un mainframe separado para propósitos prácticos.
 - Cada LPAR es independiente.
- Cada LPAR corre su propio sistema operativo.
- Dispositivos pueden compartirse a través distintos LPARs.
- Procesadores pueden ser dedicados o compartidos.
- La memoria no puede ser compartida entre LPARs.

Lámina 40

Roberto Gómez C.



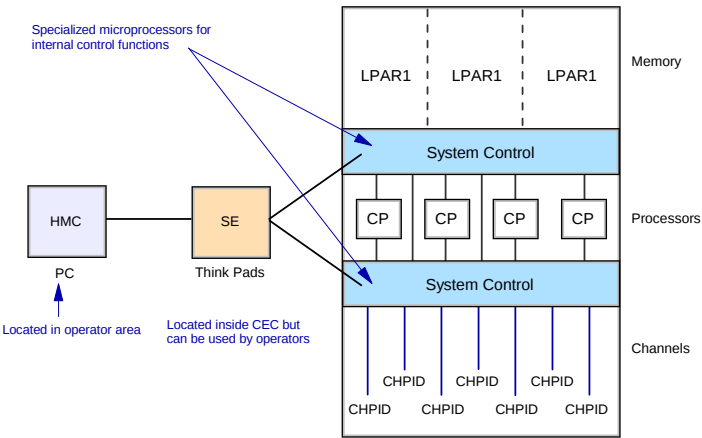
Definiendo una LPAR

- Administrador de sistema asigna
 - Memoria.
 - Procesadores.
 - CHPIDs ya sean dedicados o compartidos.
- Las especificaciones de particionamiento se almacenan, parte en el IOCDS y parte en un profile del sistema
 - Definido usando uno de los Support Elements.
 - Generalmente esto es actualizado a través del HMC.
 - HMC: Hardware Management Console.
- Cambiar el profile del sistema y el IOCDS usualmente requiere un power-on reset (POR) pero algunos cambios son dinámicos.

Lámina 41
Roberto Gómez C.



Control sistema y particionamiento

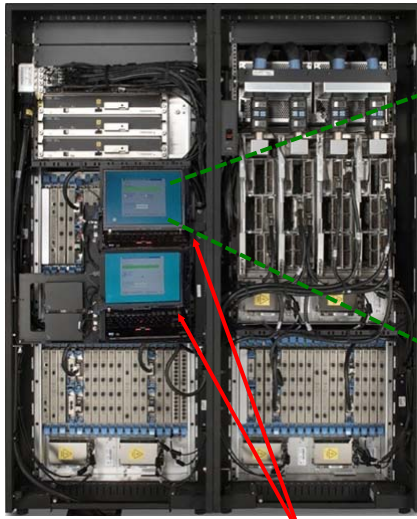


The diagram illustrates the control and partitioning architecture of a z/OS system. On the left, a PC (located in the operator area) connects to an HMC (Hardware Management Console). The HMC connects to an SE (Support Element, located inside the CEC but usable by operators). The SE then connects to the System Control blocks within the LPARs. The LPARs are shown as vertical columns, each containing a System Control block, four CP (Central Processor) blocks, and four CHPID (Channel Program Identifier) blocks. The CP blocks are labeled 'CP' and the CHPID blocks are labeled 'CHPID'. The LPARs are connected to the Channels, which are represented by vertical lines at the bottom of the diagram.

CP: Procesadores del CPC que se pueden compartir
 HMC: Hardware Management Console
 SE: Support Element, Notebook/Laptop dentro del CPC
 System Control: conjunto de controladores internos: microprocesadores de organización simple y data set más simple que los procesadores zSeries conocidos como controladores para evitar confusión con procesadores zSeries

Lámina 42
Roberto Gómez C.

El SE y el HMC

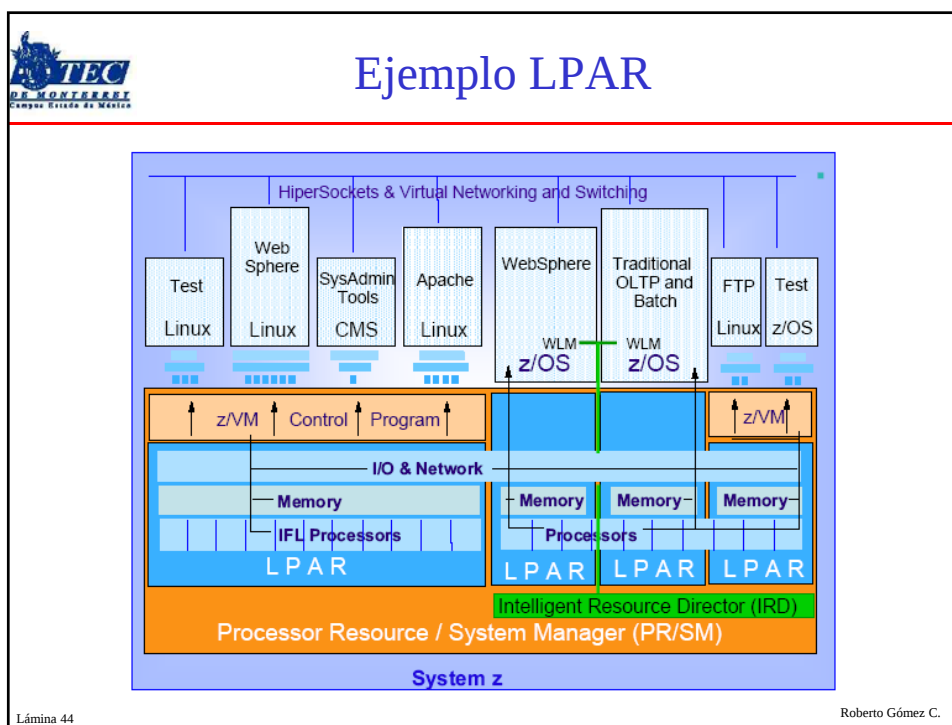


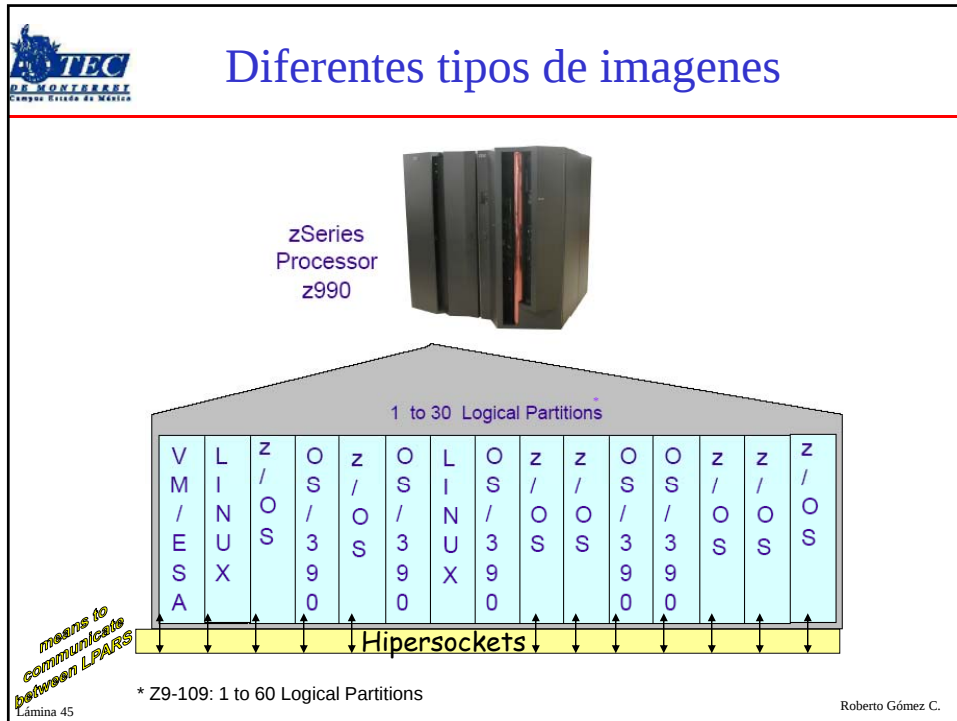
Support Elements

Hardware Management Console

Lámina 43

Roberto Gómez C.







¿Porqué un HMC?

- Servidores se están volviendo más virtualizados
 - Sistemas Operativos contarán con menos acceso directo y control sobre el hardware de control real.
 - Lugar para almacenar aplicaciones de manejo avanzado de toda la plataforma
 - Configuración servidor antes de instalar el sistema operativo.
 - Servicio cuando el sistema operativo no se encuentra disponible.
 - Coordinación de interacción a través de diferentes plataformas.
 - Estas funciones deben contar con un interfaces comunes independientemente del sistema operativo.
 - Soporte de operaciones locales y remotas.

Lámina 47

Roberto Gómez C.



Opciones

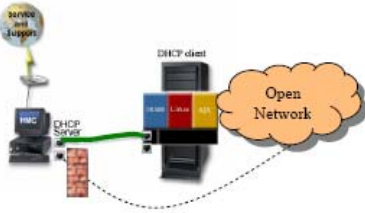
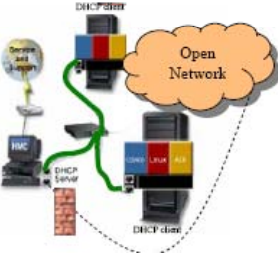
- Local HMC local
 - Conectado directamente al sistema, administrando vía una red privada.
 - Red privada = HMCs y sistemas administrados
- HMC remoto
 - Usado para acceder otro HMC o sistema de administración.
 - Presentes en una red abierta.
- Web-based System Manager Remote Client
 - PC que cuenta con el software Web-based System Manager
 - Utilizados para acceder otros HMCs de forma remota.

Lámina 48

Roberto Gómez C.



Abierto vs cerrado

- Red privada
 - HMC = Servidor DHCP
 - Sistemas Administrados = Clientes DHCP
- Red abierta
 - HMC = Cliente DHCP o IP estática

Lámina 49

Roberto Gómez C.



Redundancia del HMC (1)

- Dos HMCs con un solo hub
 - HMC-A
 - Configurado como un servidor DHCP
 - HMC-B
 - Configurado como un cliente DHCP
 - Sistemas Administrados
 - Configurados como clientes DHCP
- En el HMC-B
 - Usa “Add Managed System”
 - Para introducir el sistema administrado al HMC

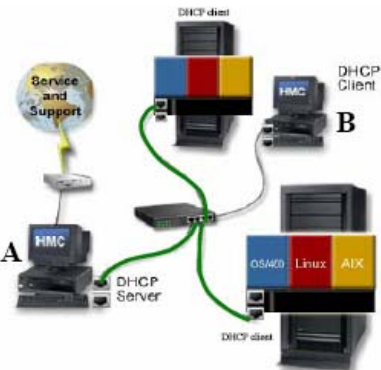


Lámina 50

Roberto Gómez C.



Redundancia del HMC (2)

- Cada HMC con su propio hub
 - HMC: Servidor DHCP
 - Sistemas Administrados: Clientes DHCP
- Cableado
 - Líneas solidas
 - Red privada 1
 - Líneas punteadas
 - Rede privada 2

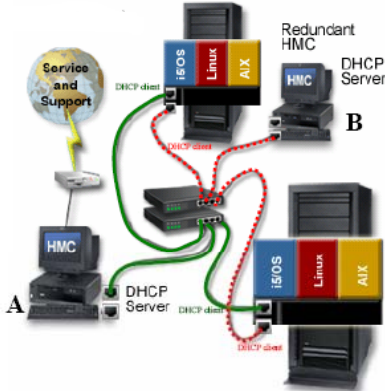


Lámina 51
Roberto Gómez C.



Interfaces usuarios y acceso HMC

- GUI local
 - Lanzado automáticamente al arranque del HMC.
- GUI remoto
 - Cliente WebSM sobre Windows o Linux.
 - Bajado directamente del HMC usando un WebBrowser.
- Línea comandos local
 - Restringida a un conjunto de comandos soportados por el HMC.
- Línea comandos remota
 - Acceso vía SSH.
 - Posible definir archivos con llaves para autenticación para evitar prompts de passwords.

Lámina 52
Roberto Gómez C.



