

Documento de tecnología

Comparaciones de interfaces de SSD para empresas

Introducción

PCI Express (PCIe) es una interfaz de bus de uso general que se utiliza en aplicaciones informáticas cliente y empresariales. Las interfaces de almacenamiento masivo existentes (SATA, SAS) se conectan al equipo host a través de adaptadores de host que a su vez se conectan a la interfaz PCIe.

La interfaz SATA se diseñó como una interfaz para unidades de disco duro y la interfaz SAS se diseñó como una interfaz de dispositivo y como una interfaz / infraestructura de subsistema de almacenamiento. A medida que han ido evolucionando los requisitos de las unidades de disco duro y los sistemas, en los que se requieren interfaces más rápidas y nuevas funciones, las interfaces SATA y SAS se han sometido a diversas revisiones.

Las unidades de estado sólido (SSD) han añadido rápidamente a estas interfaces nuevos requisitos de rendimiento significativos a medida que las velocidades de transmisión de datos de las SSD han pasado de decenas de MB/s a cientos y ahora miles de MB/s. Además del aumento de las velocidades de transmisión de datos, la ausencia de movimiento mecánico en las SSD también ha incrementado el número de operaciones de entrada y salida por segundo (IOPS) que pueden realizar estos dispositivos de almacenamiento.

Con este desarrollo ha surgido la necesidad de mejorar la implementación de los estándares existentes, así como mejorar los estándares de interfaces existentes, para gestionar los nuevos requisitos de rendimiento a la vez que se mantiene la compatibilidad con la arquitectura existente del sistema.

Este documento analiza las diferentes interfaces y compara las diversas compensaciones halladas en cuanto al rendimiento y la compatibilidad.

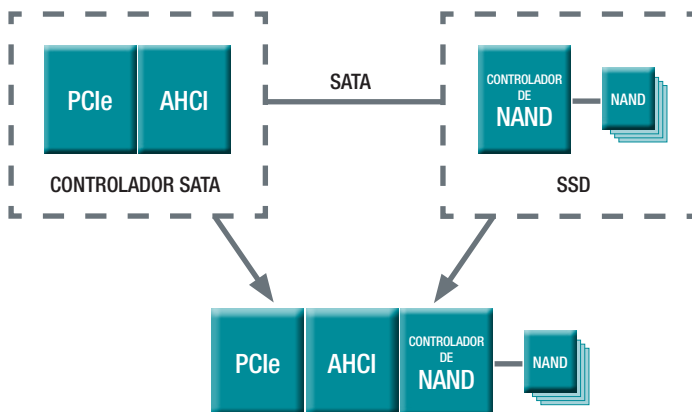
Comparaciones de interfaces de SSD para empresas



Interfaz SATA

SATA es una interfaz de bajo coste diseñada para una conexión de punto a punto mediante un cable o una trayectoria de placa de circuito impreso (PCB). La conexión de host se realiza a una interfaz de controlador de host avanzada (AHCI), que a menudo se encuentra en el conjunto de chips del host, como el adaptador de host del bus PCIe. Existen algunos problemas de diseño con esta interfaz que pueden crear una carga de trabajo del bus de 1 μ s (o más) para cada comando. No es un problema importante para las unidades de disco duro en las que una transferencia de 4 KB se realiza en 10 μ s, pero las SSD pueden transferir 4 KB de datos en 2 μ s (o menos), por lo que la carga de trabajo se vuelve significativa y la interfaz SATA deja de ser tan interesante como interfaz de almacenamiento masivo de alto rendimiento.

SATA sigue siendo adecuada como una interfaz de bajo coste para SSD donde el coste y no el rendimiento sea un factor decisivo importante. La arquitectura de SATA también puede consolidarse en un adaptador de host que gestione el conjunto de comandos de SATA sin incluir la interfaz física SATA (PHY) (Figura 1).



Fuente: Seagate Technology, 2011
Figura 1. Consolidación de arquitectura

Interfaz SAS

SAS también es una interfaz serie, conectada al host a través de un adaptador de host, pero existen diferencias significativas que hacen que sea adecuada como interfaz de SSD:

- Menor carga de trabajo en el hardware
- Velocidades de transferencia de datos más rápidas
- Puertos anchos
- Interfaces de controlador eficaces

Además, SAS incluye funciones que no se encuentran en SATA y que mejoran la fiabilidad y disponibilidad de los dispositivos conectados a la interfaz:

- Protocolo de serie resistente
- Compatibilidad con varios hosts
- Integridad de los datos de extremo a extremo
- Capacidad de puerto dual
- Grados elevados de concurrencia y agregación

