
Artículo

[Mario Sanchez Macias](#) · 19 jul, 2021 Lectura de 17 min

Plataformas de datos de InterSystems y su rendimiento - Parte 3: La CPU

Siguiendo la serie de artículos de mi compañero Murray vamos a centrarnos en el artículo donde analizamos la CPU.

Un cliente me pidió que le aconsejara sobre el siguiente escenario: sus servidores de producción se están acercando al final de su vida útil y es el momento de actualizar el hardware. También están pensando en consolidar los servidores por medio de la virtualización y quieren ajustar la capacidad, ya sea con servidores de hardware dedicado o virtualizados.

Hoy analizaremos la CPU. En artículos posteriores explicaré el enfoque para dimensionar correctamente otros "grupos alimenticios de hardware": la memoria y las Entradas/Salidas.

Entonces las preguntas son:

- ¿Cómo se traducen los requisitos de un procesador de hace más de cinco años a los procesadores actuales?
- De los procesadores actuales, ¿cuáles son adecuados?
- ¿Cómo afecta la virtualización a la planificación de la capacidad de la CPU?

Añadido en junio de 2017:

Para profundizar en los aspectos específicos de las consideraciones y la planificación de la CPU en VMware y para conocer algunas preguntas y problemas comunes, puedes consultar esta publicación: [Virtualización de grandes bases de datos: Planificación de la capacidad de la CPU en VMware](#)

[Aquí puedes ver un listado con otros artículos de esta serie >>](#)

Cómo comparar el rendimiento de la CPU usando los análisis de rendimiento (benchmarks) spec.org

Para traducir el uso de la CPU entre los distintos tipos de procesadores para aplicaciones desarrolladas con las plataformas de datos de InterSystems (Caché, Ensemble, HealthShare), puedes usar los análisis de rendimiento de SPECint como un cálculo aproximado para escalar entre procesadores. La página web <http://www.spec.org> ofrece resultados fiables de un conjunto de análisis de rendimiento estandarizados, ejecutados por proveedores de hardware.

Específicamente, SPECint es una forma de comparar distintos modelos de procesadores de los mismos proveedores y de diferentes proveedores (por ejemplo: Dell, HP, Lenovo e Intel, AMD, IBM POWER y SPARC). Puedes utilizar SPECint para entender los requisitos esperados de la CPU para tu aplicación cuando se vaya a actualizar el hardware o si tu aplicación se implementará en varios hardwares de distintos clientes y necesitas establecer una línea de base para una métrica de tamaño, por ejemplo, máximo número de transacciones por núcleo de la CPU para Intel Xeon E5-2680 (o cualquier procesador que elijas).

En el sitio web de SPECint se utilizan varios análisis de rendimiento, pero los resultados de SPECint_{rate}base2006 son los mejores para Caché, ya que se han confirmado a lo largo de muchos años analizando datos de clientes y en nuestros propios análisis de rendimiento.

Como ejemplo, en este artículo vamos a comparar la diferencia entre el servidor Dell PowerEdge de los clientes con procesadores Xeon 5570, y un servidor Dell actual con procesadores Intel Xeon E5-2680 V3. La misma metodología se puede aplicar con los procesadores de servidor Intel Xeon V4 disponibles de forma generalizada.

Ejemplo: Comparando procesadores

Busca SPECint2006Rates en la base de datos de spec.org, por nombre del procesador, por ejemplo E5-2680 V3. Puedes afinar aún más los resultados de tu búsqueda si conoces la marca y el modelo de tu servidor objetivo (por ejemplo, Dell R730); si no, utiliza un proveedor popular. En mi experiencia, creo que los modelos de Dell o HP son buenas bases de referencia de servidores estándar, generalmente no hay gran variación entre los procesadores de diferentes proveedores de hardware.

Al final de esta publicación mostraré un ejemplo paso a paso de cómo buscar resultados con el sitio web spec.org

Supongamos que buscaste en spec.org y encontraste el servidor actual y un posible servidor nuevo, así:

Actual: Dell PowerEdge R710 con Xeon 5570, 2.93 GHz: 8 núcleos, 2 chips, 4 núcleos/chips, 2 hilos/núcleos: SPECintratebase2006 = 251

Nuevo: PowerEdge R730 con Intel Xeon E5-2680 v3, 2.50 GHz: 24 núcleos, 2 chips, 12 núcleos/chips, 2 hilos/núcleos: SPECintratebase2006 = 1030

No sorprende que el nuevo servidor de 24 núcleos aumente más de 4 veces el rendimiento del análisis de rendimiento SPECintratebase2006 con respecto al servidor anterior de 8 núcleos, a pesar de que el servidor nuevo tiene una menor velocidad del reloj. Ten en cuenta que los ejemplos son servidores de dos procesadores que tienen ambas ranuras del procesador ocupadas.