Seguridad en el HMC

- Shell restrictivo
 - Proporciona acceso a los comandos y/o funciones soportados por el HMC
 - Accesible remotamente por un cliente SSH.
 - También accesible via un prompt de comandos en el mismo HMC
- Control de acceso usuario granular
 - Definir tareas y roles de recursos que definen listas de tareas de usuarios y recursos
 - Asigna roles a usuarios para definir su derechos de acceso
 - Por ejemplo: acceso limitado a una sola partición
- Controles seguridad firewall red
 - Definir que servicios de red deben ser accesibles en las diferentes interfaces físicas de red
 - Por ejemplo, limitar WebSM o SSH a una sola interfaz o a ninguna

Lámina 53

Roberto Gómez C.



Ejemplo del Hardware Management Console (HMC)

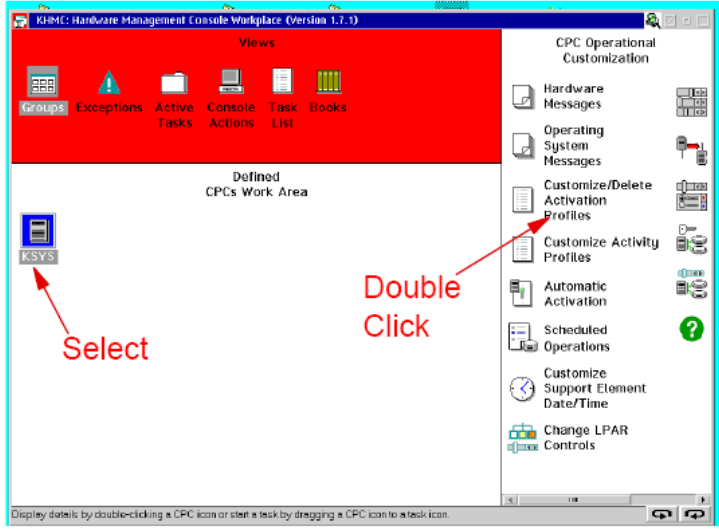


Lámina 54

Roberto Gómez C.

El Hardware Management Console (HMC)

Profile name: ZSYSRESET0308 Assigned for activation

Description: This is the default Reset profile.

Input/Output Configuration Data Set	Type	Allow Dynamic I/O	Partitions
A0 03.27.01	Partition	Yes	OSP1 OSP2
A1 03.13.01	Partition	Yes	OSP1 OSP2
A2 03.26.01	Partition	Yes	OSP1 OSP2
A3 03.27.01	Partition	Yes	OSP1 OSP2
D0 DIAGNOSE	ESA/390	No	OSP1 OSP2

Use Active IOCDS: Currently A0

Mode: Logically partitioned
ESA/390
ESA/390 TPF

Load delay for power sequencing: 0 minutes 0 seconds

Storage assignment

Allocate PUS

OS

i.e assign a profile to a Linux partition

General Storage Dynamic Options CP/SAP Partitions

Save Copy notebook Paste notebook Assign profile Cancel Help

Lámina 55

Roberto Gómez C.

Opciones imagenes

Image options

Minimum input/output (I/O) priority: 10

Maximum input/output (I/O) priority: 3

Defined capacity: 75

CP management cluster name: TESTPLEX

MSUs - WLC

General Processor Security Storage Options Load

Save Copy notebook Paste notebook Assign profile Cancel Help

Lámina 56

Roberto Gómez C.



Opciones particiones

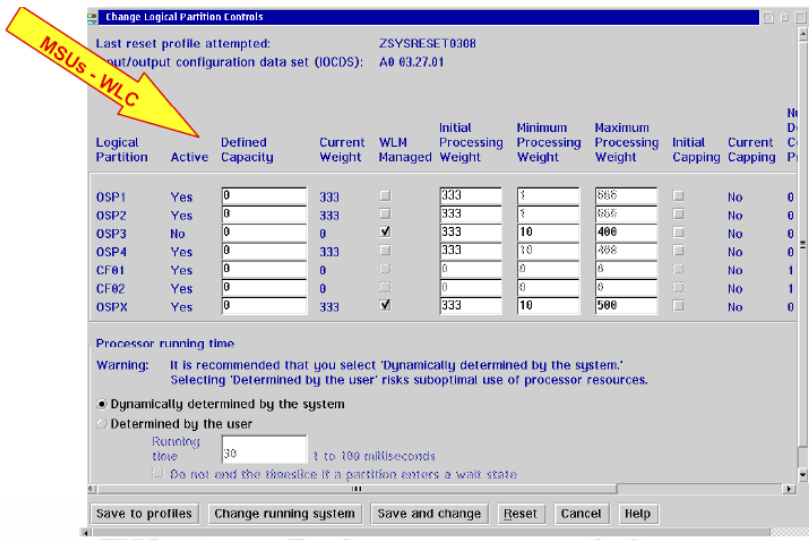


Lámina 57
Roberto Gómez C.



Unidades de procesamiento

- Existen diferentes tipos de procesadores en un sistema que puede ser usado para diferentes propósitos.
 - Varios de estos propósitos están relacionados a control del costo de software, mientras que otros son más fundamentales.
- Todos empiezan como unidades de procesamiento (PUs)
 - PU: procesador que no ha sido caracterizado para un uso.
- Cada uno de los procesadores empieza como un PU y es caracterizada por IBM durante la instalación o en un tiempo posterior.

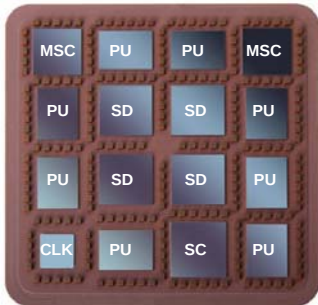


Lámina 58
Roberto Gómez C.



LIC: Licensed Internal Code

- Término usado por IBM para referirse al microcódigo (firmware).
- Cuando se compra un servidor de la familia de las series Z, se adquiere una licencia para usar el microcódigo que se da con la máquina.
- Si alguien se “mete” con el microcódigo IBM, puede, y en algunas ocasiones lo hace, puede llevar a cabo acciones legales contra dicha persona.
- No existe un equivalente en el mundo PC, lo más cercano sería el BIOS.

Lámina 59

Roberto Gómez C.



MSU: Million Services Units

- Medida de la cantidad de trabajo de procesamiento que una computadora puede hacer en una hora.
- Término asociado con mainframes IBM.
- Algunas veces es usado para definir costos, con respecto a los MSUs consumidos o la capacidad total del sistema en MSUs.
- No es una medida exacta.
 - Parecido a una taza de café, tubo de pasta de dientes.
- IBM publica rangos para cada modelo de mainframes
 - Por ejemplo, zSeries z890 Model 110 es un sistema de 4 MSU.
- Los costos de software no son lineales con los MSUs
 - Decrementar o incrementar el número de MSUs no provocan un cambio proporcional en los costos del software

Lámina 60

Roberto Gómez C.



Ejemplos referencias a MSU

- “Our accounting department will need 6 MSUs on a System z9-109 from 10 p.m. to 1:00 a.m. each night in order make sure our quarterly financial statements arrive on time.”
- “You will need 8 more DB2 MSUs for your z900 to handle this year’s Christmas sales rush. Since your current z900 configuration doesn’t have enough capacity to add 8 DB2 MSUs, you need to add another engine. It might be less expensive to upgrade to a z9 because of the double technology dividend.”

Lámina 61

Roberto Gómez C.



CP

- Central Processor.
- Ejecuta código ordinario.
- Procesador disponible para un sistema operativo normal y software de aplicación.

Lámina 62

Roberto Gómez C.



SAP: System Assistance Processor

- Ejecuta código de E/S.
- Todos sistema moderno de mainframe cuenta con al menos un SAP que proporciona un subsistema de E/S.
- Tareas
 - Traducir números de dispositivos a direcciones reales de los CHPIDs, direcciones de unidades de control y números de dispositivos.
- Administra múltiples rutas a unidades de control y lleva a cabo recuperación de errores para errores temporales.
- Sistemas Operativos y aplicaciones no detectan SAPs y los SAPs no usan memoria “normal”.

Lámina 63

Roberto Gómez C.



IFL: Integrated Facility for Linux

- Procesador normal, con una o dos instrucciones deshabilitadas que solo son usadas por z/OS.
- Linux no usa estas instrucciones y puede ser ejecutado por un IFL.
- Linux también puede ser ejecutado en un CP.
- La diferencia es que un IFL no es tomado en cuenta cuando el numero de modelo del sistema es especificado.
 - Esto puede crear una diferencia substancial en los costos del software.

Lámina 64

Roberto Gómez C.



zAAP: zSeries Application on Assistant Processor

- También conocido como IFA: Integrated Facility for Applications.
- Procesador que tiene deshabilitada un número de funciones de tal forma que ningún sistema operativo completo puede ejecutarse en el procesador.
- Sin embargo z/OS puede detectar la presencia de procesadores zAAP y usarlos para ejecutar código de Java.
- El mismo código Java puede ejecutarse en un CP.
- Compra y mantenimiento son menores que con un procesador estándar.

Lámina 65

Roberto Gómez C.



zAAP

- Los trabajos de Java ejecutados están supeditados a la JVM.
- Solo se puede contar con un zAAP por sistema.
- No procesa interrupciones de E/S.
- No puede ejecutar instrucciones del usuario, aquellas que no se encuentran bajo el control de la JVM.
- Solo disponible para z890, z990 y superiores.
- Los procesadores zAAP no se cuentan cuando se especifica el número de modelo del sistema.
- Al igual que los IFLs solo existen para controlar costos en el software.

Lámina 66

Roberto Gómez C.



¿Cómo trabaja el zAAP?

- Trabajo elegible para ser procesado en zAAP debe ser procesado en LPAR con uno más CPs normales.
- JVM decide si el trabajo Java es elegible para ser ejecutado en el zAAP.
- JVM notifica al despachador de z/OS que el trabajo elegible para zAAP esta lista para ser ejecutado.
- Trabajo elegible para zAAP es despachado al zAAP (o al CP).
- Especificaciones del usuario controlan
 - Cuando el trabajo se ejecuta en un zAAP o en un CP.
 - Cuando el trabajo compite con trabajo normal global.

Lámina 67

Roberto Gómez C.



zIIP: z9 Integrated Information Processor

- Procesador especializado para soportar cargas de trabajo de bases de datos.
- Diseñado para contar con software de bajo costo para trabajos seleccionados en el mainframe.
 - Por ejemplo: Business Intelligence (BI), Enterprise Resource Planning (ERP) y Customer Relationship Management (CRM).
- Proporciona acceso directo a DB2 siendo más efectivo en costo y reduciendo la necesidad de contar con múltiples copias de los datos.

Lámina 68

Roberto Gómez C.



zIIP

- Desde un punto de vista de configuración una limitante es que el número de zIIPs por mainframe no puede ser mayor que el numero de procesadores de propósito general.
- Desde un punto de vista de software, el computo en el zIIP no cuenta como MSU.
- Esto significa que el procesamiento hecho en el zIIP no se incluye en los MUS que se cargan al cliente que tradicionalmente paga cuando se usa el CPU del mainframe.

Lámina 69

Roberto Gómez C.



ICF: Integrated Control Facility

- Ejecuta código de acoplamiento.
- Solo ejecuta Licensed Internal Code.
- No se encuentran visibles a un sistema operativo normal o aplicaciones.
- Se puede ver como un gran bloque de memoria usado por varios sistemas para coordinar trabajo.
- ICFs son asignados a LPARs y después son pueden convertirse en coupling facilities.

Lámina 70

Roberto Gómez C.



Spare

- Algunas funciones de PU funcionan como repuesto.
- Si los controladores del sistema detectan un CP o un SAP que no funciona, este puede ser reemplazado con por un PU de reemplazo.
- En varios de los casos, esto puede hacerse sin ninguna interrupción al sistema, aún si la aplicación estaba corriendo en el procesador que falló.

Lámina 71

Roberto Gómez C.