Menor carga de trabajo en el hardware

No hay ninguna interfaz de host universal para SAS que sea equivalente a la AHCI de SATA. Por el contrario, varios proveedores compiten en el mercado de adaptadores de host SAS en el que el rendimiento es un factor clave: para las interfaces de unidades de discos duros individuales y también para diversos sistemas RAID en los que se añaden las velocidades de transferencia de datos de varios giros de unidades de discos duros para mejorar la velocidad de transferencia. Asimismo, los adaptadores de host SAS están diseñados para gestionar SSD y unidades de disco duro de mayor rendimiento (como unidades de 15.000 rpm con formateo parcial). Ya que el adaptador de host de hardware y el controlador del dispositivo que gestiona dicho adaptador de host están diseñados como un sistema, están empezando a estar disponibles nuevos diseños optimizados para SSD y están apareciendo nuevas mejoras, tanto en las velocidades de transferencia de datos como en las IOPS.

Velocidades de transferencia de datos más rápidas

Los puertos SAS admiten actualmente velocidades de transmisión de datos de hasta 6 Gb/s. Empresas como LSI y PMC-Sierra están presentando muestras de diseños actualmente en desarrollo que admiten velocidades de transmisión de datos de 12 Gb/s y más de 2 millones de IOPS, con la posibilidad de alcanzar los 24 Gb/s en el futuro.

Puertos anchos

El concepto de los puertos anchos es inherente a la arquitectura de SAS. Esto significa que pueden agregarse varios enlaces para permitir varias rutas simultáneas entre uno o más hosts y un dispositivo. El conector actual de la unidad SAS define dos puertos para la unidad. Como opción de diseño, las unidades de disco duro actuales no admiten el puerto ancho, únicamente el puerto dual, en el que cada puerto tiene una dirección SAS diferente, lo que impide su configuración como puerto ancho.

Las propuestas aceptadas para SAS-3 (12 Gb/s) permiten un incremento en el número de puertos de la unidad a cuatro, todos los cuales pueden conectarse al mismo dominio, o por parejas a diferentes dominios. Un número muy limitado de SSD admite el puerto ancho además del puerto dual en un dispositivo con dos puertos.

Comparaciones de interfaces de SSD para empresas



Protocolo de serie resistente

El protocolo de serie SAS permite la formación de los transmisores y receptores de serie. Esto mejora la calidad de la señal del cable o el panel de conexión compensando la longitud del canal, la discrepancia en la impedancia y la interferencia entre símbolos. El protocolo de serie SAS también gestiona la detección de errores y la retransmisión a nivel de hardware. Esto permite una recuperación más rápida de problemas de señalización intermitente.

Compatibilidad con varios hosts

La interfaz SAS y la trama de conmutación permite que varios hosts accedan al mismo dispositivo. Esta función puede utilizarse para gestionar fallos de host, así como fallos en la ruta de los datos, y así mejorar la disponibilidad de los datos.

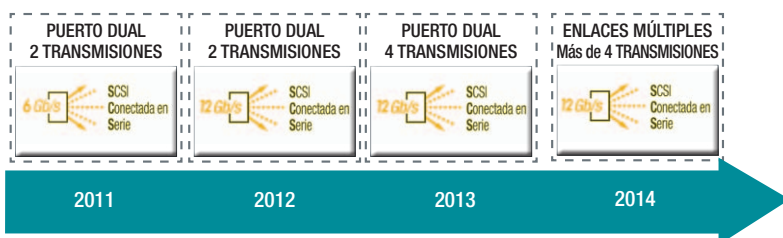
Integridad de los datos de extremo a extremo

La interfaz SAS puede verificar la integridad de los datos mediante comprobaciones de redundancia cíclica (CRC) de los datos desde el momento de su creación en el búfer de datos del host, pasando por la transferencia a través de las interfaces PCIe y SAS, hasta que se almacenan en el dispositivo y se vuelven a leer y transferir al búfer de datos del host. Esto permite establecer múltiples puntos de control a lo largo de la ruta desde las aplicaciones hasta los controladores RAID y en los dispositivos. A veces esta función se denomina protección de la información (PI).

Capacidad de puerto dual

Los dispositivos SAS de destino admiten el funcionamiento de puertos duales. Esto permite crear dos dominios predeterminados y aumenta la disponibilidad. Aunque se produzca un fallo en una de las rutas a un puerto que impida el acceso por esa ruta, todavía se podrá acceder a un dispositivo mediante el segundo puerto.

