

MORSE



**IBM Power5: Tecnología
para la era on demand**

“

Nuestra estrecha relación con IBM, unida a las inversiones realizadas en conocimientos técnicos, instalaciones para pruebas de rendimiento y soporte de servidores IBM eServer, almacenamiento, software y servicios, nos permiten ofrecer las mejores soluciones informáticas integradas para ayudar a nuestros clientes en su trayectoria de evolución hacia las tecnologías de la información adaptadas a la empresa.

”



We'll take you there

INTRODUCCIÓN

En los años setenta IBM desarrolló los primeros procesadores RISC¹ que evolucionaron hacia una generación mejorada **POWER**² de procesadores superescalares con un juego de instrucciones ampliado, que se utiliza desde los años noventa en los servidores **RISC System 6000 (RS/6000)**, que a principios de esta década pasaron a llamarse **pSeries**.

Power4, la cuarta generación POWER, que salió en el año 2001 y fue mejorado en 2002/3 como Power4+, marcó un antes y un después en muchos sentidos. Desde su salida al mercado, IBM va ganando cuota en el mercado Unix. Servidores con tan sólo 32 procesadores superan el rendimiento de modelos de otros fabricantes con 64 procesadores o más. IBM convirtió Power4 en la familia de procesadores única, no sólo para los servidores pSeries, donde sustituyó al Power3 y al RS64-IV (ambos ya de 64 bits), sino también para la familia iSeries, descrita más adelante. Fue el primer chip dual del mercado Unix e introdujo las particiones lógicas y la capacidad bajo demanda como principales novedades, así como una serie de mejoras arquitectónicas y de RAS³.

Power5, anunciado en el verano de 2004 y descrito en detalle en el siguiente capítulo, continuará la historia de éxito del Power4. Debido a unas modificaciones en su arquitectura, un incremento de la caché y una mayor frecuencia de reloj, proporciona un rendimiento muy superior al Power4, que se ve incrementado aún más por el uso de *Simultaneous Multi-Threading*. Dará soporte a sistemas SMP⁴ de hasta 64 vías. La principal novedad del Power5 es la virtualización. La última generación Power5 de esta línea de servidores se denomina **p5**.

Los pSeries / p5 son una de las siguientes cuatro líneas de servidores de IBM:

- **zSeries**, anteriormente conocidos como S/390 (*mainframe*), son servidores famosos por su fiabilidad, tal como refleja la “z” (*nearly zero downtime*).
- **pSeries**, anteriormente llamados RS/6000, son servidores Unix de alto rendimiento (representado por la “p”, *high performance*).
- **iSeries**, anteriormente denominados AS/400, son servidores para soluciones integradas de negocio (representado por la “i”). La última generación Power5 de esta línea de servidores recibe el nombre i5.
- **xSeries**, anteriormente conocidos como Netfinity, son servidores x86 (Intel y AMD) donde la “x” representa “**X-Architecture**”, que es como IBM denomina el reaprovechamiento de tecnologías de sus otras líneas de servidores en esta familia.

Las primeras tres líneas se basan en tecnologías de procesador propias de IBM, tratándose en el caso de los pSeries y los iSeries de la tecnología POWER. Con el fin de concentrar los esfuerzos de investigación y desarrollo, las cuatro líneas de servidores comparten buena parte de sus tecnologías. Así, por ejemplo, los servidores pSeries e iSeries, y en menor medida también los xSeries, se benefician de las sofisticadas tecnologías mainframe de la gama zSeries. Por otro lado, los servidores pSeries (p5) e iSeries (i5) están convergiendo, de forma que en la actualidad comparten totalmente su plataforma hardware.

CONTENIDO

Introducción	3
Arquitectura	4
Sistemas operativos.....	8
Particiones y virtualización.....	8
Capacidad bajo demanda (CoD).....	12
Reliability, Availability, Serviceability (RAS).....	13
Morse e IBM	14
¿Por qué Morse?.....	14

¹ “Reduced Instruction Set Computing”, con las instrucciones implementados a nivel de transistores y no a nivel de microcódigo, lo cuál acelera su ejecución, haciendo más productivo cada ciclo de reloj.

² “Performance Optimization with Enhanced RISC”.

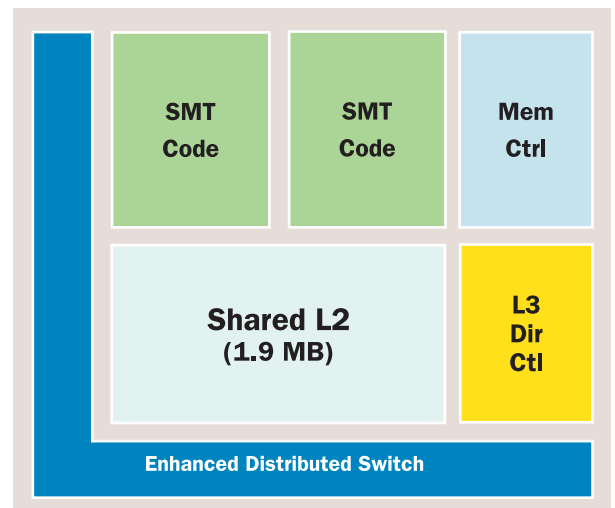
³ Reliability, Availability, Serviceability.

⁴ Symmetric Multi-Processor.

ARQUITECTURA

Diseño del chip Power5

Tras Power4, Power5 es la segunda generación POWER que incorpora dos procesadores RISC de 64 bits en un sólo chip, cada uno con múltiples unidades de ejecución y su propia caché L1 de instrucciones y de datos⁵. Además, incluye la caché L2 compartida entre ambas CPUs, y el directorio y la controladora de la caché L3. La propia caché L3 es demasiado grande para formar parte del chip por lo que se encuentra fuera de él. Además, el chip alberga también las controladoras de memoria y del **fabric switch** distribuido. El *fabric switch* proporciona la interconexión (de un ancho de banda muy elevado) entre los diferentes chips en un SMP y con la memoria y el subsistema de entrada / salida descrito más adelante.



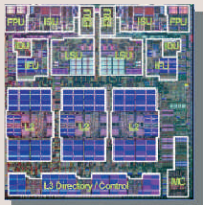
El chip Power5

Power5 proporciona una mayor velocidad que Power4, estando diseñado para soportar frecuencias de reloj de hasta 3 GHz. La disipación del calor, que suele marcar el límite para las frecuencias de reloj, se reduce mediante el novedoso **Dynamic Power Management**.

Además, cuenta con una caché mayor (1,9 MB L2 y 36 MB L3), lo cual reduce la necesidad de acceso a memoria.