Ejemplo de CP compartidos

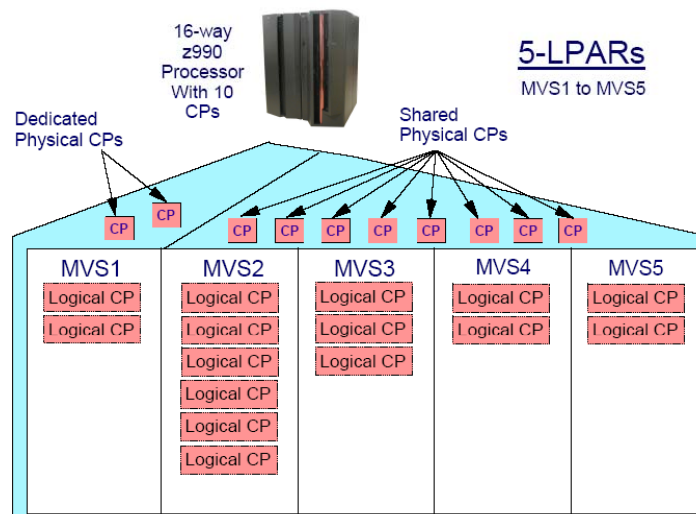
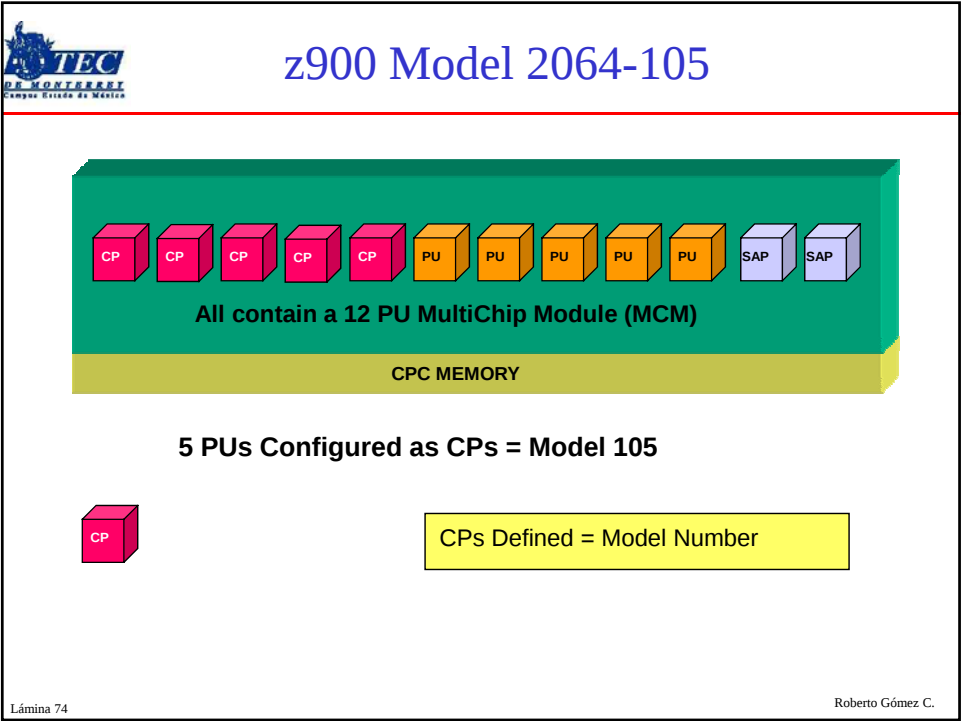
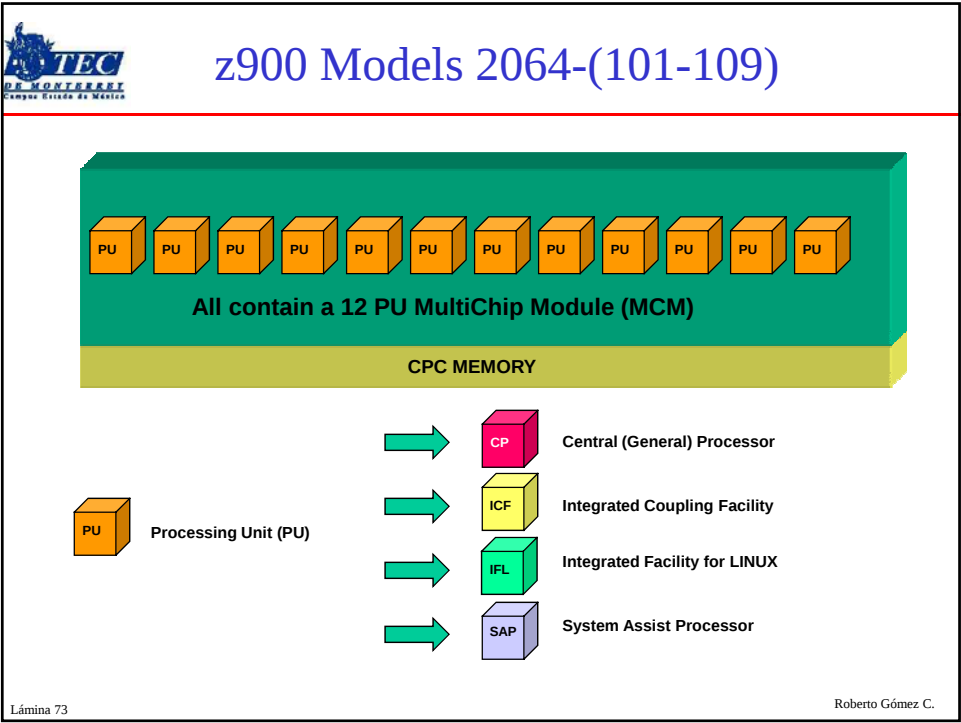
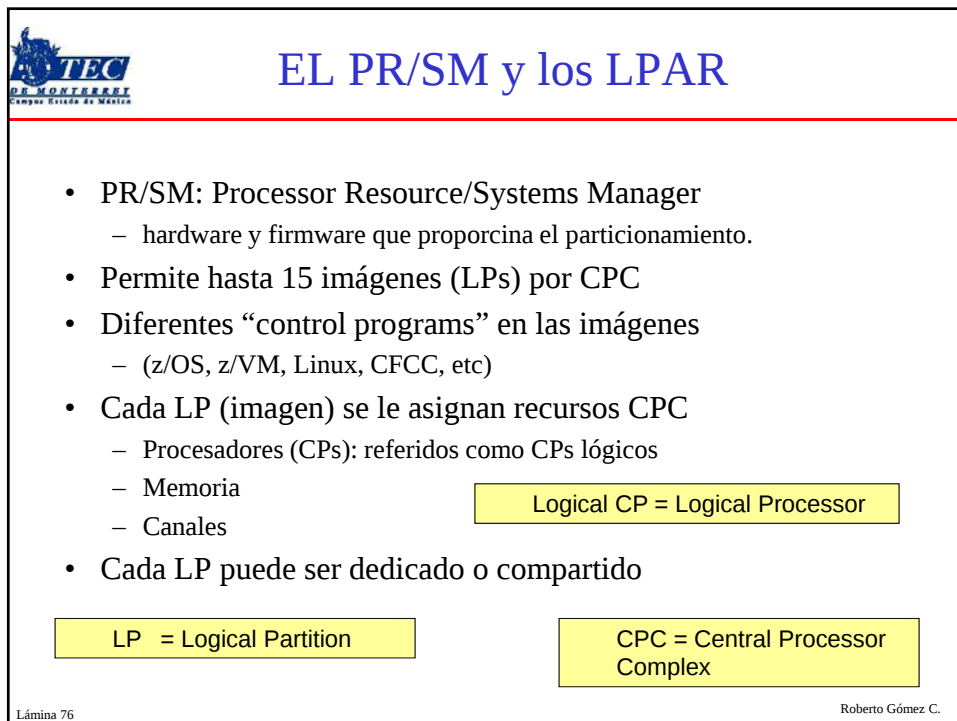
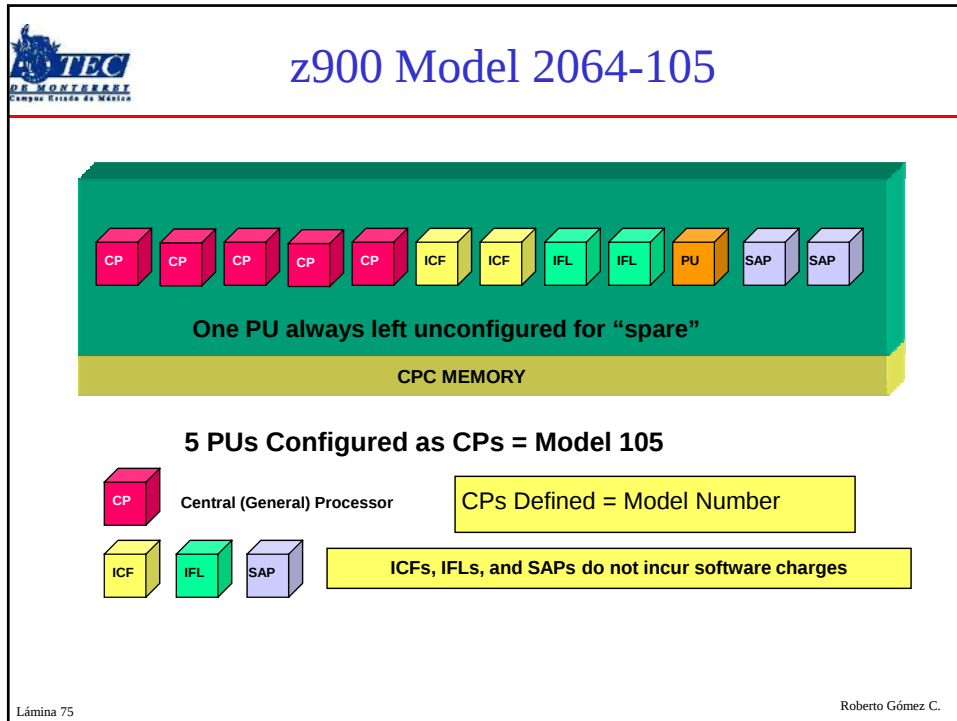


Lámina 72

Roberto Gómez C.







Beneficios de particiones

- Protección/aislamiento de aplicaciones críticas del negocio de cargas de trabajo no críticas.
- Aislamiento de sistema operativos de prueba
- Balanceo de carga
- Diferentes sistemas operativos - - mismo CPs
- Habilidad de garantizar mínimo porcentaje de CP compartido en cada particion
- Más espacio en blanco
 - Habilidad para manejar picos y demandas no predecidas

Lámina 77

Roberto Gómez C.



Desiciones a tomar en la configuración de LPARS

- Definiciones de los LP
 - Dedicado o no dedicado (compartido).
 - Procesadores lógicos (iniciados, reservados).
 - Peso (inicial, mínimo, máximo).
 - Capped o not capped.
 - Asignacion memoria CPC.
 - Configuracion/distribucion canales E/S.
 - Más...

LP = Logical Partition

HMC = Hardware
Management Console

CPC = Central Processor
Complex

Lámina 78

Roberto Gómez C.

Definiendo LPARs dedicados a través del HMC

LPs lógicos son asignados permanentemente a CPs físicos del CPC.

Menos overhead de LPASrs (que LPs compartidos).

LPs dedicados desperdician ciclos de procesadores físicos (CPC) a menos que estén 100% ocupados.

Cuando se encuentran en menos de 100% ocupados los procesadores físicos CPs son asignados a LPs inactivos.

Logical CP = Logical Processor

LP = Logical Partition

CPC = Central Processor Complex

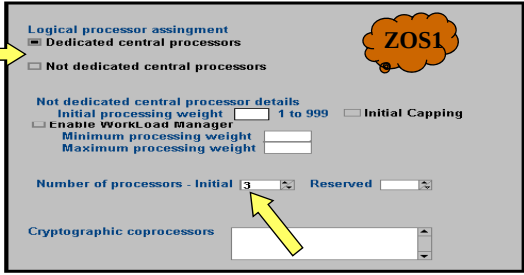
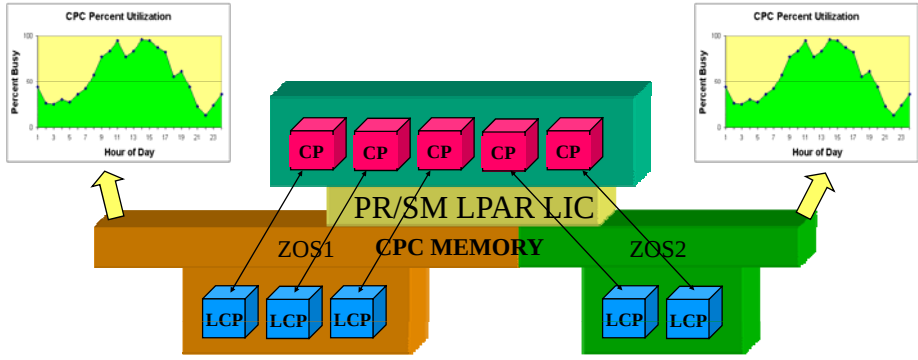


Lámina 79

Roberto Gómez C.