¿Por qué se utiliza SPECintratebase2006 para Caché?

El sitio web spec.org tiene explicaciones de las distintos análisis de rendimiento, pero el resumen es que el análisis de rendimiento SPECintrate2006 es un punto de referencia completo a nivel de sistema que utiliza todos los CPU con HyperThreading.

Para un análisis de rendimiento SPECintrate2006 en particular, se reportan dos métricas: base y pico. Base es un análisis de rendimiento conservador, mientras que Pico es agresivo. Para planificar la capacidad, utiliza los resultados de SPECintratebase2006.

¿Un SPECintratebase2006 cuatro veces mayor significa una capacidad cuatro veces mayor para usuarios o transacciones?

Posiblemente si se utilizaran los 24 núcleos, el rendimiento de la aplicación pudiera escalar a cuatro veces la capacidad del servidor anterior. Sin embargo, hay varios factores que pueden hacer que los resultados sean distintos. SPECint te permitirá conocer el tamaño y el rendimiento que debería ser posible, pero hay algunas advertencias.

Aunque SPECint ofrece una buena comparación entre los dos servidores del ejemplo anterior, no garantiza que el servidor E5-2680 V3 tenga un 75% más de capacidad para los picos de usuarios concurrentes o para el pico de rendimiento de las transacciones que el servidor anterior basado en Xeon 5570. Hay que tener en cuenta otros factores, como el hecho de que los demás componentes de hardware de nuestros "grupos alimenticios de hardware" estén actualizados, por ejemplo, si el almacenamiento nuevo o el actual es capaz de dar servicio al aumento del rendimiento (pronto publicaré un análisis profundo sobre el almacenamiento).

Basado en mi experiencia analizando el rendimiento de Caché y observando los datos de rendimiento de los clientes, Caché es capaz de escalar linealmente a tasas de rendimiento extremadamente altas en un solo servidor a medida que se añaden recursos informáticos (núcleos de la CPU). Esto es aún más cierto con las mejoras que Caché recibe año tras año. Dicho de otro modo, veo un escalado lineal del rendimiento máximo de la aplicación, por ejemplo de las transacciones de la aplicación o reflejado en las glorefs (referencias globales) de Caché conforme se añaden núcleos en la CPU. Sin embargo, si hay cuellos de botella en las aplicaciones, pueden empezar a aparecer con tasas de transacción más altas y repercutir en el escalado lineal. En artículos posteriores explicaré dónde se pueden buscar síntomas de cuellos de botella en las aplicaciones. Una de las mejores medidas que se pueden tomar para mejorar la capacidad de rendimiento de las aplicaciones es actualizar Caché a la última versión.

Nota: Para Caché, los servidores Windows 2008 con más de 64 núcleos lógicos no son compatibles. Por ejemplo, un servidor de 40 núcleos debe tener desactivado el Hyper Threading. Para Windows 2012 se admiten hasta 640 procesadores lógicos. No hay límites en Linux.