POWER5 UN NUEVO ESTÁNDAR

- Segunda generación de chips duales
- 276 millones de transistores (130 nm.)
- Gestión inteligente de consumo energético



Chip p5

- SMT inteligente de 2 vías
- Gestión de capacidad sin impacto en el rendimiento
- Litografía de 130 nm
 - 276 millones de transistores
 - 8 capas de metal

MCM

- SMP de 8 vías que aparece como de 16 vías para el software
- Sólo se requiere DIMMs y E/S
- 95 mm de largo
- Rendimiento equivalente a más de 4 MCMs del p690

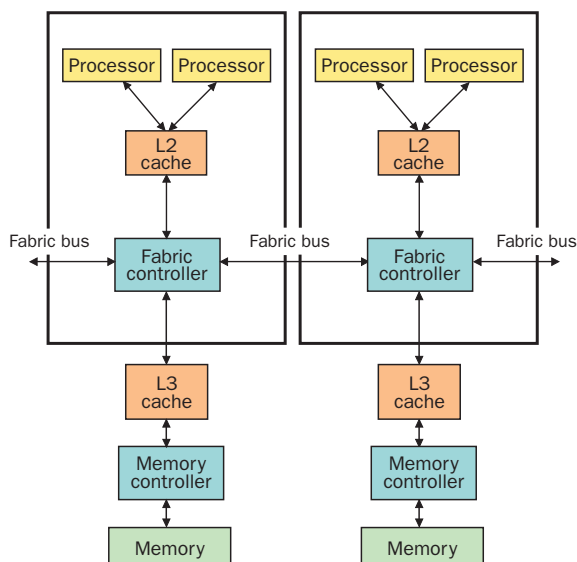


Power5 sigue en buena parte los principios de diseño del Power4, que más allá de centrarse en el diseño del propio chip, revolucionaron la arquitectura de servidores en su conjunto. En este sentido, igual de importante que el rendimiento del propio procesador es el enlace entre los procesadores que gracias al *fabric switch* distribuido reduce las colisiones en las interconexiones evitando su saturación. Además, proporciona una interconexión elástica, cuya velocidad tiene una relación fija con la frecuencia de reloj del procesador, escalando con ella. De esta forma se obtiene una arquitectura equilibrada en la que un intercambio de los procesadores por otros de mayor frecuencia de reloj supone que el sistema en su conjunto incremente su rendimiento sin dejar cuellos de botellas en las interconexiones (*Server-on-a-Chip*).

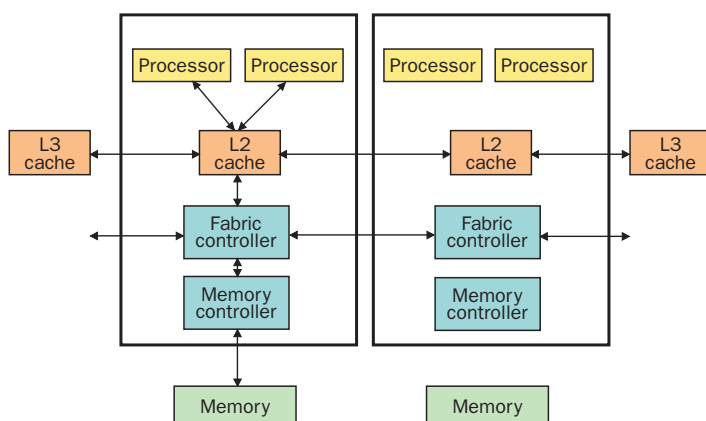
⁵ Para la gama de entrada existen también chips con solo una CPU y la otra deshabilitada (sin posibilidad de una posterior habilitación).

Power5 llevará la escalabilidad SMP, más allá de las 32 vías de la generación Power4, hasta 64 vías. Para evitar que el correspondiente incremento de tráfico sature el *fabric switch* de interconexión, a diferencia del Power4, la caché L3 cuenta con su propio bus y ya no se accede a través del fabric. Ello no sólo reduce la latencia en el acceso a esta caché, sino que también libera al fabric del tráfico correspondiente, tal como se muestra en la siguiente figura.

ESTRUCTURA POWER4



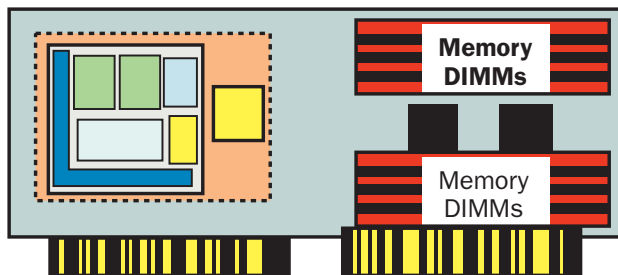
ESTRUCTURA POWER5



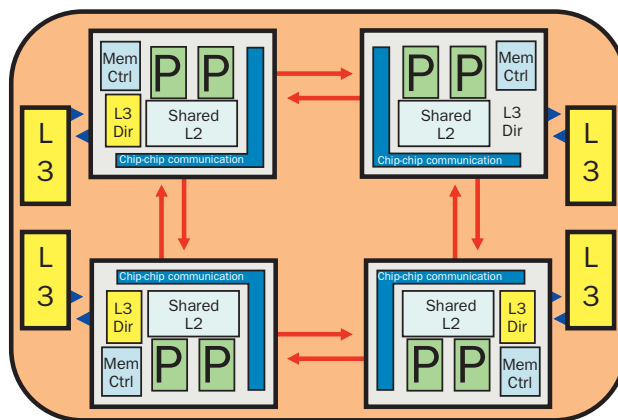
Además, se ha integrado la controladora de memoria, que en los sistemas Power4 requería su propio chip, en el mismo chip Power5, lo cual, unido al hecho de que ya no se accede pasando por la caché L3, reduce drásticamente la latencia en el acceso a memoria. La controladora se conecta a las interfaces de memoria síncronos (SMI), ubicados fuera del chip.

Empaquetamiento

Al igual que en los sistemas Power4, existen dos formas básicas de empaquetamiento de los chips Power5. Ambos se instalan en el denominado **complejo electrónico central (CEC)**.



El **Dual Chip Module (DCM)**, muy eficiente en términos de coste, es la base de los sistemas de gama media-baja. Se compone del chip Power5 (proporcionando por tanto dos vías⁵), la caché L3⁶, dos interfaces de memoria y ocho ranuras de memoria, que permiten uno o dos quads⁷ de RAM.



En el **Multi Chip Module (MCM)** se basan los sistemas más potentes.

Se compone de cuatro chips Power5 (proporcionando por tanto ocho vías), cuatro unidades de caché L3 (que suman un total de 144 MB) y se conecta con las ranuras de memoria, que no forman parte del propio MCM.

Entrada / Salida (I/O)

Los recursos de entrada / salida se conectan a través de la tecnología RIO-2 (segunda generación de la tecnología Remote I/O). En los nuevos equipos, las ranuras PCI son de tecnología PCI-X que permite el intercambio en caliente de las tarjetas, y los discos internos SCSI, también intercambiables en caliente, de tecnología Ultra320. A través de las apropiadas tarjetas PCI, también se pueden conectar discos externos, por ejemplo de tecnología FC ó SCSI.

En los equipos de entrada y de gama media, el propio CEC permite, aparte de los procesadores y de la memoria, unos recursos I/O que en muchos casos son más que suficientes, dando lugar a unas configuraciones económicas y compactas en un sólo cajón. Para casos de altos requerimientos de I/O (que se dan típicamente en sistemas particionados sin virtualización del I/O) existen cajones de expansión de entrada/salida (**I/O drawers**), que añaden más recursos I/O. En los equipos de gama alta, que vienen con su propio bastidor y presentan un diseño modular de varios cajones, el CEC no está preparado para recursos I/O y siempre requiere por lo menos un I/O drawer⁸.