Modelos LPARs dedicado



ZOS1 Image - 3 Dedicated Logical Processors ZOS2 Image - 2 Dedicated Logical Processors

Same problem as basic mode - Unused cycles wasted

LCP = Logical CP = Logical Processor

Lámina 80

Roberto Gómez C.

Definiendo LPARs compartidos

Logical processor assingment

☐ Dedicated central processors

☒ Not dedicated central processors

Not dedicated central processor details

Initial processing weight **400** 1 to 999 ☐ Initial Capping

☐ Enable WorkLoad Manager

Minimum processing weight

Maximum processing weight

Number of processors - Initial **5** Reserved

Cryptographic coprocessors

Lámina 81

Roberto Gómez C.

Definiendo LPARs dedicados

Logical processor assingment

☐ Dedicated central processors

☒ Not dedicated central processors

Not dedicated central processor details

Initial processing weight **100** 1 to 999 ☐ Initial Capping

☐ Enable WorkLoad Manager

Minimum processing weight

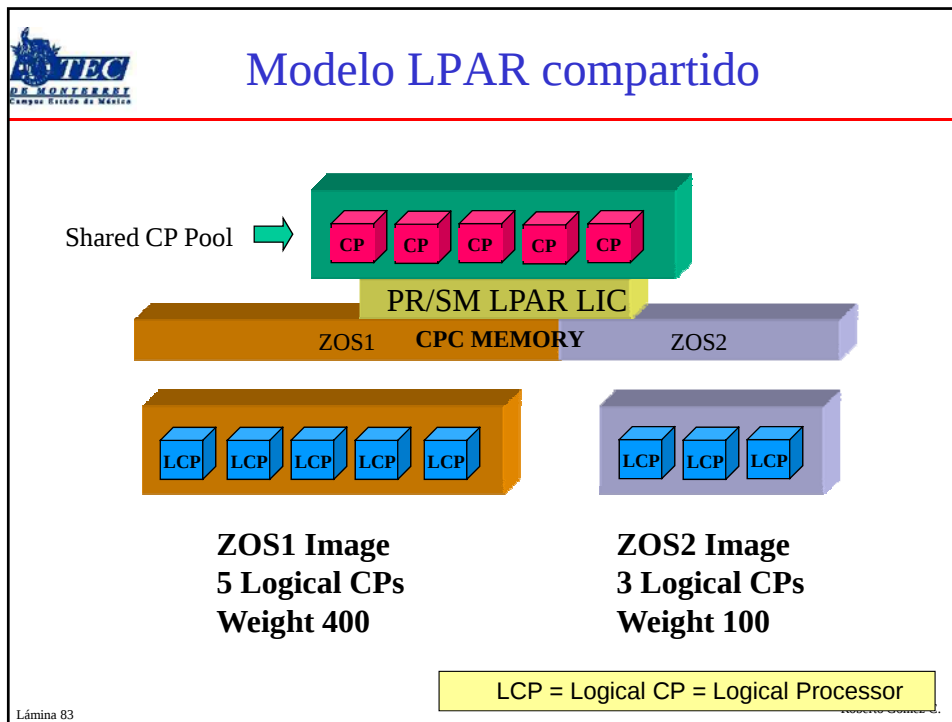
Maximum processing weight

Number of processors - Initial **3** Reserved

Cryptographic coprocessors

Lámina 82

Roberto Gómez C.



El LPAR LIC y el dispatching

¿Qué hace el LLIC (LPAR Licensed Internal Code)?

- LCP (Logical Central Procesor) son consideradas unidades de trabajo calendarizables.
- LCPs se encuentran en una cola de listos
- LLIC ejecuta en un CP físico
 - selecciona un LCP listo
 - lo atiende en CPs reales
- z/OS ejecuta en CP físicos hasta que su tiempo expira (12.5 a 25 milisegundos) o hasta que z/OS entra a un estado de espera.
- Ambiente almacenado, LLIC ejecuta en un CP libre
- Si LCP no termino ejecución, se vuelve a poner en la cola de listos

LLIC = LPAR Licensed Internal Code

CP = Central Processor LCP = Logical CP = Logical Processor

Lámina 84

Roberto Gómez C.



Seleccionando LCPs (CPs lógicos)

- La prioridad en la cola de “listos” es determinada por el PR/SM LIC
 - Basado en el uso “actual” de un CP lógico contra el uso “calculado” (targetted utilization)
- Uso “calculado” es determinado como una función del número de LCPs y el peso de LP
 - El peso de LP es un número especificado por el usuario entre 1 y 999 (se recomiendan 3 dígitos)

LP = Logical Partition

LLIC = LPAR Licensed Internal Code

LCP = Logical CP = Logical Processor

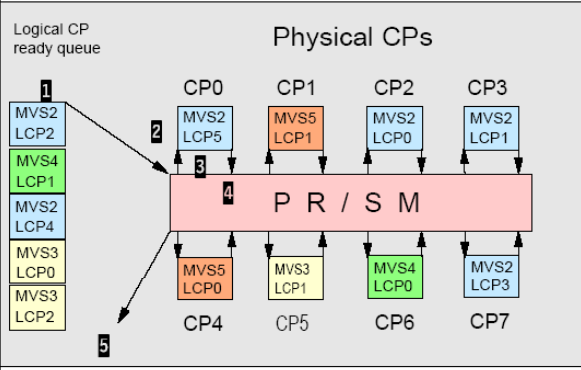
CP = Central Processor

Lámina 85
Roberto Gómez C.



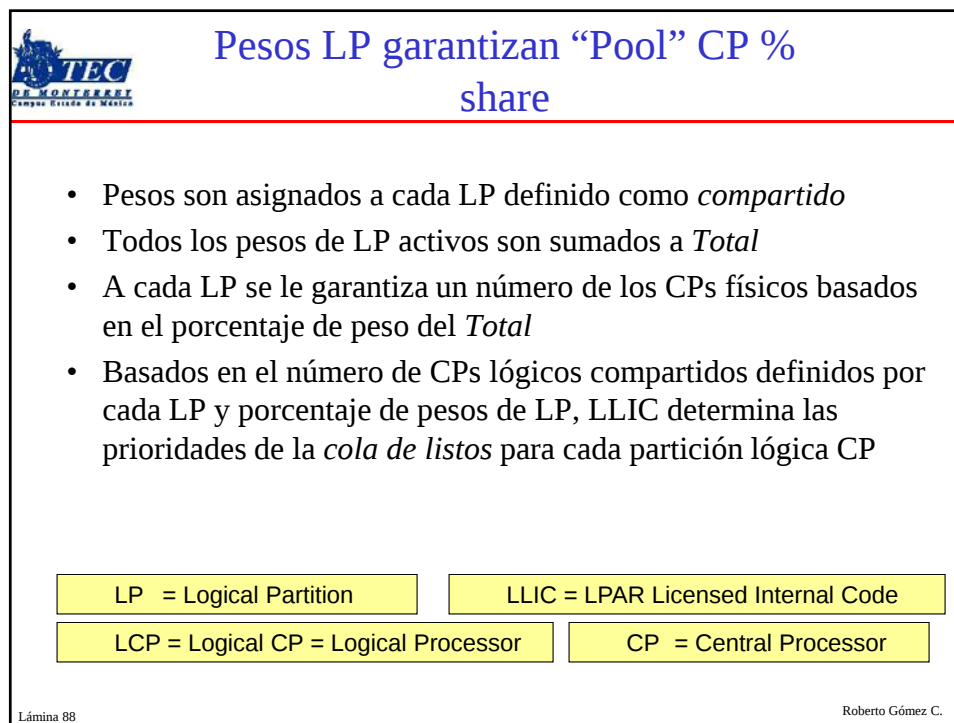
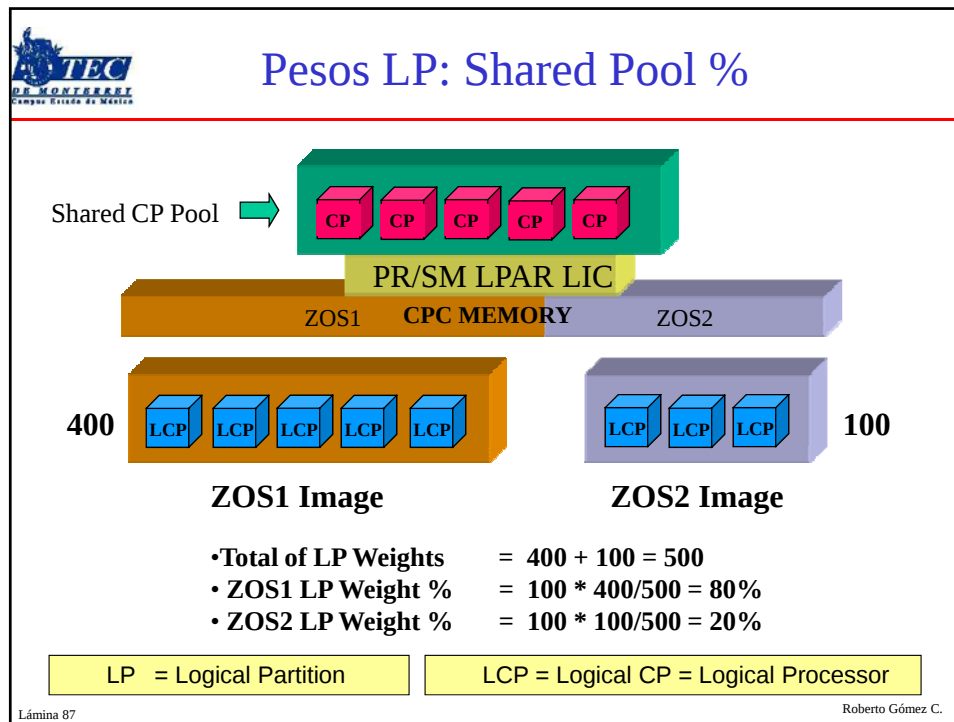
LPAR Logical Dispatching