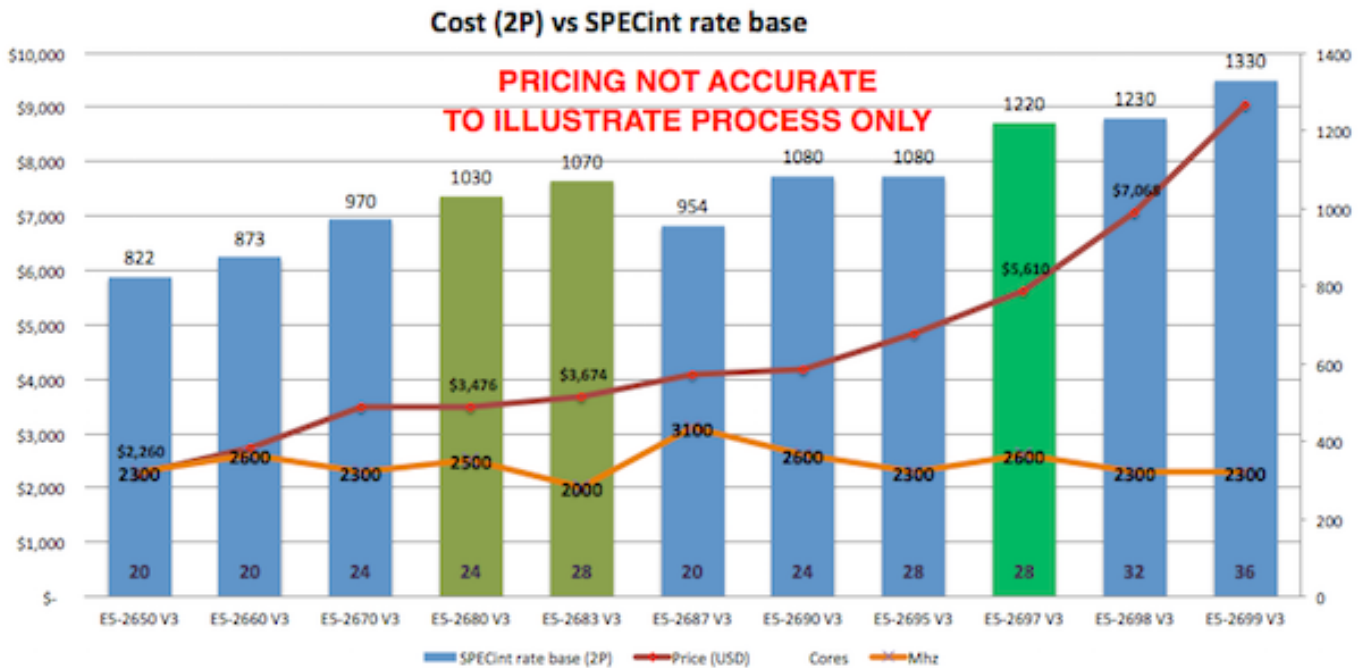
¿Cuántos núcleos necesita la aplicación?

Las aplicaciones varían y cada uno conoce el perfil de sus propias aplicaciones, pero el enfoque común que utilizo cuando planifico la capacidad de la CPU para un servidor (o máquina virtual) es a partir de la monitorización cuidadosa del sistema, comprender que un cierto número de núcleos de CPU de un determinado procesador "estándar" puede sostener una tasa de transacción máxima de n transacciones por minuto. Estas podrían ser episodios, o encuentros, o pruebas de laboratorio o lo que sea. El punto es que el rendimiento del procesador estándar se basa en las métricas que ha recogido en tu sistema actual o un sistema de clientes.

Si conoces el pico de uso de recursos actual de la CPU en un procesador conocido con n núcleos, se puede traducir al número de núcleos necesarios en un procesador más nuevo o diferente para la misma tasa de transacciones utilizando los resultados de SPECint. Con un escalado lineal esperado, $2 \times n$ transacciones por minuto se traduce aproximadamente en $2 \times$ el número de núcleos necesarios.

Cómo elegir un procesador

Como se puede ver en spec.org o analizando la oferta de tu proveedor, hay muchas opciones de procesadores. El cliente de este ejemplo está contento con Intel, por lo que me limitaré a recomendarle servidores Intel actuales. Una forma de proceder es buscar la mejor relación coste-beneficio, o la mejor relación SPECint_{rate}base2006 por dólar y por núcleo. Por ejemplo, el siguiente gráfico muestra los servidores básicos de Dell - tu precio variará, pero esto ilustra que hay puntos más favorables en el precio y recuentos de núcleo más altos adecuados para la consolidación de los servidores utilizando la virtualización. Creé el gráfico fijando el precio de un servidor de producción de calidad, por ejemplo, el Dell R730, y luego examinando distintas opciones de procesadores.



Según los datos del gráfico y la experiencia en sitios de los clientes, el procesador E5-2680 V3 presenta un buen rendimiento a un precio adecuado por SPECint, o por núcleo.

También entran en juego otros factores, por ejemplo, si estás buscando procesadores de servidor para la implementación virtualizada, puede ser más barato aumentar el número de núcleos por procesador, lo que tiene un coste mayor pero logra reducir el número total de servidores necesarios para soportar todas tus máquinas virtuales. De esta forma ahorrarás en software (p. ej. VMware o sistemas operativos) que cobran licencias por ranura de procesador. También tendrás que equilibrar el número de servidores con los requisitos de alta disponibilidad (HA). En artículos posteriores volveré a hablar de VMware y HA.