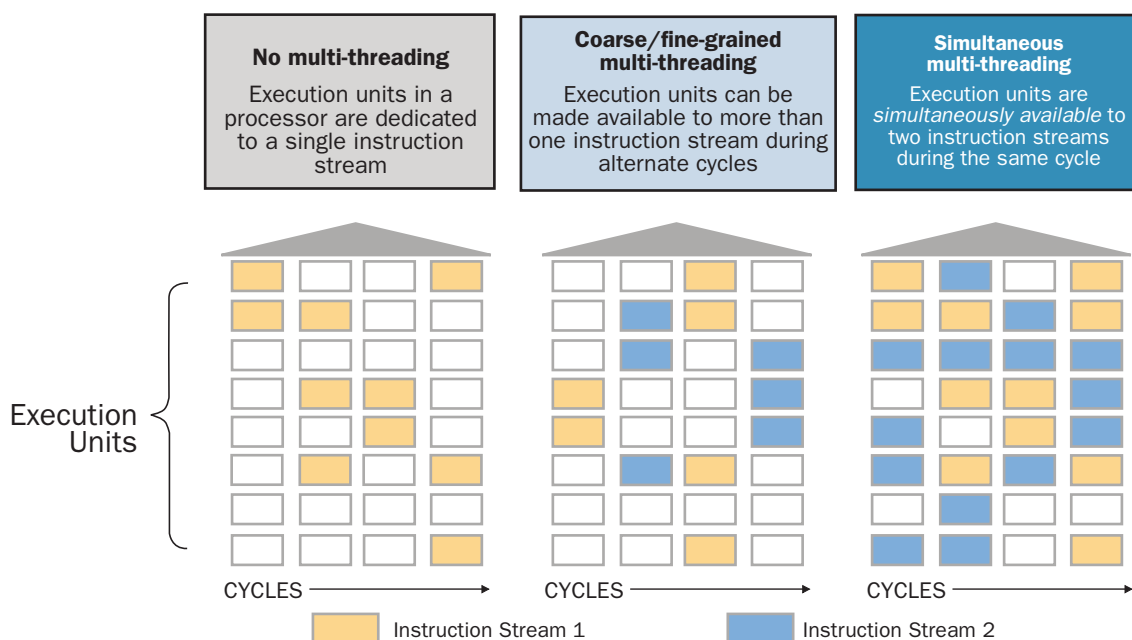
⁶ Los DCMs de sólo una vía, empleados en configuraciones de bajo coste, no llevan caché L3.

⁷ Grupo de cuatro módulos idénticos de memoria. En algunos modelos de entrada se permiten también sólo dos módulos.

⁸ Para que el subsistema I/O no se quede atrás frente a las demás mejoras arquitectónicas de los sistemas, la industria está trabajando en la sustitución de la arquitectura PCI actual por un nuevo estándar **InfiniBand**, en el que la periferia del ordenador aparece como una red de alta velocidad cuyos nodos son los recursos I/O. IBM prevé introducir InfiniBand en los nuevos I/O drawers que probablemente saldrán con la segunda generación de servidores p5 basada en Power5+.

Simultaneous Multi-Threading (SMT)

Una importante novedad incorporada en los procesadores Power5, sólo disponible con AIX 5.3 y determinadas versiones de Linux, es la tecnología SMT (Simultaneous Multi-Threading), que aumenta el aprovechamiento de los procesadores. Como procesadores superescalares, los procesadores Power5 cuentan con múltiples unidades de ejecución, cada una de las cuales soporta una pipeline, que normalmente no está saturada por el thread que está ejecutando. SMT aprovecha estos huecos en la pipeline, para ejecutar simultáneamente instrucciones de dos threads diferentes, y, a diferencia del multi-threading convencional, en un mismo ciclo de reloj, llevando así el paralelismo hasta el nivel de instrucción.



Simultaneous Multi Threading

La figura, que representa como filas las diferentes unidades de ejecución, y como columnas los ciclos de reloj, visualiza de la izquierda a la derecha el funcionamiento de un procesador en modo single threaded (por ejemplo un Power4 ó un Power5 con SMT deshabilitado), un procesador con multi-threading convencional (como un Intel Xeon), y un procesador Power5 con SMT.

Con SMT, un procesador físico (o virtual en el caso de las micro-particiones, descritas más adelante), aparece como dos procesadores **lógicos**. Así, un chip dual Power5 dedicado, parece como un SMP de cuatro vías.

El incremento del rendimiento que SMT supone depende sensiblemente del tipo de aplicación. Aplicaciones intensivas en términos de cálculo y acceso a memoria, como se pueden dar en el cálculo científico optimizado de alto rendimiento, ya no sacan mucho provecho de SMT. En cambio, una carga de trabajo variable, y que no aprovecha toda la capacidad de cálculo del procesador o el ancho de banda a memoria, como es el caso en muchas aplicaciones comerciales, puede verse acelerada por SMT en alrededor de un 30%.

“ En opinión de los analistas, para el final del año 2008, AIX habrá incrementado su cuota de mercado hasta convertirse en el sistema operativo UNIX número uno. ”

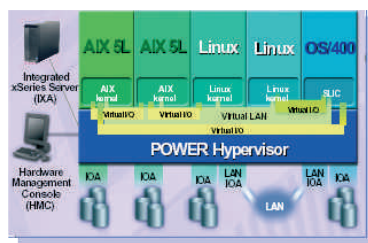
Sistemas operativos

AIX

El sistema operativo Unix de IBM se denomina AIX (Advanced Interactive eXecutable), actualmente en su quinta generación, **AIX 5L**, donde la “L” representa la afinidad con Linux⁹. Actualmente están disponibles tres versiones de AIX 5L.

Los primeros servidores de la generación Power4 salieron con la primera versión de AIX 5L, 5.1. Su principal novedad era el soporte de las particiones lógicas, descritos más adelante. AIX 5.1 no está soportado sobre Power5.

Las siguientes versiones 5.2 ó 5.3 están soportadas tanto sobre Power4 como sobre Power5¹⁰. La versión 5.2 introdujo la capacidad de modificación de las particiones lógicas de manera dinámica. AIX 5.3, que ha salido simultáneamente con Power5, añade además (sobre Power5) la virtualización y SMT como principales novedades.



Linux

IBM apuesta fuertemente por Linux y no sólo soporta determinadas versiones en xSeries, es decir sobre la arquitectura x86, la plataforma habitual de Linux, sino también en particiones lógicas de zSeries e iSeries, y en sus servidores Unix, tanto en particiones lógicas como nativo¹¹.

i5/OS

También está soportado (con ciertas restricciones) el sistema operativo i5/OS, propio de los servidores iSeries/i5, en particiones de determinadas configuraciones de modelos p5 de gama media y alta.

Particiones y virtualización

Dos importantes hitos en la consolidación de servidores, y que conjuntamente con la estrategia on Demand, descrita más adelante, pueden contribuir a incrementar significativamente el aprovechamiento de los recursos, son las particiones lógicas, introducidas con Power4, y la virtualización, opcionalmente disponible con todos los modelos Power5. IBM tiene una experiencia de décadas en este campo en el mundo mainframe, gracias a la cual ahora es pionera en el mercado Unix en estas tecnologías.

Particiones lógicas (LPAR)

El particionado lógico permite la existencia de varios servidores lógicos en una máquina, y de manera más flexible que el particionado físico, en el que los límites de las particiones tienen que coincidir con los límites entre los bloques de fabricación del sistema.

⁹ Supone que un código fuente Linux puede ser recompilado en AIX 5L.

¹⁰ Con el nivel de mantenimiento adecuado de AIX 5.2 para Power5.

¹¹ No todas las funcionalidades descritas en este documento están soportadas con Linux, al tratarse de un sistema operativo estándar no específicamente desarrollado para POWER.

En el particionado físico, todos los recursos (procesadores, memoria e I/O) de un bloque tienen que pertenecer a la misma partición, lo cual suele causar un alto grado de desaprovechamiento de los recursos.

En cambio, en las particiones lógicas los procesadores (incluso los dos de un mismo chip) y las controladoras de I/O (y por tanto generalmente las tarjetas PCI) pueden asignarse de manera independiente a diferentes particiones. El mínimo de memoria de una partición es de 128 MB en el caso de Power5 y 1 GB en el caso de Power4, que puede ser incrementado en intervalos de 16 MB en el caso de Power5 y de 256 MB en caso de Power4.