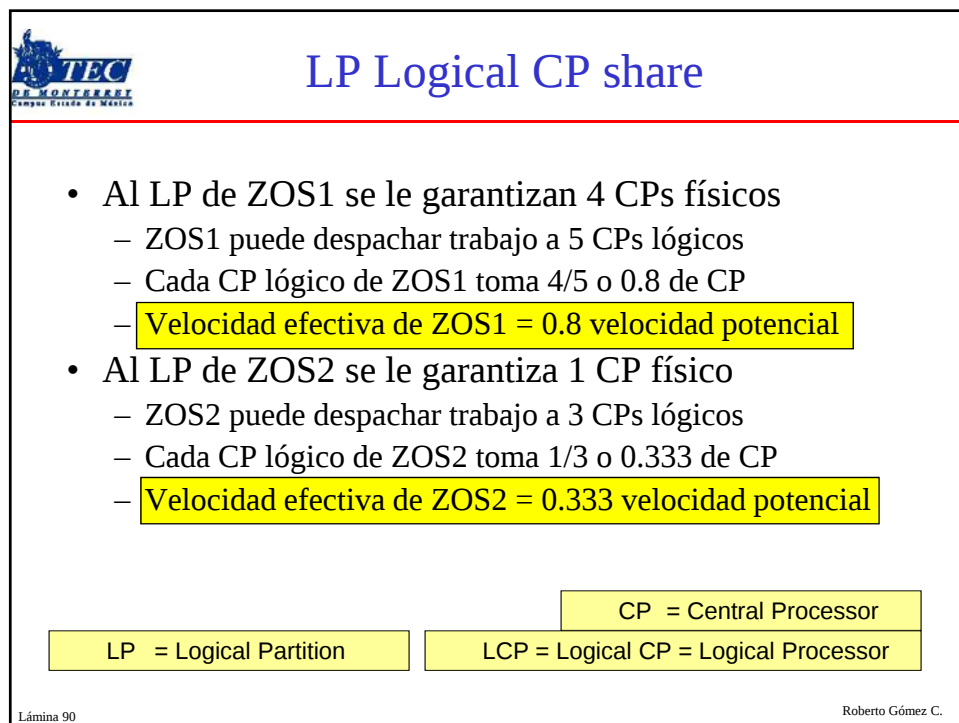
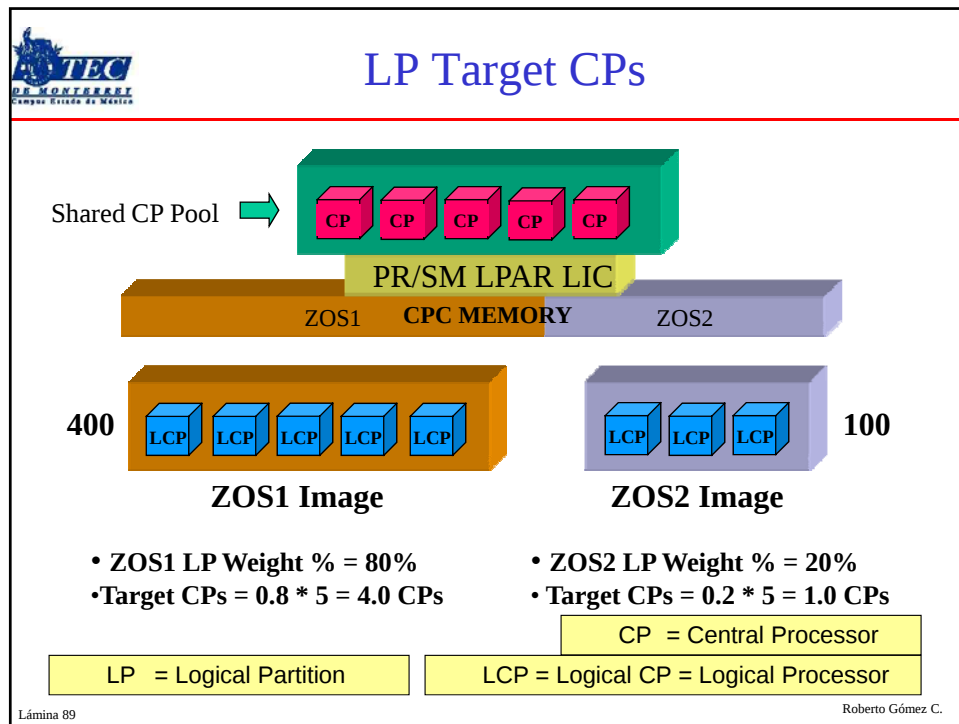
CPC in LPAR Mode



- El próximo CP lógico a atender es elegido del CP que se encuentre listo.
- LPAR LIC atiende el CP seleccionado en un CP físico en el CPC
- La unidad z/OS se ejecuta en el procesador lógico CP0. Ejecuta hasta que su tiempo expira, o se suspende.
- Cuando se termina su tiempo, ambiente CP0 se almacena y control se pasa a LPAR LIC, que empieza a ejecutar en CPO de nuevo
- LPAR LIC determina el siguiente CP lógico a utilizar.

Lámina 86
Roberto Gómez C.







Impacto del cambio de pesos

- El peso de un LP activo puede ser cambiado sin perjudicar al sistema usando la consola del sistema
 - Incrementar el peso de un LP por x, sin cambiar otras configuraciones, incrementa la parte que comparte en el pool a expensas de otros LPs compartidos
 - Esto se debe a que el TOTAL del peso compartido de LP incrementó, mientras que los pesos de los otros LPs permanecen constantes

$$\frac{\text{Peso LP}_n}{\text{Total}} > \frac{\text{Peso LP}_n}{\text{Total} + x}$$

LP = Logical Partition

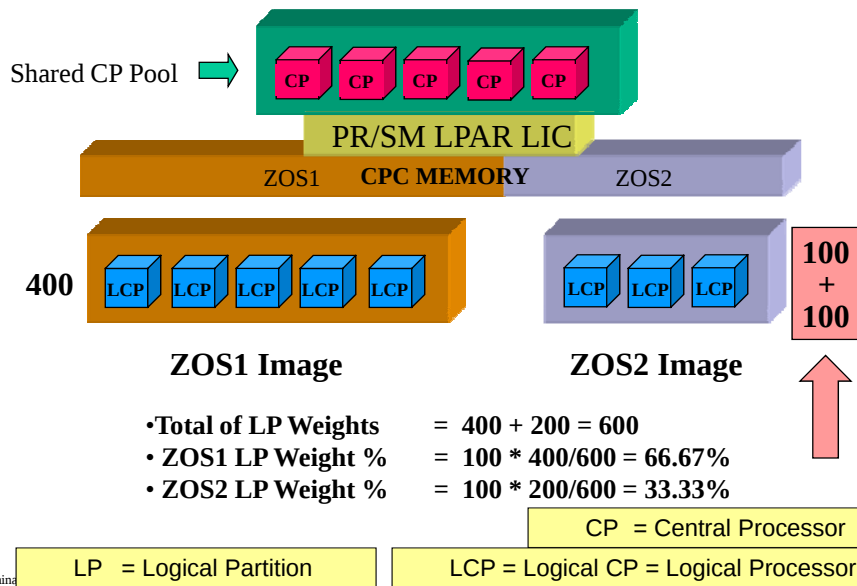
CP = Central Processor

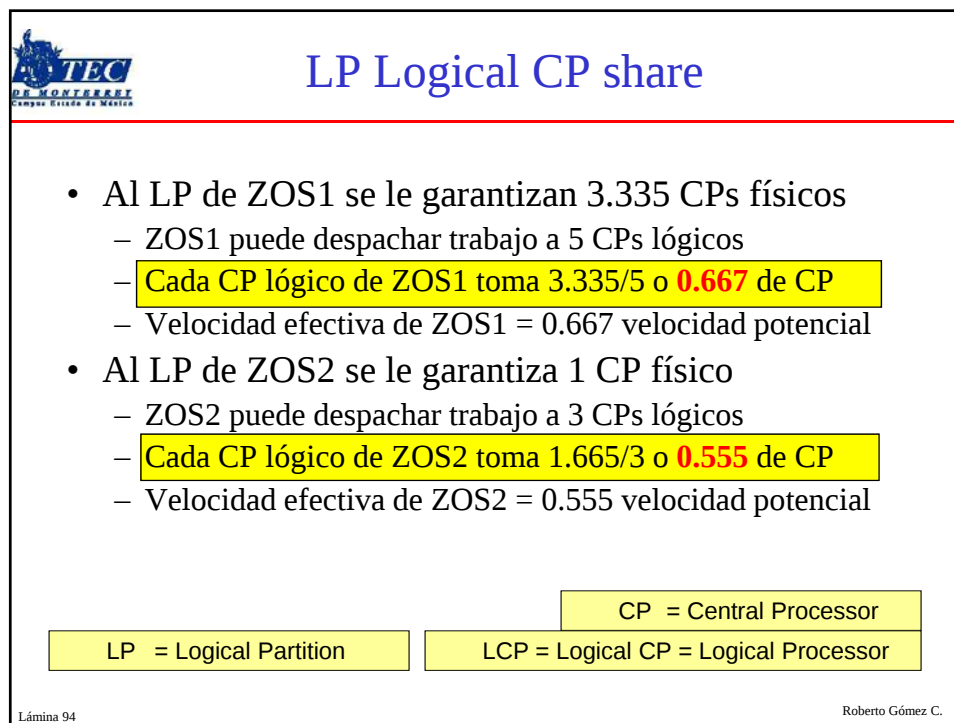
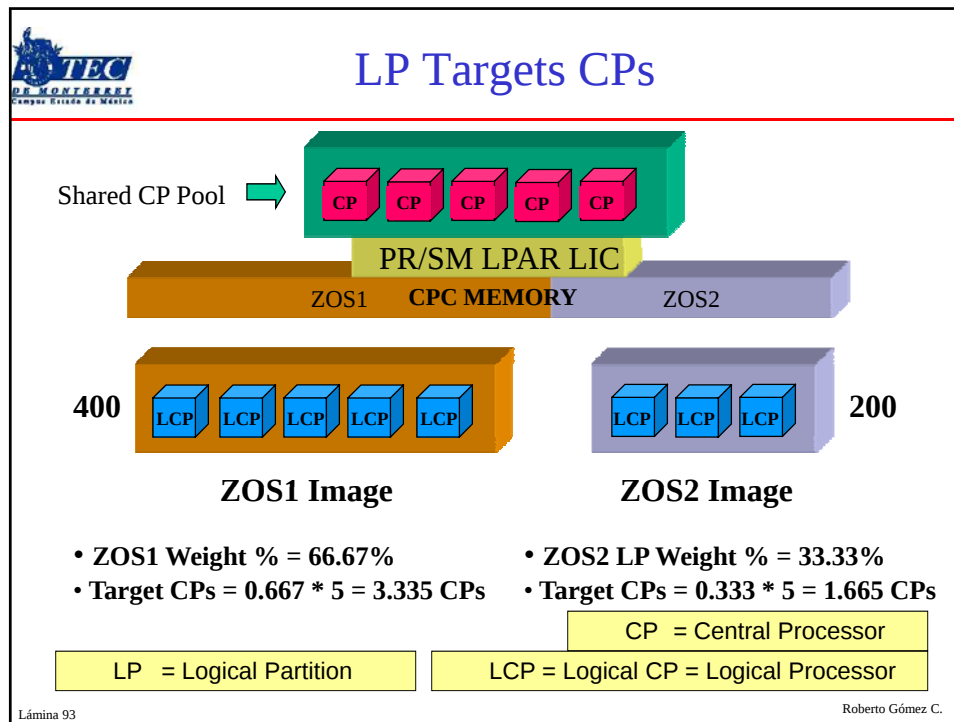
Lámina 91

Roberto Gómez C.



Cambiando los pesos del LPAR







Cambiando el número de CPs lógicos

- Se pueden incrementar o reducir el número de CPs lógicos
- Cambiar el número de CPs lógicos para un LP compartido incrementa o decrementa el trabajo potencial de un LP
 - Cambia el overhead de z/OS y PR/SM
 - No cambia el % CPC del pool share
 - Cambia la “velocidad efectiva” del CP en el LP lógico

CPC = Central Processor Complex

CP = Central Processor

LP = Logical Partition

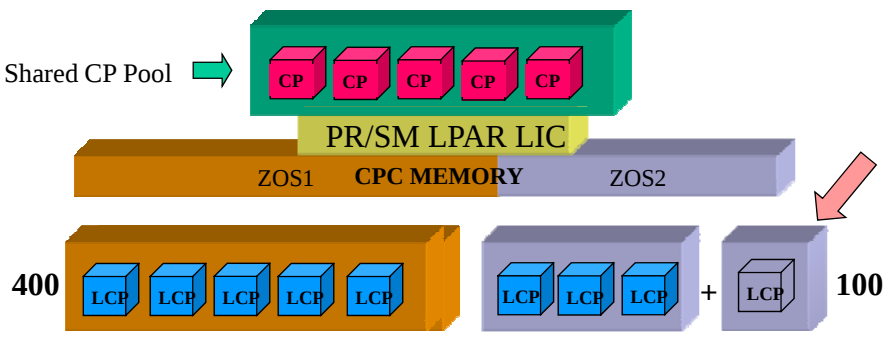
LCP = Logical CP = Logical Processor

Lámina 95
Roberto Gómez C.