Históricamente, Seagate ha liderado la adopción de interfaces en el mercado. Seagate está colaborando con la SCSI Trade Association (STA) y otros líderes del sector para sacar el máximo partido de la infraestructura SAS existente ampliamente extendida (Figura 2). La tabla 1 muestra de qué modo SAS de 12 Gb/s y los enlaces múltiples benefician a los creadores de sistemas y las organizaciones de usuarios finales.



Fuente: Seagate Technology, 2011
Figura 2. Evolución de la interfaz SAS

Interfaz PCI Express

PCI Express (PCIe) es la interfaz fundamental que conecta los dispositivos periféricos al procesador host y, a través de un controlador de memoria, a la arquitectura de la memoria del sistema. Tanto la interfaz SATA como la interfaz SAS tratadas anteriormente se conectan a través de una interfaz PCIe (o adaptador de host) al procesador host y la memoria.

Enlaces múltiples (bytes escritos)	X4 (4 x 600 MB/s)
Potencia disponible	25 W (2,5 pulgadas)
Latencia total	Muy baja
Protocolo de varios hosts	Sí
Alta disponibilidad	Sí (puerto dual)
Escalabilidad	Excelente
Pila de protocolos resistente y probada	Sí
Intercambiable en caliente	Sí
Compatible con software de gestión existente	Sí

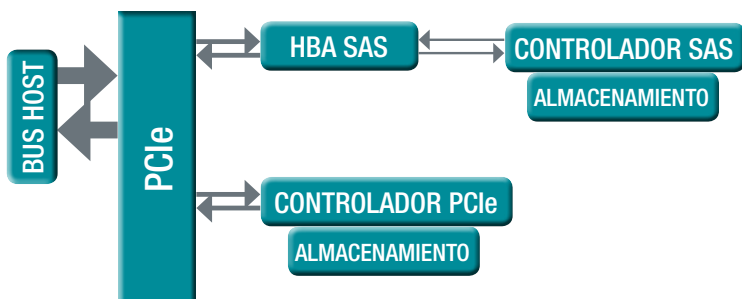
Fuente: Seagate Technology, 2011

La interfaz PCIe es una implementación en serie de la interfaz PCI original que proporcionaba una conexión paralela de direcciones/datos entre periféricos y el procesador host/la memoria. La interfaz PCIe se comunica a través de una o más líneas compuestas por una interfaz serie de transmisión y una de recepción por cada línea. Se pueden utilizar hasta 32 líneas para conectar un host a un dispositivo. La velocidad de transmisión de datos en serie de cada línea depende de la versión del estándar de PCIe implementada, la versión actual es 3.0 y la velocidad de transmisión de datos es de aproximadamente 1 GB/s.

Para un servidor de 1U, la interfaz PCIe está diseñada para utilizar un único conector en una placa base (cliente), o un adaptador de ángulo recto de dos conectores en una placa base (servidor). También hay disponible un sistema de cableado (aunque apenas se utiliza). Un servidor de 2U, 4U o 7U tiene muchas más ranuras PCIe, de manera similar a las implementaciones de clientes. La especificación PCIe también utiliza formación de los transmisores (y receptores) para adaptar las variaciones de impedancia de una configuración, pero su objetivo son canales de transmisión de menor longitud que en SAS.

Los interruptores PCIe pueden adaptarse a los métodos de virtualización de E/S de raíz única (SR-IOV) y virtualización de E/S de raíz múltiple (MR-IOV) utilizados para mejorar el rendimiento del controlador en sistemas virtuales (hipervisor) con uno o varios hosts. SR-IOV comienza a estar disponible de manera general en los adaptadores; sin embargo, puede que VMware todavía no saque partido de esto. MR-IOV normalmente no es compatible con los adaptadores.

Comparaciones de interfaces de SSD para empresas



Fuente: Seagate Technology, 2011
Figura 3. Evolución de la interfaz SAS

Los dispositivos de almacenamiento que se conectan mediante la interfaz PCIe lo hacen a través de una conexión de registro directa o a través de un adaptador de host que, a continuación, se conecta al dispositivo a través de un cableado adicional o una interfaz del tipo panel de conexión.

Actualmente existen diversas implementaciones diferentes en ambas arquitecturas. SATA utiliza una implementación de adaptador de bus host en el conjunto de chips del sistema (*Southbridge*), ya sea AHCI Intel o AMD, que requiere diferentes controladores de AHCI pero que se asigna a implementaciones anteriores de IDE compatibles. Estas interfaces también implementan diversas funciones de gestión de RAID.

SAS tiene varios proveedores de adaptadores de bus host, con expansores y controladores RAID adicionales disponibles, todos los cuales utilizan BIOS y controladores de dispositivos patentados para cubrir diferentes necesidades de rendimiento y capacidad de configuración.