Todos los modelos Power5 (y también los Power4 salvo algunos modelos de entrada), soportan las particiones lógicas, sin coste adicional. Las particiones están aisladas entre sí y cada una cuenta con su propia imagen de sistema operativo, que es independiente de las demás, pudiendo coexistir diferentes versiones de AIX y Linux en una sola máquina física. Se instalan sobre una capa de firmware denominada **Hypervisor** que controla la distribución de los recursos sobre las particiones. Se gestiona desde una **Consola de Gestión de Hardware (HMC)**.

Los recursos se pueden reasignar entre las diferentes particiones. En el caso de particiones con AIX 5.2 ó 5.3, esta reasignación se puede llevar a cabo de manera dinámica, por lo que se denominan **particiones dinámicas (DLPAR)**¹².

Consola de Gestión de Hardware (HMC)

La Hardware Management Console (HMC) es obligatoria para la gestión de todos los sistemas particionados (y en los sistemas de gama alta aunque no estén particionados), aunque no tiene que estar activa para que funcione el sistema. Es un PC, con una interfaz gráfica de usuario (GUI) basada en Java sobre Linux, disponible en formato pedestal y bastidor (1 U).

Se pueden gestionar varios servidores físicos desde la misma HMC, aunque no se pueden gestionar sistemas Power4 y Power5 desde la misma HMC. La conexión entre los sistemas y la HMC se lleva a cabo por Ethernet en el caso de sistemas Power5 y por conexión serie en el caso de sistemas Power4. Además, todas las particiones y la HMC que les administra deben conectarse a una misma red Ethernet.

La HMC lleva también a cabo otras tareas, como por ejemplo la gestión de los recursos inactivos de CoD, upgrades de microcódigo, supervisión del estado del sistema, etc.

Virtualización

La virtualización es la gran novedad de Power5 e incrementa más aún la flexibilidad de las LPARs y el grado de aprovechamiento de los recursos. Oculta los dispositivos físicos bajo una capa de abstracción sobre la que se crean los dispositivos virtuales que son visibles para el software. De esta manera, los recursos físicos se pueden compartir entre las particiones, salvo la memoria, que no se virtualiza. Ello no implica una restricción en la práctica, ya que, como se vio anteriormente, con Power5 puede ser asignada de manera prácticamente continua a las diferentes particiones.

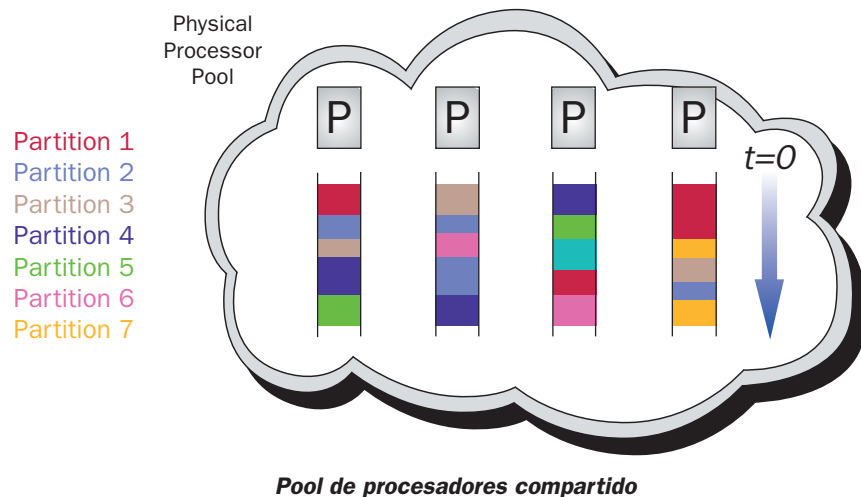
La virtualización requiere la adquisición de la **Advanced Power Virtualization (APV)** incluida por defecto en los modelos de gama alta y disponible opcionalmente en los demás, que proporciona soporte para micro-particiones (con procesadores virtuales), Virtual I/O Server (compuesto por Shared Ethernet Adapter y Virtual SCSI Server) y el Partition Load Manager.

¹² AIX 5.1, sólo soportado sobre Power4, no permite DLPAR, es decir una partición AIX 5.1 involucrada en una reasignación, tiene que reiniciarse para que tenga efecto. Las particiones con Linux son parcialmente dinámicas.

“ ¿Qué es virtualización?
Es la agrupación de recursos TI de forma que se oculte la naturaleza física y los límites de esos recursos a los usuarios. ”

Micro-particiones: virtualización de procesadores

La APV permite la creación de un pool de procesadores compartidos, sobre el que se pueden definir procesadores virtuales entre los que se reparten los ciclos de reloj de los procesadores físicos, tal como muestra la siguiente figura.



A todas las particiones con AIX 5.3 o la apropiada versión de Linux se les pueden asignar procesadores virtuales de este pool con un mínimo equivalente a la décima parte de un procesador físico, por lo que el número máximo de particiones soportado en un sistema, que sin APV coincide generalmente con el número de procesadores físicos, se multiplica con APV por diez, aunque con un límite de 254 particiones por sistema. A partir de este mínimo se puede incrementar la capacidad de procesadores en intervalos equivalentes a la centésima parte de un procesador físico, lo cual permite aumentar mucho el grado de aprovechamiento de los procesadores.

El pool también puede contener procesadores no asignados a ninguna partición, a los que en caso de necesidad las particiones autorizadas para ello pueden acceder. (Véase también Reserve Capacity on Demand, descrito más adelante.)

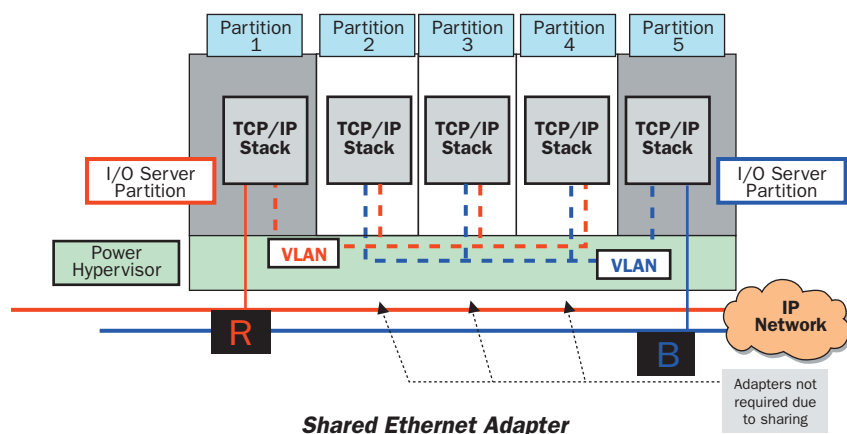
No todos los procesadores del sistema se tienen que asignar al *pool* compartido, es decir pueden coexistir procesadores virtuales del *pool* compartido, y procesadores físicos, dedicados a una sola partición. De hecho, una partición con AIX 5.2 sólo permite procesadores dedicados. Las demás particiones pueden funcionar o bien con procesadores virtuales del pool compartido, o bien con procesadores dedicados (pero no con una mezcla de ambas).