Añadiendo CPs lógicos

Shared CP Pool →



ZOS1 Image

400 LCPs

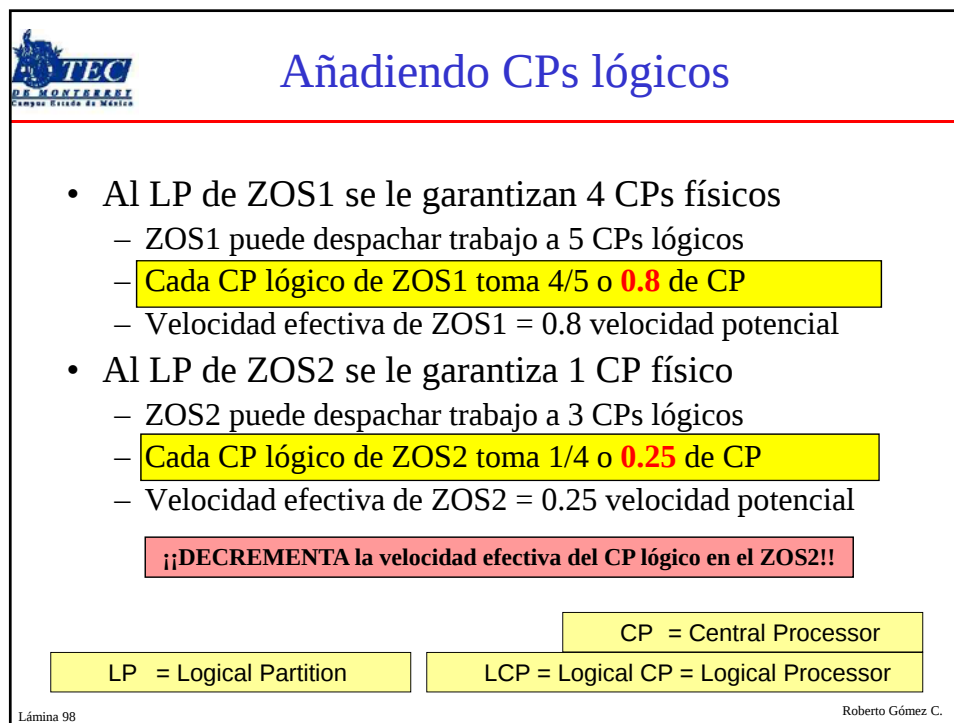
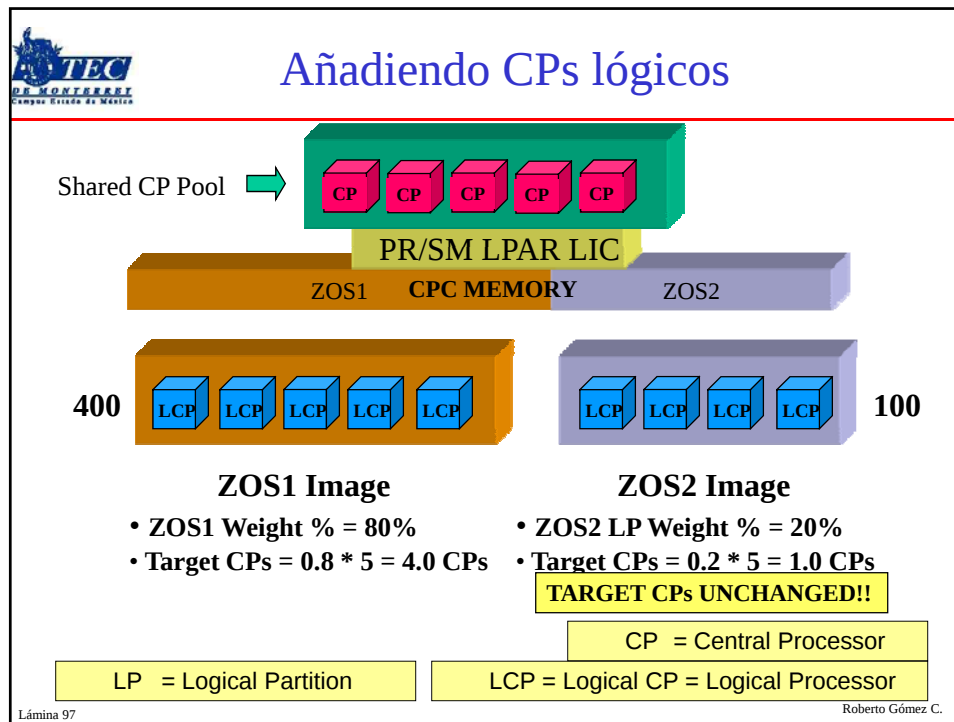
ZOS2 Image

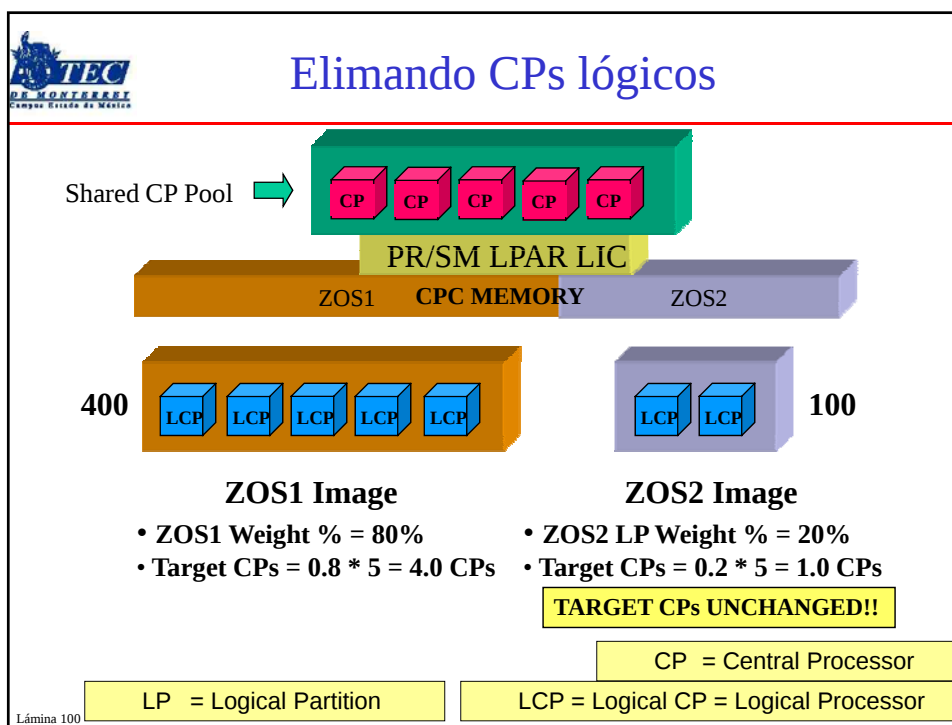
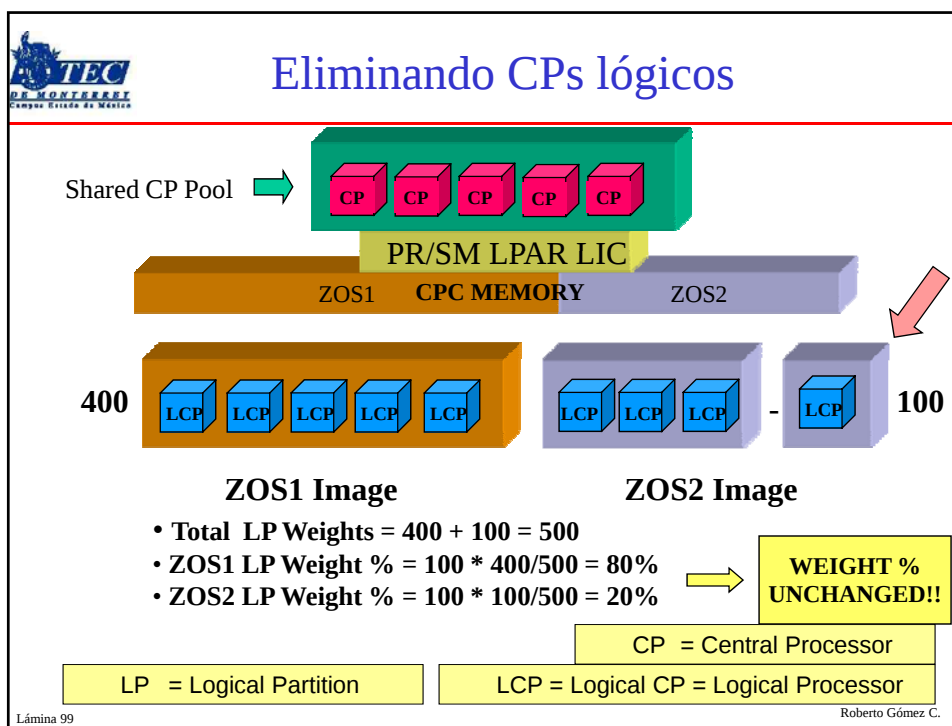
100 LCPs

- Total LP Weights = 400 + 100 = 500
- ZOS1 LP Weight % = $100 * 400 / 500 = 80\%$
- ZOS2 LP Weight % = $100 * 100 / 500 = 20\%$

→ **WEIGHT % UNCHANGED!!**

Lámina 96
Roberto Gómez C.







Eliminando CPs lógicos

- Al LP de ZOS1 se le garantizan 4 CPs físicos
 - ZOS1 puede despachar trabajo a 5 CPs lógicos
 - Cada CP lógico de ZOS1 toma 4/5 o **0.8** de CP
 - **Velocidad efectiva de ZOS1 = 0.8 velocidad potencial**
- Al LP de ZOS2 se le garantiza 1 CP físico
 - ZOS2 puede despachar trabajo a 2 CPs lógicos
 - Cada CP lógico de ZOS2 toma 1/2 o **0.5** de CP
 - **Velocidad efectiva de ZOS2 = 0.5 velocidad potencial**

¡¡INCREMENTA la velocidad efectiva del CP lógico en el ZOS2!!

LP = Logical Partition

LCP = Logical CP = Logical Processor

Lámina 101

Roberto Gómez C.



Entonces, ¿cuántos CPs lógicos?

- Tanto el overhead de z/OS y PR/SM disminuyó cuando el número de LCP fue igual a los requerimientos físicos del CP de la carga de trabajo ejecutada.
- El número de LCPs en línea a un LP es correcto ... a veces ...
 - Cuando el LP esta limitada por el CPU, muy pocos.
 - Cuando el LP esta desocupada, muchos.
 - Cuando el LP esta cerca de 100% ocupado, es lo ideal
- Idealmente, la velocidad efectiva del LCP es 1.0

Lámina 102

Roberto Gómez C.



LP “Hard” Capping

HMC Image Profile

- Peso inicial impuesto
- El LLIC no permitirá al LP usar más de porcentaje del pool compartido, aún si otros LPs se encuentra desocupados.
- Cambio dinámico al status del capping.
- Dynamic change to capping status
 - Capped o NO capped
 - Valor del peso capped

Logical processor assignment

☐ Dedicated central processors

☒ Not dedicated central processors

Not dedicated central processor details

Initial processing weight 1 to 999 ☒ Initial Capping

☐ Enable WorkLoad Manager

Minimum processing weight

Maximum processing weight

Number of processors - Initial Reserved

Cryptographic coprocessors

¡En general no es recomendado!

LLIC = LPAR Licensed Internal Code

LP = Logical Partition

Lámina 103

Roberto Gómez C.



Clustering

Tipos y características

Lámina 104

Roberto Gómez C.



Clustering

- Se ha llevado a cabo por muchos años en formas diversas, desde los tiempos del S/360.
- Tres niveles de clustering
 - Basic shared DASD
 - Anillos CTC
 - Parallel Sysplex
- La mayor parte de las instalaciones z/OS de hoy en día usan uno o más de estos niveles.
 - Una instalación aislada de z/OS es relativamente rara.

Lámina 105

Roberto Gómez C.



El concepto de imagen

- Usado para describir un solo sistema z/OS el cual puede ser standalone o un LPAR en una caja más grande.
- Una imagen puede existir en un S/390, o en un servidor zSerie con LPARs, o puede existir en un LPAR, o bajo z/VM.
- Un sistema con seis LPARs, cada uno con un sistema z/OS por separado, cuenta con seis imágenes z/OS.
- Se utiliza el termino de imagen para indicar que no interesa donde un sistema z/OS se encuentra corriendo.

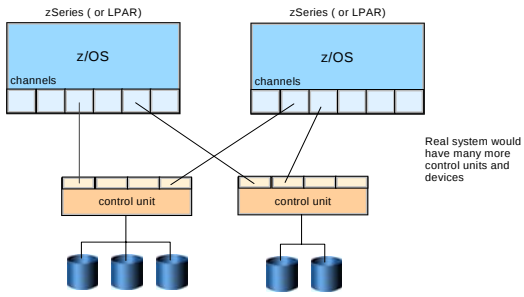
Lámina 106

Roberto Gómez C.