La interfaz de controlador de PCIe se implementa en la especificación NVM Express y en la especificación SOP propuesta.

La arquitectura consolidada de SATA descrita anteriormente es otro ejemplo de la conexión de registro directa de PCIe.

SSD PCIe en la actualidad

Existen dos tipos principales de SSD PCIe en el mercado en la actualidad: nativa y agregadora. El controlador nativo se conecta con el bus PCIe host y, a continuación, controla directamente varios buses de memoria flash. Estos a menudo utilizan una interfaz de software patentada por el fabricante y utilizada únicamente para el dispositivo específico. Algunas de estas implementaciones trasladan la carga de la conversión de direcciones y otras funciones a la CPU y la memoria del host. A cambio, esto produce una reducción en los recursos del sistema para las aplicaciones cuando los dispositivos se utilizan con fuertes cargas de trabajo. Además, al ser relativamente nuevas en el mercado, estas exclusivas combinaciones de unidades y hardware suelen presentar inestabilidades, ya que sus ecosistemas siguen evolucionando.

El modelo agregador ofrece un enfoque diferente del diseño. Este enfoque utiliza un controlador RAID SAS o SATA, al que se conectan varias SSD SAS o SATA. Se entregan en el mismo paquete en una única tarjeta PCIe. El controlador RAID añade el rendimiento de varios dispositivos para ofrecer niveles elevados de rendimiento. Al estar basados en interfaces de hardware y software de clase empresarial existentes que han sido probadas, estos diseños son muy estables y están muy consolidados. Asimismo, estos diseños utilizan controladores inteligentes que realizan conversiones de direcciones y otras funciones, lo que permite

que las aplicaciones puedan utilizar por completo los ciclos y la memoria de la CPU del sistema, incluso con fuertes cargas de trabajo de E/S.

El futuro de las SSD PCIe

Tanto el enfoque de SOP como el de NVMe son parecidos en cuanto a su arquitectura. Sin embargo, NVMe se está desarrollando en un grupo de trabajo del sector, mientras que SOP se está desarrollando en un reconocido foro de estándares abiertos. NVMe está dirigido únicamente al uso de dispositivos de memoria no volátil, mientras que SOP también está dirigido al uso para adaptadores de bus host y controladores RAID con funciones para establecer conexiones entre los diversos dispositivos SOP. Por otra parte, SOP aprovecha al máximo las arquitecturas y funciones existentes del sector, mientras que NVMe utiliza un nuevo conjunto de instrucciones muy limitado y una interfaz de colas.

Ventajas y problemas de las interfaces

Todas y cada una de las arquitecturas de almacenamiento descritas ofrecen ciertas ventajas al igual que plantean algunos problemas. En función del diseño general del sistema, las ventajas de utilizar una arquitectura específica pueden compensar los problemas asociados a dicha arquitectura, por lo que es necesario realizar un análisis detallado para tomar la decisión adecuada. Dicha decisión también debe tener en cuenta la compatibilidad con el diseño existente del sistema.

Por ejemplo, la actualización de un equipo portátil con una unidad de disco duro SATA de 2,5 pulgadas existente y una SSD únicamente funcionaría con una SSD del mismo tamaño físico y con la misma interfaz SATA (o una más reciente). Habrá un límite relativo a la velocidad permitida para la SSD en este caso; superar la velocidad de la interfaz SATA del host existente no aumentará el rendimiento del sistema.

En una situación similar, un servidor empresarial que esté utilizando una unidad de disco duro SAS de 15.000 rpm con formateo parcial para almacenar un índice de base de datos puede actualizarse mediante una SSD SAS, que aumentará el rendimiento total del sistema, pero que también formará un cuello de botella en otro factor del sistema (CPU, memoria, red, adaptadores, etc.).

En la arquitectura de un sistema nuevo, la adición de almacenamiento de estado sólido puede aumentar significativamente el rendimiento del sistema, pero únicamente en la medida en que

Comparaciones de interfaces de SSD para empresas



Tabla 2. Comparativa de SSD PCIe nativa y agregadora

	Nativa	Agregadora
Comandos / transporte	Patentado (FTL ¹ en el host / memoria principal)	SCSI o SATA (SSD múltiples, controlador en tarjeta)
Comité	Ninguno	Ninguno
Basado en estándares	No	Sí
Rendimiento con flash	Alto	Alto
Carga de trabajo de la CPU	Alta	Baja
Latencia con cola corta	Muy baja	Baja
Latencia con cola profunda	Moderada	Baja
Extensibilidad de casos de uso	No	Sí (RAID, HBA, etc.)
Consolidación	En evolución	Basado en probadas arquitecturas del sector
Conjunto de funciones empresariales (PI, seguridad, gestión, etc.)	No	Depende de la implementación