Entrada / salida virtual (Virtual I/O Server)

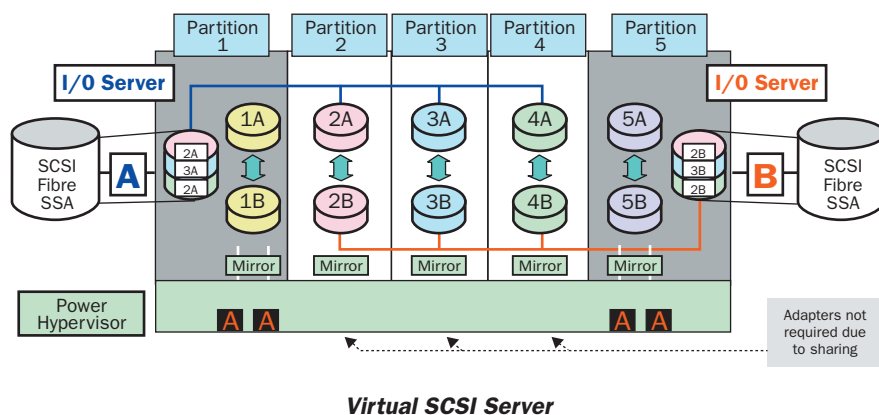
APV permite también la consolidación de los recursos Ethernet y disco (interno y externo) mediante la creación de una o varias particiones con AIX 5.3 dedicada a la funcionalidad de **Virtual I/O Server**. Otras particiones, con AIX 5.3 o una apropiada versión de Linux, pueden prescindir de recursos Ethernet y disco propios, haciendo de cliente de una partición Virtual I/O Server (o de varias en redundancia), que sirven estos recursos a sus clientes.

Los sistemas Power5 soportan (sin necesidad de adquirir APV) la tecnología Virtual Ethernet que permite la interconexión entre sus particiones (con AIX 5.3 o la adecuada versión de Linux) a través de memoria, de tal forma que aparecen al sistema operativo como interconexiones de Ethernet de alta velocidad. La tecnología VLAN garantiza que cada partición sólo tenga acceso a los datos destinados a ella.

Cuando se establecen particiones Virtual I/O Server con APV, éstas se comunican con sus clientes por Virtual Ethernet. El Virtual I/O Server proporciona un adaptador Ethernet compartido (**Shared Ethernet Adapter**) que enruta el tráfico de red entre los adaptadores Ethernet virtuales con los que se comunica con sus clientes, y la red Ethernet física, a la que se conecta con un adaptador Ethernet físico (o con varios, trunking). La siguiente figura muestra un sistema particionado con dos particiones Virtual I/O Server en redundancia, que comparten sus recursos Ethernet con las demás particiones, que son sus clientes y no cuentan con adaptadores Ethernet físicos.



La otra funcionalidad de las particiones **Virtual I/O Server**, es la de **Virtual SCSI Server**, que consiste en la creación de unidades lógicas de disco que se exportan a las particiones clientes. Los discos físicos subyacentes del Virtual I/O Server pueden ser discos internos o externos (SCSI y FC). Por tanto no sólo proporcionan la consolidación de discos internos sino también de adaptadores PCI (SCSI ó FC). La siguiente figura muestra un sistema particionado con dos particiones Virtual I/O Server en redundancia, que comparten sus discos con las demás particiones, que son sus clientes y que no tienen acceso propio a discos. Cada cliente crea un RAID1 (mirror) entre las unidades lógicas que le proporcionan los dos servidores virtuales.



En un sistema e incluso en una misma partición pueden coexistir recursos I/O físicos dedicados, y recursos virtualizados proporcionados por los I/O Servers. Las particiones de AIX 5.2 sólo pueden funcionar con recursos dedicados, y según el caso puede ser recomendable también en particiones AIX 5.3 ó Linux, por ejemplo por su alto volumen de tráfico I/O.

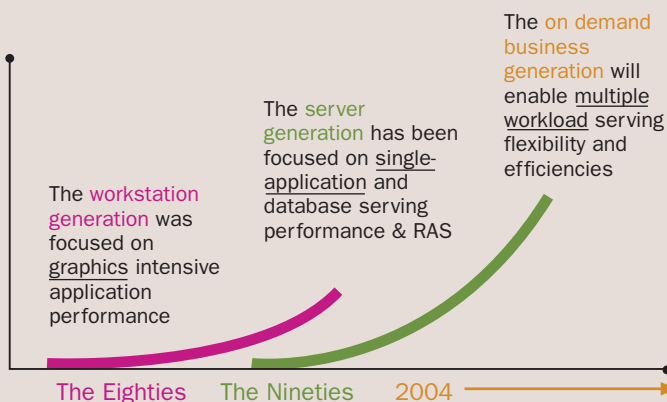
Partition Load Management

AIX incluye por defecto el **Workload Manager (WLM)**, que redistribuye automáticamente los recursos (procesadores, memoria e I/O) entre distintos procesos en un servidor o una partición, basado en una política definida por el administrador. No permite una redistribución entre diferentes particiones.

El nuevo **Partition Load Manager (PLM)**, proporcionado con APV, complementa esta herramienta, permitiendo una redistribución automática de procesadores y memoria entre los procesos de un grupo de diferentes particiones AIX de un sistema Power5¹³. Esta reasignación, basada en una política definida por el administrador, requiere la comunicación con la HMC. El PLM se puede ejecutar tanto en una partición del mismo sistema cuyos recursos gestiona, como en otro sistema p5 con AIX, siempre y cuando tiene una conexión Ethernet tanto con las particiones gestionadas como con la HMC.

ANNOUNCING THE 3RD GENERATION OF UNIX® COMPUTING...

ON DEMAND BUSINESS



Capacidad bajo demanda (CoD)

Los servidores actuales ofrecen un amplio rango de configuraciones, que permiten ajustarlas a las necesidades concretas del cliente, y modificarlas a lo largo del ciclo de vida si es necesario. Para facilitar este proceso, IBM ofrece dentro de su estrategia **on Demand** para casi todos los modelos Power5 y para varios modelos Power4 la posibilidad de disponer de recursos de procesadores y memoria inactivos (a un precio muy inferior que los recursos activos), cuya activación puede ser adquirida en cualquier momento, proporcionando así una ampliación inmediata de los recursos, sin necesidad de esperar a un suministro físico. Además, supone en algunos modelos de gama media-alta una granularidad más fina (1 procesador, 1 GB en el caso de Power5; 2 procesadores, 4 GB en el caso de los Power4) que la de los recursos físicamente instalados. Las activaciones de los procesadores y la memoria son operaciones independientes. Se denomina **Capacity on Demand (CoD)** y existen diferentes formas para ello:

- **Capacity Upgrade on Demand** supone la activación permanente de procesadores y/o memoria inactivos mediante la adquisición de unas claves.
- **Trial Capacity on Demand:** El usuario dispone para todos los recursos inactivos de 30 días de uso gratuito de forma única. Eso le permite una activación inmediata de los recursos, sin esperar la adquisición definitiva de las claves mencionadas en el apartado anterior. Durante estos 30 días se puede tramitar la compra de las claves, o bien decidir no llevarla a cabo. Así se evitan por un lado los tiempos de espera, y por el otro se puede probar el rendimiento con los nuevos recursos antes de decidir sobre la adquisición de su activación definitiva.
- **On/Off Capacity on Demand** supone un pago por uso que determinados clientes pueden contratar con IBM para sus modelos p5. Consiste en el uso discontinuo con una granularidad de un día de los recursos inactivos, tanto procesadores como memoria, que se reporta mensualmente a IBM para su facturación. En sistemas Power4 esta característica sólo está disponible para procesadores y requiere una adquisición por adelantado de su uso discontinuo.

¹³ Todas las particiones de un grupo tienen que tener o bien procesadores virtuales o bien procesadores dedicados; en un mismo grupo no está permitida la mezcla de ambos tipos.