Basic shared DASD

- Las imágenes z/OS pueden ser cualquier versión anterior del sistema operativo, o dos LPARs en el mismo sistema.
 - No hay diferencia en el concepto u operación.



Real system would have many more control units and devices

Lámina 107
Roberto Gómez C.



Características Basic shared DASD

- Capacidad limitada.
- El sistema operativo automáticamente usa comandos RESERVE y RELEASE sobre un DASD antes de interactuar con él.
 - Comando RESERVE limita acceso al sistema que ejecuto el comando y esto perdura hasta que el comando RELEASE se ejecuta.
- Comandos trabajan bien por periodos limitados.
 - Actualización de datos.
 - Aplicaciones los pueden utilizar para proteger data sets durante la duración de un aplicación.

Lámina 108
Roberto Gómez C.



Usos

- Usado por el staff de operaciones para controlar que trabajo actúa sobre que sistema y asegurarse que no hay conflictos.
- Ambiente muy útil para pruebas, recuperación y balanceo de carga.
- Otros tipos de dispositivos o unidades de control pueden atarse a ambos sistemas.
 - Por ejemplo: Unidad control cinta, con varios drives de cintas.

Lámina 109

Roberto Gómez C.



Anillos CTC

- Mismo principio que shared DASD, pero con conexiones CTC entre los sistemas
 - CTC: Channel To Channel
- Esto se conoce como Anillo CTC.
- El aspecto de anillo se aprecia mas cuando hay más de dos sistemas.

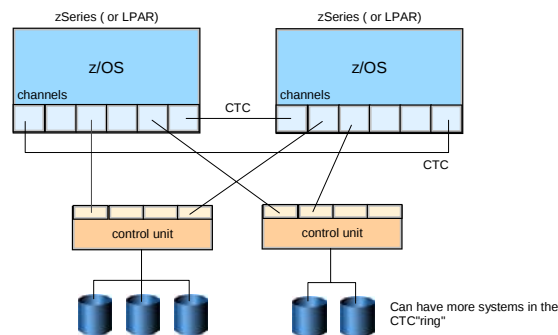


Lámina 110

Roberto Gómez C.



Información transmitida entre sistemas

- Usado para pasar información de control entre sistemas vía el anillo CTC.
- La información incluye:
 - Información de uso y bloqueo de data sets en disco. Lo que permite al sistema prevenir de forma automática acceso duplicado a los data sets.
 - Bloqueo basado en especificaciones JCL
 - Información de colas de trabajos para que todos los sistemas en el anillo puedan aceptar trabajos de una sola cola.
 - Por ejemplo todos los sistemas pueden enviar a imprimir a una sola cola de salida
 - Controles de seguridad que permiten decisiones de seguridad a través de todo el sistema.

Lámina 111 Roberto Gómez C.



Sysplex

- Colección de sistemas z/OS que cooperan, usando ciertos productos de software y hardware para procesar trabajo.
- Tecnología de clustering que puede proporcionar disponibilidad.
- Sysplex vs sistemas grandes
 - Sistemas computacionales convencionales grandes también utilizan hardware y software para cooperar en el procesamiento de trabajos.
 - Diferencia mayor con sysplex: potencial de crecimiento y nivel de disponibilidad en el sysplex.

Lámina 112 Roberto Gómez C.



Parallel Sysplex

- Es un sysplex que usa una tecnología de datos compartidos entre varios sistemas.
- Permite acceso concurrente de lectura/escritura de datos compartidos para todos los nodos (o servidores) de la configuración.
 - Cada nodo puede almacenar datos en la memoria de un procesador local.
- Resultado: peticiones de trabajo asociadas con una carga simple como transacciones de negocios o peticiones bases de datos pueden ser dinámicamente distribuidas para ejecución paralela en el cluster sysplex basadas en la capacidad disponible del procesador.

Lámina 113

Roberto Gómez C.



Componentes hardware de un Parallel Sysplex

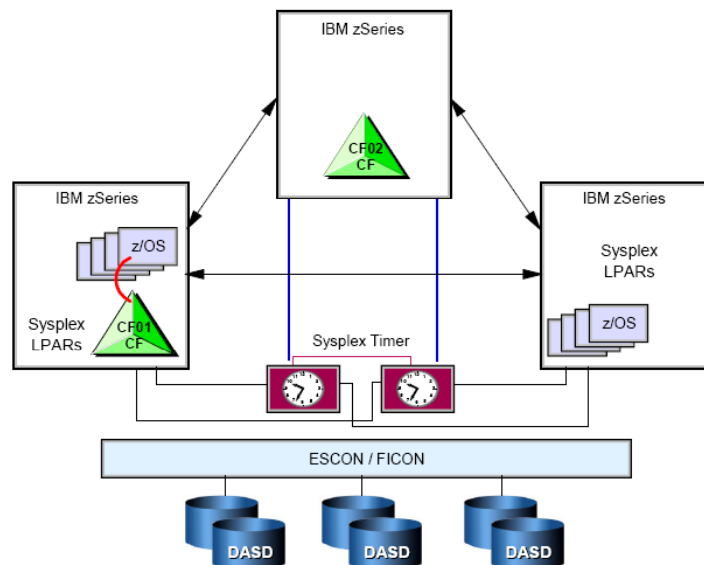


Lámina 114

o Gómez C.



Características Parallel Sysplex

- Parece un único sistema grande.
- Cuenta con solo una interfaz que controla todos los sistemas.
- Con una buena planeación y operación (ninguna de las cuales es trivial), cargas complejas pueden ser compartidas por cualquier o todos los sistemas y la recuperación puede ser automática para muchas cargas.

Lámina 115

Roberto Gómez C.



La facilidad de acoplamiento

- Componente básico de un Parallel Sysplex
- Procesador del mainframe, con memoria y canales especiales y un sistema operativo propio.
- No cuenta con dispositivos de E/S, aparte de los canales especiales, y el sistema operativo es muy pequeño.
- Es utilizado para los siguientes propósitos:
 - Poner un candado sobre información que es compartida entre todos los sistemas atados.
 - “Cachar” (cache) información (como de base de datos) que es compartida entre todos los sistemas “atados”.
 - Información lista datos que es compartida entre todos los sistemas “atados”.

Lámina 116

Roberto Gómez C.



Coupling Facility

- La información reside en memoria y un CF típicamente cuenta con una gran cantidad de memoria.
- Puede ser un sistema por separado y un LPAR puede ser usado como un CF.

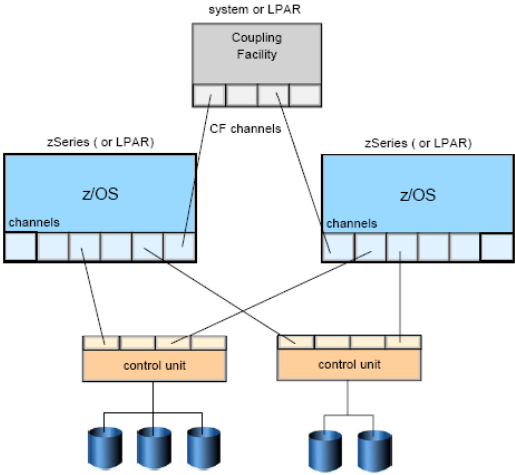


Lámina 117
Roberto Gómez C.



Tecnologías de clustering para mainframes

- Tecnología Parallel Sysplex ayuda a mantener continuidad.
- Permite unir hasta 32 servidores con escalabilidad lineal.
- Cada servidor puede ser configurado para acceder a recursos y una instancia “clonada” de una aplicación puede correr en cada servidor.
- Posible añadir y cambiar sitios dentro del Parallel Sysplex.
- Puntos a tomar en cuenta
 - Shared data clustering
 - Nondisruptive maintenance

Lámina 118
Roberto Gómez C.



Shared data clustering

- Cada servidor del Parallel Sysplex tiene acceso a todos los recursos y cada aplicación clonada puede correr en cualquier servidor.
- El enfoque “shared data” permite balancear la carga dinámicamente entre todos los servidores.
- En el caso de una sobrecarga del sistema, programada o no, las cargas de trabajo puede ser dinámicamente redirigidas a servidores disponibles, proporcionando disponibilidad de servicio.

Lámina 119

Roberto Gómez C.



Nondisruptive maintenance

- Mantenimiento de hardware y software sin interrumpir el trabajo.
- Servidores pueden ser removidos o añadidos al cluster de forma dinámica, permitiendo actividades de instalación y mantenimiento mientras que los sistemas restantes continúan con el procesamiento del trabajo.
- Actualizaciones de hardware y software pueden ser introducidas en un sistema a la vez.
- Esto permite rotar los cambios en todos los sistemas, a un ritmo que tenga sentido para el negocio.

Lámina 120

Roberto Gómez C.



Ejemplos sistemas típicos mainframes

- Tres niveles generales de configuración
 - Sistemas muy pequeños.
 - Sistemas medios.
 - Sistemas grandes.

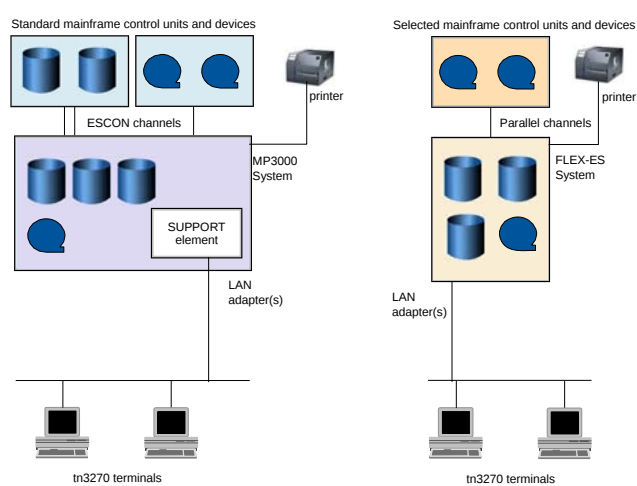
Lámina 121

Roberto Gómez C.



Ejemplo de dos sistemas pequeños

- Dos sistemas pequeños
 - Izquierda: Multiprise 3000, diseñado para pequeñas instalaciones con drives de discos internos
 - Derecha: es un sistema emulador de FLEX-ES, que corre en una PC corriendo Linux o UNIX



Mainframe refers more to a style of computing rather than to unique hardware.

Lámina 122

Roberto Gómez C.



TEC
DE MONTERREY
Campus Ciudad de México

Configuración mediana de un mainframe

- Implementación completamente moderna sin dispositivos viejos.
- Un z890 con dos controladores de disco externo, drivers de cintas, impresoras, conexiones a LAN y consolas.

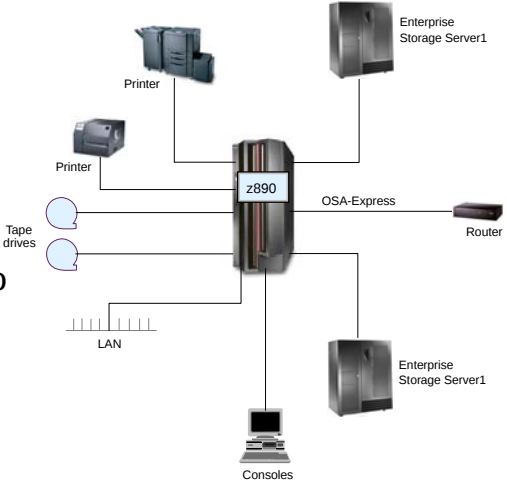


Lámina 123

Roberto Gómez C.



TEC
DE MONTERREY
Campus Ciudad de México

Características del sistema

- Sistema pueden correr varios LPARs
 - Una imagen de producción z/OS con aplicaciones interactivas
 - Una segunda imagen de producción LPAR dedicada a trabajos batch
 - Un imagen de prueba z/OS LPAR para probar software nuevo
 - Uno o mas LPARs corriendo Linux, tal vez corriendo aplicaciones web

Lámina 124

Roberto Gómez C.