Por ejemplo, un clúster de VMware HA formado por tres servidores host de 24 núcleos proporciona una buena disponibilidad y una potencia de procesamiento significativa (por número de núcleos), permitiendo configuraciones flexibles de máquinas virtuales de producción y de no producción. Recuerda que VMware HA tiene un tamaño de N+1 servidores, por lo que tres servidores de 24 núcleos equivalen a un total de 48 núcleos disponibles para tus máquinas virtuales.

Núcleos vs GHz: ¿Qué es lo mejor para Caché?

Si tienes que elegir entre núcleos de CPU más rápidos o más núcleos de CPU, debes considerar lo siguiente:

- Si tu aplicación requiere muchos hilos/procesos de cache.exe, un mayor número de núcleos permitirá que más de ellos se ejecuten exactamente al mismo tiempo.
- Si tu aplicación tiene menos procesos, desearás que cada uno se ejecute lo más rápido posible.

Otra forma de ver esto es pensar que si tienes una aplicación de tipo cliente/servidor con muchos procesos, por ejemplo, uno o más por usuario simultáneo, querrás tener más núcleos disponibles. Para aplicaciones basadas en el navegador que utilizan CSP, en las que los usuarios se agrupan en menos procesos de servidor CSP muy ocupados, tu aplicación se beneficiaría de tener un número potencialmente menor de núcleos, pero más rápidos.

En un mundo ideal, ambos tipos de aplicaciones se beneficiarían de contar con muchos núcleos rápidos, asumiendo que no hay contención de recursos cuando varios procesos cache.exe se ejecutan simultáneamente en todos esos núcleos. Como señalé anteriormente, pero merece la pena repetirlo, cada versión de Caché presenta mejoras en el uso de los recursos de la CPU, por lo que actualizar las aplicaciones a las últimas versiones de Caché puede beneficiarse de más núcleos disponibles.

Otra consideración importante es maximizar los núcleos por servidor cuando se utiliza la virtualización. Es posible que las máquinas virtuales individuales no tengan un elevado número de núcleos, pero en conjunto hay que

encontrar un equilibrio entre el número de servidores necesarios para la disponibilidad y minimizar el número de hosts para la administración y la consideración de los costes mediante el aumento del número de núcleos.

Virtualización de VMware y CPU

La virtualización de VMware funciona bien para Caché cuando se utiliza con los componentes de almacenamiento y del servidor actuales. Al seguir las mismas reglas que la planificación de la capacidad física, no hay un impacto significativo en el rendimiento usando la virtualización de VMware en el almacenamiento, la red y los servidores configurados correctamente. La compatibilidad con la virtualización es mucho mejor en procesadores Intel Xeon posteriores, en concreto solo deberías considerar la virtualización en Intel Xeon 5500 (Nehalem) y versiones posteriores, es decir: Intel Xeon 5500, 5600, 7500, serie E7 y serie E5.

Ejemplo: Actualización de hardware, cómo calcular los requerimientos mínimos de CPU

Poniendo en práctica los consejos y procedimientos anteriores, nuestro ejemplo es una actualización de servidor de una carga de trabajo que se ejecuta en Dell PowerEdge R710 con 8 núcleos (dos procesadores de 4 núcleos Xeon 5570).

Al hacer un gráfico del uso actual de la CPU en el servidor de producción primario en el cliente, vemos que el servidor está alcanzando un máximo de menos del 80% durante la parte más activa del día. La fila de ejecución no está bajo presión. La Entrada/Salida y la aplicación también funcionan bien, por lo que no hay cuellos de botella que ~~superen~~ supriman la CPU.

Regla general: Comienza por dimensionar los sistemas para un uso máximo del 80% de la CPU al final de la vida útil del hardware, teniendo en cuenta el crecimiento esperado (por ejemplo, un aumento de usuarios/transacciones). Esto permite crecimientos inesperados, eventos inusuales o picos inesperados de actividad.

Para que los cálculos sean más claros, vamos a suponer que no se espera un crecimiento del rendimiento durante la vida del nuevo hardware:

El escalado por núcleo se puede calcular como: $(251/8)$ $(1030/24)$ o un aumento del 26% en el rendimiento por núcleo.