- **Reserve Capacity on Demand** está sólo disponible con sistemas p5 con micro-particiones. Supone la adquisición por adelantado para el uso discontinuo de procesadores que forman parte del pool compartido, y a los que todas las particiones autorizadas para ello pueden acceder en caso de necesidad, tal como se describió en el apartado de las microparticiones. Una vez consumida la capacidad adquirida, los procesadores desaparecen del pool.
- **Dynamic processor sparing:** Los procesadores inactivos de un sistema se activan sin coste en caso de fallo de un procesador activo. Como se describirá a continuación, en la mayoría de los casos las características RAS desactivan un procesador defectuoso antes de que su fallo definitivo pueda causar una caída del sistema y el sistema sigue funcionando con los demás procesadores. En este caso, si existe un procesador inactivo, éste sustituirá al procesador desactivado de manera dinámica¹⁴.

Reliability, Availability, Serviceability (RAS)

Las tecnologías RAS están encaminadas a elevar la **fiabilidad**, la **disponibilidad** y las **facilidades de servicio** de los sistemas y son otro de los puntos fuertes de los servidores Unix de IBM, ya que incorporan muchas características del mundo mainframe.

La fiabilidad empieza por el diseño y la fabricación de los componentes, que reduce su tasa de fallos. Además, la redundancia de componentes evita que un fallo de uno de ellos cause la caída del sistema. Un procesador de servicio incorporado supervisa los componentes. Mediante una tecnología llamada First Failure Data Capture, única de IBM e introducida con Power4, este procesador es capaz de diagnosticar los fallos en tiempo real, lo cual evita el laborioso proceso de reproducción de fallos (sobre todo en el caso de fallos intermitentes).

Los fallos recuperables se resuelven sin necesidad de intervención externa, por tecnologías como por ejemplo ECC chipkill para la memoria.

Basado en los síntomas que les suelen anticipar, la mayoría de los pocos fallos irre recuperables, que causarían una caída del sistema, se pueden predecir. Gracias a ello, un componente crítico susceptible a fallar en breve de manera irre recuperable es desactivado y se utilizan las redundancias internas para mantener el sistema funcionando. Automáticamente se emiten las correspondientes llamadas de servicio. La posterior sustitución de los componentes defectuosos se puede realizar en caso de muchos componentes sin parada de servicio (hot swap). Pero incluso en el caso de los componentes cuya sustitución requiere una parada, como ésta se puede llevar a cabo de manera planificada en un momento conveniente, siempre será preferible a la caída inesperada en pleno servicio.

Para cubrirse también frente a la posibilidad de una caída del sistema entero, o para servicios que son tan críticos que ni permiten una parada planificada, dos o más servidores se pueden configurar en un cluster de alta disponibilidad, por ejemplo mediante el software HACMP de IBM para AIX. Este software soporta opcionalmente también distancias ilimitadas entre los nodos, lo cual proporciona alta disponibilidad incluso frente a la destrucción física de un CPD.

¹⁴ La sustitución es dinámica en sistemas Power5 y en particiones dinámicas de Power4; en otros entornos Power4 requiere un rearranque del sistema o de la partición.

MORSE E IBM

Nuestra oferta completa de servicios, software e infraestructura nos permite proporcionar sistemas en distintos entornos empresariales. Con consultores especializados en el diseño, implementación y gestión de sistemas de IBM, planificamos las estrategias tecnológicas mediante el análisis del entorno y requerimientos con anterioridad a la propuesta técnica.

Desde 1998, nuestra red europea de Centros Empresariales de Cálculo ha aportado una serie de entornos controlados en los cuales se pueden realizar pruebas sobre escalabilidad e interacción entre productos en entornos de plataformas heterogéneos. De esta forma se reducen los riesgos asociados al despliegue de soluciones complejas. Completamente equipados con hardware de gama alta, podemos realizar pruebas y demostraciones de la última tecnología de IBM (incluyendo iSeries, pSeries, zSeries, xSeries, almacenamiento y software), llevamos a cabo pilotos sobre servidores y almacenamiento y también desarrollamos pruebas de "porting" de aplicaciones, pruebas de rendimiento y cursos de formación.

¿Por qué Morse?

Morse es un integrador tecnológico. Ayudamos a nuestros clientes a conseguir más por menos mediante unas tecnologías de la información adaptadas a la empresa, definiendo e implementado complejas soluciones integradas y que se ajusten a las necesidades de negocio. Para conseguir esto, utilizamos nuestros propios recursos y servicios, la estrecha relación que tenemos con los grandes fabricantes de tecnología, entre los que se encuentra IBM, y los 20 años de experiencia que tenemos en entornos críticos. Con más de 1.200 empleados y oficinas en el Reino Unido, Alemania, Francia, España e Irlanda, contamos con una fuerte presencia en Europa.

Para afrontar proyectos complejos que cubran las necesidades de nuestros clientes es fundamental elegir el compañero de viaje adecuado. Morse no basa su filosofía en simples componentes tecnológicos sino que también aporta:

- Un importante equipo humano con muchos años de experiencia y un gran nivel de conocimiento.
- Una metodología propia especializada en proyectos de integración tecnológica basada en un seguimiento y un control escrupuloso de los mismos.
- Una actitud de compromiso y seriedad en el establecimiento de expectativas y en la implantación de los proyectos.
- Una visión dinámica de la tecnología y de su evolución diaria para ofrecer capacidad de asesoramiento en las múltiples tendencias emergentes en cada momento.

es.morse.com/ibm
91 766 90 69

Morse Integración de Sistemas

Avda. Diagonal, 569, 2ª Plta
Edif. L'Illa Diagonal
08029 Barcelona

Tel.: 93 363 25 90
Fax: 93 321 76 64

Avda. de Burgos, 17, Planta 11
Edif. Triada A
28036 Madrid

Tel.: 91 766 90 69
Fax: 91 766 90 63

Creemos en las TI adaptadas a la empresa, ayudando a nuestros clientes a conseguir más por menos. A este recorrido lo llamamos A2e.

Definimos y proporcionamos soluciones tecnológicas integradas que se adaptan a las necesidades de negocio de nuestros clientes. Colaborando estrechamente con IBM, ofrecemos la mejor tecnología e integración on demand.

Servidores pSeries - Power4

Los servidores Power4+¹ proporcionan una plataforma hardware a aquellos clientes que aún necesitan entornos AIX 5.1, una versión del sistema operativo que no está soportado sobre Power5.

Tecnología DCM

Los modelos de entrada **pSeries 615** y **630**, así como el **pSeries 650** de gama media se basan en los *Dual Chip Modules*². El 630 y el 650 soportan opcionalmente el I/O drawer D20, que ocupa 4 U de espacio vertical en un bastidor estándar de 19 pulgadas. Proporciona 7 ranuras PCI-X y opcionalmente 2 backplanes SCSI, cada uno con 6 bahías de disco, que según las necesidades pueden ser controlados por la misma tarjeta SCSI o por 2 tarjetas independientes³. Para el 650 existe como alternativa al D20 el modelo D10, que también requiere 4 U de altura en un bastidor estándar, pero sólo ocupa media anchura, permitiendo por lo tanto 2 unidades del D10 aparejadas en el espacio de 4 U. No permite discos internos, sino que sólo proporciona 6 ranuras PCI-X.