ITEC
DE MONTERREY
Campus Ciudad de México

Configuración moderadamente larga

- Dos máquinas
 - una nueva 2990 y
 - una vieja 9672
- Parallel Sysplex con Coupling Facility
- Múltiples ESS con un viejo DASD conectado via un switch
- Controladores comunicación 3745 para una red SNA
- Drivers cinta 3490 retenidas para compatibilidad
- Conexiones OSA Express a varias LANs
- Consolas

Lámina 125
Roberto Gómez C.

ITEC
DE MONTERREY
Campus Ciudad de México

Comentarios sistemas grandes

- Mezcla de generaciones de dispositivos que puede encontrarse en cualquier empresa.
- Nuevos dispositivos son incorporados, pero viejos dispositivos se encuentran aún disponibles hasta que alcancen el final de su vida útil.
- Generalmente z/OS corre en dispositivos viejos hasta que cambios arquitecturales fuerzan su retiro
- Este tipo de cambio es publicado con anterioridad y se dan a conocer incompatibilidades que usualmente coinciden con el inicio del fin de vida útil de la máquina.

Lámina 126
Roberto Gómez C.



Continuidad mainframe con Parallel Sysplex

- Concurrencia para facilitar mantenimiento.
- DASD con sistemas de espejo y tecnología RAID.
- Tecnología de red con conexiones tolerantes a fallas.
- Subsistemas de E/S soportan paths múltiples de E/S y switcheo dinámico.
- Componentes software z/OS permiten contar con versiones nuevas y antiguas de componentes de software
- Aplicaciones orientadas a datos compartidos
- Procesos de recuperación y operación son completamente automatizados y transparentes para los usuarios.

Lámina 127

Roberto Gómez C.



Beneficios Parallel Sysplex

- No puntos de falla únicos.
- Capacidad y escalamiento.
- Balanceo dinámico de carga.
- Facilidad de uso.
 - WLM: Workload Management
 - SFM: Sysplex Failure Manager
 - ARM: Automatic Restart Manager
 - Clonnig and symbolics
 - zSeries resource sharing
- Imagen de un solo sistema.
- Cambios compatibles y crecimiento que no interrumpe.
- Compatibilidad aplicaciones.
- Recuperación de desastres.

Lámina 128

Roberto Gómez C.



Manejo del peso en WLM LPAR

- Dinámicamente se cambian los pesos del LP
- WLM evalúa todas las cargas de trabajo del SYSPLEX.
- Suffering Service Class Periods (SSCPs)
 - High (>1) SYSPLEX Performance Index (PI)
 - High Importance
 - CPU delays

Lámina 129

Roberto Gómez C.



El IRD Intelligent Resource Director

- Aumentar fortalezas plataforma a través de la integración
 - Workload Manager
 - Parallel Sysplex
 - PR/SM
 - Channel Subsystem
- Ver un cluster de LPs en una zSerie como un solo repositorio de recursos computacionales
 - Recursos manejados inicialmente: CPU y E/S
 - Requiere Parallel Sysplex, WLM Goal Mode, WLM Structure y Level 9 Coupling Facility
 - z/OS V1,2, z/VM y Linux para zSeries soporte para manejo de cargas LPAR

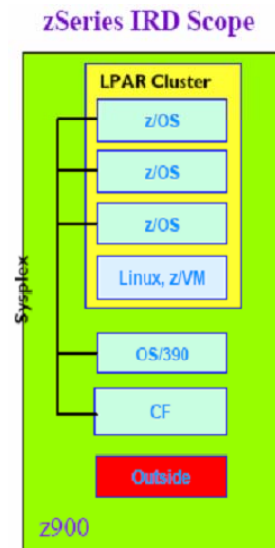
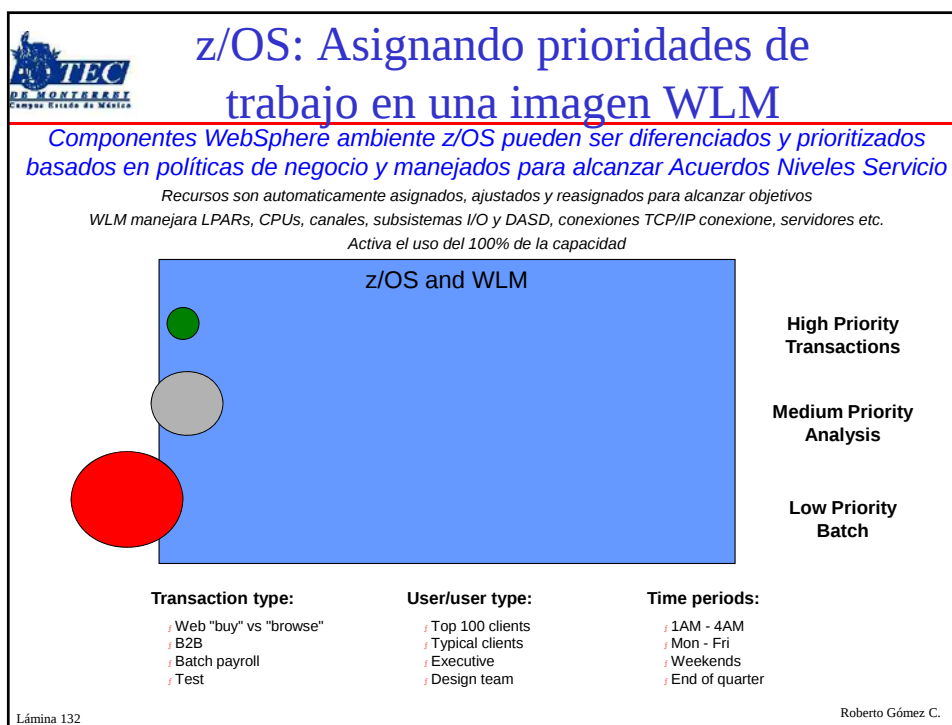
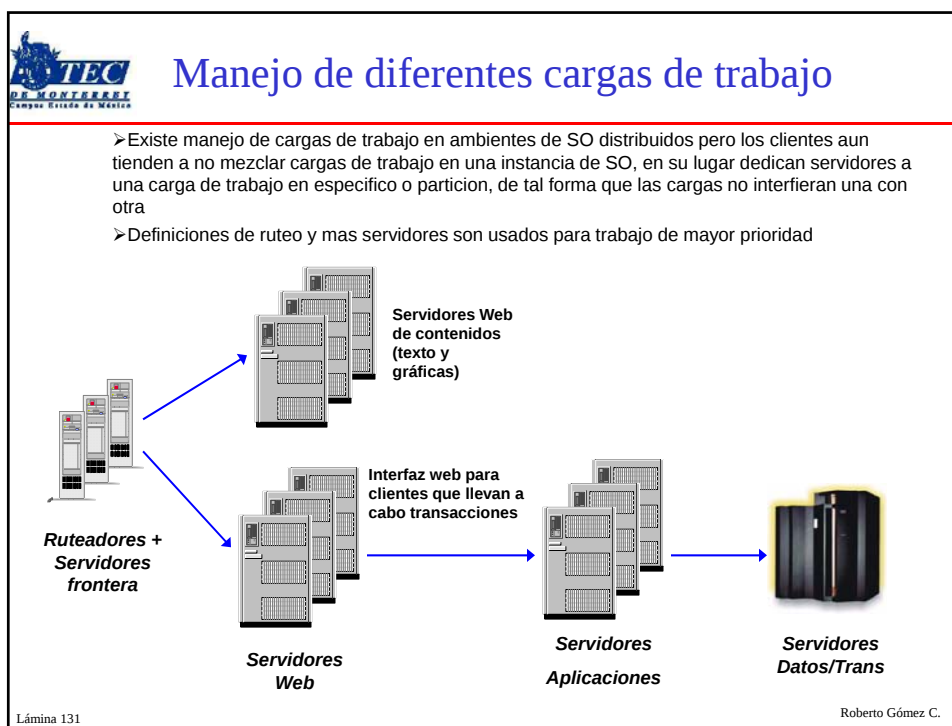


Lámina 130

Roberto Gómez C.



z/OS: Asignando prioridades de trabajo en un servidor - IRD

"Intelligent Resource Director (IRD)" diferencia de forma fuerte a z/OS con su habilidad para manejar recursos a través de múltiples particiones en un servidor

PR/SM, IRD y WLM trabajan juntos para asegurar que los recursos del servidor están correctamente balanceados para habilitar trabajos para cumplir objetivos establecidos en las políticas

MAS

The diagram illustrates a zSeries server architecture. On the left, a z/VM partition (yellow) contains a Linux instance (red text). A green arrow labeled 'MAS' points to this partition. To the right, there are three z/OS partitions (blue). The first two are labeled 'Producción' and contain 'WAS', 'DB2', and 'CICS'. The third is labeled 'Pruebas & trabajos en lote' and 'Low Priority'. The entire setup is labeled 'zSeries' at the bottom.

Recursos procesador, ancho banda y colas decisiones de E/S son perfectamente balanceados a través del servidor para manejar diferentes cargas de trabajo dentro de los parámetros de los objetivos establecidos

Lámina 133 Roberto Gómez C.

Manejo de diferentes cargas de trabajo

The diagram shows a large light blue rectangle representing a workload. A green circle is positioned in the top-left corner, indicating high priority work. To the right of the rectangle, there are three labels: 'Transacciones de alta prioridad', 'Análisis de mediana prioridad', and 'Trabajos en lotes baja prioridad'.

Transacciones de alta prioridad

Análisis de mediana prioridad

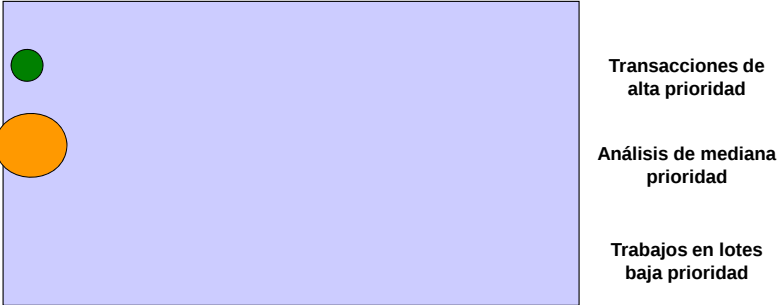
Trabajos en lotes baja prioridad

Cargas de trabajo pueden afectarse entre Una carga de trabajo lenta de baja prioridad puede afectar a cargas de trabajo de mayor prioridad.

Lámina 134 Roberto Gómez C.



Manejo de diferentes cargas de trabajo



Transacciones de alta prioridad

Análisis de mediana prioridad

Trabajos en lotes baja prioridad

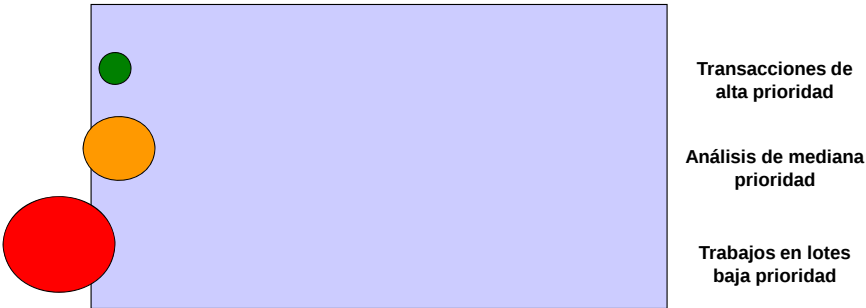
Cargas de trabajo pueden afectarse entre Una carga de trabajo lenta de baja prioridad puede afectar a cargas de trabajo de mayor prioridad.

Lámina 135

Roberto Gómez C.



Manejo de diferentes cargas de trabajo



Transacciones de alta prioridad

Análisis de mediana prioridad

Trabajos en lotes baja prioridad

Cargas de trabajo pueden afectarse entre Una carga de trabajo lenta de baja prioridad puede afectar a cargas de trabajo de mayor prioridad.

Lámina 136

Roberto Gómez C.

 **Manejo de diferentes cargas de trabajo**



Transacciones de alta prioridad

Análisis de mediana prioridad

Trabajos en lotes baja prioridad

Cargas de trabajo pueden afectarse entre Una carga de trabajo lenta de baja prioridad puede afectar a cargas de trabajo de mayor prioridad.

Lámina 137 Roberto Gómez C.

 **Arquitectura mainframes familia Z**

Roberto Gómez Cárdenas
rogomez@itesm.mx

Lámina 138 Roberto Gómez C.