El 80% de la CPU con 8 núcleos en el servidor anterior equivale aproximadamente al 80% de la CPU con 6 núcleos en los nuevos procesadores E5-2680 V3. Así, el mismo número de transacciones podría estar soportado por seis núcleos.

El cliente tiene varias opciones, puede comprar nuevos servidores hardware dedicados que cumplan el requisito mínimo de seis núcleos de CPU E5-2680 V3 o su equivalente, o avanzar con sus planes para virtualizar su carga de trabajo de producción en VMware.

Virtualizar tiene sentido para aprovechar las ventajas de la consolidación, flexibilidad y alta disponibilidad de los servidores. Dado que hemos calculado los requisitos de la CPU, el cliente puede avanzar con confianza para ajustar correctamente el tamaño de las máquinas virtuales de producción en VMware. Como nota al margen, comprar servidores actuales con una baja cantidad de núcleos es una opción cara o de difícil acceso, lo cual hace que la virtualización sea una opción aún más atractiva.

Virtualizar también ofrece ventajas si se espera un crecimiento significativo. Los requisitos de la CPU se pueden calcular en función del crecimiento en los primeros años. Si se realiza una monitorización constante, una estrategia válida es añadir recursos adicionales solo cuando sea necesario y antes de necesitarlos.

Consideraciones sobre la CPU y la virtualización

Como hemos visto, los sistemas de producción en Caché son dimensionados en función de los análisis de rendimiento y las medidas realizadas en los clientes. También es válido dimensionar los requerimientos de la CPU virtual (vCPU) de VMware a partir de la monitorización de los equipos de hardware dedicado. La virtualización por medio del almacenamiento compartido añade muy poca sobrecarga a la CPU comparado con el hardware dedicado**. Para los sistemas de producción, utiliza una estrategia de ajustar el tamaño inicial del sistema igual que los núcleos de CPU de hardware dedicado.

**Nota: Para implementaciones de VMware VSAN debes añadir un búfer de la CPU a nivel host del 10% para el procesamiento de vSAN.

Se deben considerar las siguientes reglas para asignar CPUs virtuales:

Recomendación: No asignes más CPUs virtuales de las necesarias para lograr un rendimiento adecuado.

- Aunque se puede asignar un gran número de CPUs virtuales a una máquina virtual, la práctica recomendada es no asignar más de las necesarias, ya que puede haber una sobrecarga de rendimiento (normalmente baja) para administrar las CPUs virtuales no utilizadas. La clave es monitorizar tus sistemas frecuentemente para asegurarse de que las máquinas virtuales tengan el tamaño adecuado.

Recomendación: Los sistemas de producción, especialmente los servidores de bases de datos, inicialmente ajustan el tamaño para 1 CPU física = 1 CPU virtual.

- Se espera que los servidores de producción, especialmente los de bases de datos, tengan un alto nivel de uso. Si necesitas seis núcleos físicos, ajusta el tamaño para seis núcleos virtuales. Consulta también la nota sobre el Hyper Treading que se encuentra más abajo.

Sobresuscripción

La sobresuscripción se refiere a varios métodos por los cuales más recursos de los disponibles en el servidor físico se pueden asignar hacia los servidores virtuales que son compatibles con ese host. En general, es posible consolidar los servidores por medio de la sobresuscripción de recursos de procesamiento, memoria y almacenamiento en máquinas virtuales.

La sobresuscripción del servidor sigue siendo posible cuando se ejecutan bases de datos de producción en Caché; sin embargo, para ajustar el tamaño inicial de los sistemas de producción se asume que la CPU virtual tiene dedicación completa del núcleo. Por ejemplo, si tienes un servidor E5-2680 V3 de 24 núcleos (2 de 12 núcleos), ajusta el tamaño para una capacidad total de hasta 24 CPUs virtuales, sabiendo que podría haber espacio libre disponible para consolidación. En esta configuración se asume que el Hyper Threading está habilitado a nivel del servidor. Una vez que hayas pasado tiempo monitorizando la aplicación, el sistema operativo y el rendimiento de VMware durante las horas pico de procesamiento, puedes decidir si es posible una mayor consolidación.