	pSeries 615	pSeries 630	pSeries 650
Formato	Pedestal o bastidor estándar de 19" (4 U)	Pedestal o bastidor estándar de 19" (4 U)	Bastidor estándar de 19" (8 U)
Nº de DCMs	1 ⁴ (de 1 ⁵ ó 2 CPUs)	1 ó 2 (de 1 ⁵ ó 2 CPUs)	1 - 4 (de 2 CPUs)
CPUs Power4	1 ó 2 de 1,2 GHz, ó 2 de 1,45 GHz	1,2 ó 4 de 1,2 ó 1,45 GHz	2, 4, 6 ó 8 de 1,2 ó 1,45 GHz
Caché L2 por DCM	1,5 MB	1,5 MB	1,5 MB
Caché L3 por DCM	8 MB	8 MB	8 MB (1,2 GHz) ó 32 MB (1,45 GHz)
Memoria (max.) ⁶	16 GB (1-16 GB)	32 GB (1-16 GB por DCM)	64 GB (1-16 GB por DCM)
Puertos Ethernet integrados ⁷	1x10/100Mbps, 1x10/100/1000Mbps	2x10/100Mbps	1x10/100Mbps
Puertos externos SCSI integrados ⁷	-	1xUltra320	1xUltra320
Ranuras I/O (CEC)	6	6	7
Bahías disco (CEC)	4 ó 8 (1 ó 2 backplanes de 4)	4 (en un backplane)	4 (opcionalmente en 2 backplanes de 2)
I/O Drawers opcionales (max.)	-	2 D20 (sólo en el modelo bastidor)	8 D20 ó D10
Ranuras I/O (max.)	6	20	63
Discos (max.) ⁸	8	28	100
LPARs (max.)	-	4 (sólo en modelo bastidor)	8
CoD	-	-	Sólo en caso de CPUs a 1,45 GHz
rPerf máximo ⁹	4,41	8,69	18,67

La **IntelliStation Power 275** es básicamente un pSeries 615 en formato pedestal con una tarjeta gráfica incorporada, la cual es opcional en otros modelos.

Servidores pSeries - Power4

Tecnología MCM

El modelo de gama alta **pSeries 690**¹⁰, así como el **pSeries 670** (básicamente un 690 limitado en escalabilidad, para posicionar un modelo entre el 650 y el 690) se basan en los Multi Chip Modules, al igual que el modelo ultradenso pSeries 655 (diseñado para concentrar una gran capacidad de cálculo en un espacio muy reducido, normalmente gestionado en un cluster de alto rendimiento). El pSeries 670 y el pSeries 690 requieren obligatoriamente una HMC para su gestión.

Para estos equipos está disponible el I/O drawer **61D**, que ocupa 4 U en los bastidores específicos de 24 pulgadas de los equipos MCM. Cada cajón se divide en 2 bloques independientes, cada uno de los cuales proporciona 10 ranuras PCI-X y 2 backplanes de 4 bahías de discos que dependen de controladoras SCSI integradas independientes. Por tanto, cada cajón cuenta con 20 ranuras PCI-X y 16 bahías de discos. Para el 670 y el 690 por lo menos un I/O drawer es obligatorio, ya que no cuentan con recursos I/O en su CEC.

	pSeries 615	pSeries 630	pSeries 650
Formato	Bastidor de 24" (2 unidades aparejados caben en 4 U)	Ocupa 1 bastidor propio entero de 24" y 42 U(8 U)	Según el nº de I/O drawers ocupa 1 ó 2 bastidores propios enteros de 24" y 42 U
Nº de MCMs	1 (de 4 ⁵ ó 8 CPUs)	1 (de 4 ⁵ ó 8 CPUs) ó 2 (de 8 CPUs)	1 - 4 (de 8 CPUs)
CPUs Power4	4 ó 8 de 1,7 GHz	4, 8 ó 16 de 1,5 GHz	8, 16, 24 ó 32 de 1,5 ó 1,7 ó 1,9 GHz
Caché L2 por MCM	4 x 1,5 MB	4 x 1,5 MB	4 x 1,5 MB
Caché L3 por MCM	128 MB	128 MB	128 MB
Memoria (max.)	64 GB (mín. 4 GB) ¹¹	256 GB (mín. 4 GB) ¹²	1024 GB (mín. 8 GB) ¹²
Puertos Ethernet int.	2x10/100 Mbps	-	-
Puertos externos SCSI int.	1 x Ultra 320	-	-
Ranuras I/O (CEC)	3	-	-
Bahías disco (CE C)	2	-	-
I/O Drawers 61D (max.)	1 (opcional)	1 - 3 (2 con 1 MCM)	1 - 8 (max. 2 por MCM)
Ranuras I/O (max.)	23	60 ¹³	160 ¹³
Discos (max.) ⁸	18	48	128
LPARs (max.)	4	16	32
CoD	-	Sí	Sí
rPerf máximo ⁹	21,87	46,79	104,17

¹ Anteriormente, estos servidores (con excepción de los pSeries 615 y 650, que salieron al mercado ya con Power4+) formaron la gama Power4.

² IBM se refiere en su documentación a los DCMs del 630 y 650 como *Single Chip Module* (SCM).

³ Cada tarjeta SCSI ocupa una ranura PCI-X.

⁴ Integrado en la placa base y no intercambiable.

⁵ En los DCMs de 1 CPU y los MCMs de 4 CPUs, cada chip tiene sólo 1 CPU.

⁶ Las 8 ranuras de memoria de cada DCM se ocupan con 1 ó 2 quads de 4 módulos de 256, 512, 1024 ó 2048 MB cada uno.

⁷ Más puertos disponibles mediante tarjetas PCI.

⁸ Los discos SCI disponibles son de 36,4 y 73,4 GB a 10.000 y 15.000 revoluciones / minuto (RPM), así como de 146,8 GB a 10.000 RPM.

⁹ Los valores rPerf son benchmarks internos de IBM para comparar el rendimiento de sus distintos modelos Unix en función de los procesadores y la caché, suponiendo la memoria máxima para la correspondiente configuración. No tienen en cuenta el I/O y deben entenderse como meramente orientativos, siendo el rendimiento real altamente dependiente de las aplicaciones ejecutadas.

¹⁰ Desde el anuncio de los procesadores Power4+ a 1,9 GHz en febrero del 2004, el benchmark TPC-C para sistemas SMP individuales (es decir no en cluste del pSeries 690 con 32 de estos procesadores no ha sido alcanzado por los modelos de la competencia con 64 y más procesadores, lo cuál es más notable aún teniendo en cuenta que se trata de la tecnología antecesora del Power5. http://www.tpc.org/tpcc/results/tpcc_perf_results.asp?resulttype=noncluster

¹¹ El 655 lleva 4 ranuras de memoria para los que están disponible módulos de 4, 8 y 16 GB.

¹² Cada MCM habilita 2 ranuras de memoria para los que están disponible módulos de 4, 8, 16, 32, 64 y (sólo en el 690) 128 GB.

¹³ Una de ellas se ocupará con una tarjeta SCSI para la conexión del cajón de soporte de almacenamiento (media drawer).

Servidores pSeries - Power5

Todos los modelos de la nueva gama Power5 soportan la expansión del I/O, son particionables y permiten la virtualización, en el caso de los modelos de tecnología DCM de forma opcional.

Tecnología DCM

Los modelos de entrada **p5-520** y **550** así como el **p5-570** de gama media se basan en los *Dual Chip Modules*. Sus configuraciones con la frecuencia de reloj más baja, 1,5 GHz, se denominan también **Express** y como se verá a continuación, aplican algunas particularidades para ellos, como el no soporte de CoD.

Los modelos 520 y 550 son los sucesores de los pSeries 615 y 630 respectivamente. Aparte de las novedades genéricas que aporta el Power5, cabe destacar la mayor escalabilidad del I/O así como el soporte CoD para procesadores en el p5-550 como novedades en la gama de entrada.

El modelo 570 cuenta con un novedoso diseño modular, inexistente en la familia Power4, que le permite llegar a unas prestaciones hasta ahora reservadas a la gama alta¹⁴. Su CEC se compone de entre 1 y 4 bloques de sistema, que se interconectan de tal manera que forman un sólo SMP con memoria plana. La configuración Express está limitada a 2 bloques, con lo que se posiciona entre el 550 y el 570 ilimitado, con un precio muy inferior a este último, sucediendo al pSeries 650. Con AIX 5.3, los modelos duplican aproximadamente el rendimiento de sus antecesores con el mismo número de procesadores Power4.