¹ FTL: Flash Translation Layer (capa de traducción flash)

Fuente: Seagate Technology, 2011

Tabla 3. Comparativa de SSD PCIe SOP y NVMe

	SOP ¹	NVMe ²
Comandos / transporte	SOP / PQI ³ (FTL en el controlador)	NVMe / NVMe (FTL en el controlador)
Comité	T10 / INCITS ⁴	Grupo de trabajo del sector
Basado en estándares	Sí (ANSI / ISO)	No
Rendimiento con flash	Muy alto	Muy alto
Carga de trabajo de la CPU	Baja	Baja
Latencia con cola corta	Muy baja	Muy baja
Latencia con cola profunda	Baja	Baja
Extensibilidad de casos de uso	Sí (RAID, HBA, etc.)	No (únicamente NVM)
Consolidación	Basado en probadas arquitecturas del sector	TBD
Conjunto de funciones empresariales (PI, seguridad, gestión, etc.)	Compatibilidad total	Limitada

¹ SOP: SCSI over PCI Express (SCSI sobre PCI Express)

² NVMe: Nonvolatile Memory Express (memoria no volátil expres)

³ PQI: PCIe Queuing Interface (interfaz de colas de PCIe)

⁴ INCITS: International Committee for Information Technology Standards (Comisión Internacional de Normas de Tecnología de la Información)

Fuente: Seagate Technology, 2011

www.seagate.com

AMÉRICA ASIA/PACÍFICO
EUROPA, ORIENTE MEDIO Y ÁFRICA

Seagate Technology LLC 10200 South De Anza Boulevard, Cupertino, California 95014, EE. UU., +1 408 658 1000
Seagate Singapore International Headquarters Pte. Ltd. 7000 Ang Mo Kio Avenue 5, Singapur 569877, +65 6485 3888
Seagate Technology SAS 16-18 rue du Dôme, 92100 Boulogne-Billancourt, Francia, +33 1 41 86 10 00

© 2012 Seagate Technology LLC. Todos los derechos reservados. Impreso en EE. UU. Seagate, Seagate Technology y el logotipo Wave son marcas comerciales registradas de Seagate Technology LLC en Estados Unidos y/o en otros países. Todas las demás marcas comerciales o marcas registradas pertenecen a sus respectivos propietarios. Seagate se reserva el derecho a modificar las ofertas o especificaciones de los productos sin previo aviso. TP625.1-1203ES, marzo de 2012

el resto de la arquitectura del sistema pueda adaptarse al aumento en la velocidad de transmisión y el ancho de banda de los datos. El aumento en la velocidad de transmisión de datos en las SSD también requiere un mayor suministro de energía al dispositivo y una mayor disipación del calor en el lugar donde se instale la SSD.

Otro factor es si resulta oportuna la disponibilidad de la compatibilidad con la BIOS y los controladores de dispositivos del sistema operativo para estas nuevas interfaces de SSD, así como la fiabilidad inicial del software.

Aspectos de la latencia de las interfaces y las SSD flash

Existen numerosos conceptos erróneos sobre qué factores añaden latencia y en qué medida afectan realmente al rendimiento de la aplicación. Al considerar este aspecto, es importante tratar de ver el conjunto como un todo, y no solo fijarse en una parte de él.

Sin duda, los principales elementos que contribuyen a la latencia de las SSD son los propios elementos flash. Los tiempos de acceso de SLC son de más de 25 μ s y los de MLC son de más de 50 μ s, suponiendo en ambos casos que no haya ninguna contención de acceso. A medida que aumentan las profundidades de colas, la contención de acceso a los elementos flash puede aumentar la latencia en gran medida.

Una vez que un elemento flash comienza su acceso, las otras solicitudes para el mismo elemento deben esperar. Hasta ocho matrices flash comparten un bus común, lo que hace que la matriz deba esperar su turno para utilizar el bus. Las actividades de mantenimiento aumentan la latencia de manera adicional (conversión de direcciones, recolección de basura, nivelación del desgaste, etc.).

Otro aspecto es el sistema operativo, que añade latencia independientemente del protocolo de acceso y la interconexión. Aquí se incluyen el sistema de archivos, el gestor de volúmenes, los controladores de clase y las cargas de trabajo de cambio de contexto.

Las diferencias en protocolos e interconexiones tienen efectos insignificantes en la latencia en lo que respecta a una aplicación (fracciones de microsegundo).