Si estás mezclando máquinas virtuales que no son de producción, una regla general que uso a menudo para ajustar el tamaño del sistema y calcular los núcleos de CPU totales es ajustar inicialmente las CPUs que no son de producción en 2:1 físicas a virtuales. Sin embargo, esta es definitivamente un área en la que los resultados podrían variar y se deberá monitorizar el sistema para ayudar a planificar la capacidad. Si tienes dudas o no tienes experiencia, puedes separar las máquinas virtuales de producción de las que no lo son a nivel de servidor o por medio de la configuración de vSphere hasta comprender las cargas de trabajo.

VMware vRealize Operations y otras herramientas de terceros ofrecen la posibilidad de monitorizar los sistemas a lo largo del tiempo, y sugerir la consolidación o alertar de que se requieren más recursos para las máquinas virtuales. En un artículo posterior hablaré sobre otras herramientas de monitorización disponibles.

La conclusión es que, en el ejemplo de nuestros clientes, pueden estar seguros de que su máquina virtual de producción con 6 CPUs virtuales funcionará bien; siempre y cuando los otros "grupos alimenticios" principales, como las Entradas/Salidas y el almacenamiento, tengan capacidad suficiente ;-D

Hyper Threading y planificación de la capacidad

Un buen punto de partida para ajustar el tamaño de las máquinas virtuales basado en reglas conocidas para los servidores físicos es calcular los requisitos de la CPU del servidor físico, con el objetivo por procesador con Hyper Threading activado, y después simplemente hacer la traducción:

una CPU física (incluido el Hyper Threading) = una CPU virtual (incluido el Hyper Threading).

Un concepto erróneo común es que el Hyper Threading duplica de alguna manera la capacidad de la CPU virtual. Esto NO es cierto para los servidores físicos o para las CPUs virtuales lógicas. Como regla general, el Hyper Threading en un servidor de hardware dedicado puede ofrecer un 30% de capacidad de rendimiento adicional comparado con el mismo servidor sin Hyper Threading. La misma regla del 30% se aplica a los servidores virtualizados.

Licencias y CPUs virtuales

En vSphere, se puede configurar una máquina virtual con una cierta cantidad de ranuras o núcleos. Por ejemplo, si tienes una máquina virtual con doble procesador, se puede configurar para que tenga dos ranuras de CPU, o una única ranura con dos núcleos de CPU. Desde el punto de vista de la ejecución no hay mucha diferencia porque el hipervisor decidirá en última instancia si la máquina virtual se ejecuta en una o dos ranuras físicas. Sin embargo, especificar que la máquina virtual de doble CPU en realidad tiene dos núcleos en lugar de dos ranuras podría marcar una diferencia para las licencias de software que no son de Caché.

Resumen

En este artículo describí cómo comparar procesadores entre proveedores, servidores o modelos usando los resultados del análisis de rendimiento SPECint. También cómo planificar la capacidad y elegir los procesadores en función del rendimiento y la arquitectura, tanto si se utiliza la virtualización como si no.

Estos son temas complejos, en los que es fácil irse por las ramas... Por eso, al igual que en mis otras publicaciones, no dejes de hacer comentarios o preguntas si deseas profundizar en algún tema relacionado.

EJEMPLO: Búsqueda de resultados de SPECintrate2006.

En la siguiente imagen se muestra la selección de los resultados de SPECintrate2006.

Utiliza la pantalla de búsqueda para reducir resultados.

Ten en cuenta que también se pueden descargar todos los registros a un archivo .csv de aproximadamente 20 MB para su procesamiento local, por ejemplo con Excel.

Los resultados de la búsqueda muestran el Dell R730.

Selecciona HTML para dar el resultado completo del análisis de rendimiento.

Puedes ver los siguientes resultados para servidores con los procesadores de nuestro ejemplo.

Dell PowerEdge R710 con 2.93 GHz: 8 núcleos, 2 chips, 4 núcleos/chips, 2 hilos/núcleos con Xeon 5570:
SPECintratebase2006 = 251

PowerEdge R730 (Intel Xeon E5-2680 v3, 2.50 GHz) de 24 núcleos, 2 chips, 12 núcleos/chips, 2 hilos/núcleos con Xeon E5-2680 v3: SPECintratebase2006 = 1030

[#Administración del sistema](#) [#Arquitecturas y Soluciones de Negocio con InterSystems](#) [#Rendimiento](#) [#Caché](#)
[#InterSystems IRIS](#) [#InterSystems IRIS for Health](#)

URL de
fuente: <https://es.community.intersystems.com/post/plataformas-de-datos-de-intersystems-y-su-rendimiento-parte-3-la-cpu>