Todos los modelos Power5 soportan opcionalmente el I/O drawer **D20**, ya descrito en el capítulo de los equipos Power4. Para el 570 existe como alternativa al D20 el modelo **D11**, que sólo se distingue de su antecesor D10, descrito en el capítulo de los equipos Power4, en que 1 de sus 6 ranuras es de 5 V para tarjetas antiguas¹⁵.

	p5-520	p5-550	p5-570
Formato	Pedestal o bastidor estándar de 19" (4 U)	Pedestal o bastidor estándar de 19" (4 U)	1-4 (con 1,5 GHz sólo 2) bloques de 4 U en bastidor estándar de 19"
Nº de DCMs	1 ⁴ (de 1 ⁵ ó 2 CPUs)	1 (de 1 ⁵ ó 2 CPUs) ó 2 (de 2 CPUs)	2 (de 2 vías) por bloque (con 1 bloque puede llevar también 1 DCM)
CPUs Power5	1 ó 2 de 1,5 GHz, ó 2 de 1,65 GHz	1 ó 2 de 1,5 GHz ó 2 ó 4 de 1,65 GHz	2, 4 ó 8 de 1,5 GHz ó 2, 4, 8, 12 ó 16 de 1,65 ó 1,9 GHz
Caché L2 / DCM	1,9 MB	1,9 MB	1,9 MB
Caché L3 / DCM	36 MB (para 1 vía no lleva)	36 MB (para 1 vía no lleva)	36 MB
Memoria (max.)	32 GB (1-32 GB; para 1,5 GHz también 0,5 GB) ¹⁶	64 GB (1-32 GB por DCM; para 1,5 GHz también 0,5 GB) ¹⁶	512 GB (2-64 GB por DCM) ¹⁷
Puertos Ethernet integrados ⁷	2x10/100/1000 Mbps	2x10/100/1000 Mbps	2x10/100/1000 Mbps por bloque
Puertos externos SCSI integrados ⁷	-	-	-
Ranuras I/O (CEC)	6	5	6 por bloque
Bahías disco (CEC)	4 ó 8 (1 ó 2 backplanes de 4)	4 ó 8 (1 ó 2 backplanes de 4)	6 por bloque
I/O Drawers (opcionales)	4 D20	8 D20 (4 por DCM)	20 D20 ó D11 (en el primer bloque 4 por DCM, después 4 por bloque)
Ran. I/O (max.)	34	60 ¹⁸	163 ¹⁸
Discos (max.) ⁸	56	104	264
LPARs (max.)	20 (2 sin APV)	40 (4 sin APV)	160 (16 sin APV)
CoD	-	Sólo para procesadores (salvo para 1,5 GHz)	Sí (salvo para 1,5 GHz)
rPerf máximo ¹⁹	9,86	19,66	77,45

El modelo **OpenPower 720**, que se comercializa dentro de la familia xSeries, es básicamente un p5-550 más económico, que sólo soporta el sistema operativo Linux y no permite CoD.

Servidores pSeries - Power5

Tecnología MCM

Los modelos de gama alta se basan en placas de 2 *Multi Chip Modules* y 16 ranuras de memoria. El **p5-590** permite hasta 2 de estas placas y el **595** hasta 4. El 590 es básicamente un 595 limitado en escalabilidad, para posicionar un modelo entre el 595 y el 570. En términos de escalabilidad se parece al pSeries 690 y a pesar de una menor frecuencia de reloj, le supera en un 50% en el rendimiento máximo. La máxima frecuencia de reloj está limitada al 595, que eleva el techo del rendimiento de la familia, anteriormente sostenido por el 690, en un factor 3²⁰. Ambos modelos requieren obligatoriamente una HMC para su gestión.

Ambos modelos requieren por lo menos un *I/O drawer*, ya que no cuentan con recursos I/O en su CEC. Se trata del mismo modelo **61D**, descrito en el capítulo correspondiente a los equipos Power4, aunque con p5 ya no lleva este nombre, sino se pide como una feature del propio servidor. Como novedad en p5 existe también una versión con sólo 8 bahías de disco de este I/O drawer.

	p5-590	p5-595
Formato	Ocupa 1 bastidor propio entero de 24" y 42U	Según el nº de <i>I/O drawers</i> ocupa 1 ó 2 bastidores propios enteros de 24" y 42U
Nº de MCMs	2 ó 4 (de 8 CPUs)	2, 4, 6 ó 8 (de 8 CPUs)
CPUs Power4	16 ²¹ ó 32 de 1,65GHz	16, 32, 48 ó 64 de 1,65 ó 1,9 GHz
Caché L2 / por MCM	4 x 1,9 MB	4 x 1,9 MB
Caché L3 / por MCM	144 MB	144 MB
Memoria (max.)	1024 GB (8-512 GB por 2 MCMs) ²²	2048 GB (8-512 GB por 2 MCMs) ²²
Puertos Ethernet integrados ⁷	-	-
Puertos externos SCSI integrados ⁷	-	-
Ranuras I/O (CEC)	-	-
Bahías disco (CEC)	-	-
I/O Drawers 61D (max.)	8 (6 con 2 MCMs)	12 (6 con 2 MCMs)
Ranuras I/O (max.)	160	240
Discos (max.) ²³	128	192
LPARs (max.)	254 (32 con procesadores dedicados)	254 (64 con procesadores dedicados)
CoD	Sí	Sí
rPerf máximo ¹⁹	151,72	306,21

En breve saldrá también un modelo de cálculo intensivo, que sucederá al pSeries 655.

¹⁴ De hecho, en el ranking mencionado en la nota 10, el p5-570 es el único modelo de 16 CPUs entre los primeros 10, siendo un HP Integrity Superdome con 64 CPUs Itanium2 la única máquina de la competencia que lo supera.

¹⁵ Los I/O drawer D10 actualmente instalados con un pSeries 650, puede ser aprovechado también para un p5-570, pero ya no está disponible para pedidos nuevos con un p5-570.

¹⁶ Los modelos p5 de gama entrada duplican la memoria máxima por DCM frente los modelos DCM Power4, gracias a los nuevos módulos de 4 GB que se añaden a los módulos de 256, 512, 1024 y 2048 MB. Aparte de 1 ó 2 *quads* de 4 módulos las configuraciones Express permiten también tan sólo 2 módulos de 256 MB.

¹⁷ El 570 cuadruplica la memoria por DCM frente a los modelos DCM Power4, gracias a los nuevos módulos de 8 GB que se añaden a los módulos de 0,5, 1, 2 y 4 GB.

¹⁸ Para más de cuatro cajones, se requiere un segundo *hub* RIO-2, que ocupa uno de las ranuras del CEC, que ha sido descontada.

¹⁹ Los valores rPerf (nota 9) que se indican para estos modelos, están medidos con AIX 5.3 y son un 30 % superior a los valores con AIX 5.2, debido al SMT.

²⁰ Hay que recordar que el *benchmark* TPC-C del pSeries 690 aún no está alcanzado por la competencia, tal como se indica en la nota 10.

²¹ Configuraciones a partir de 8 procesadores disponibles mediante CoD.

²² Existen tarjetas de 4, 16 y (a partir de abril del 2005) 32 GB, que se instalan en pares con un máximo de 8 pares por placa de 2 MCMs. Hasta abril del 2005 la capacidad máxima estará limitada a la mitad de los valores indicados aquí.

²³ Los discos SCSI disponibles son de 36,4 y 73,4 GB a 15.000 revoluciones / minuto.