Documento tecnico

Interfacce SSD classe Enterprise a confronto

Introduzione

PCI Express (PCIe) è un'interfaccia bus generica utilizzata nelle applicazioni di elaborazione client e classe Enterprise. Le interfacce di memorizzazione di massa esistenti (SATA, SAS) si collegano al computer host tramite adattatori host che, a loro volta, si collegano all'interfaccia PCIe.

L'interfaccia SATA è stata progettata come interfaccia delle unità disco (HDD), mentre l'interfaccia SAS è stata progettata sia come interfaccia del dispositivo sia come interfaccia/infrastruttura del sottosistema di memorizzazione. Man mano che le unità disco e i requisiti di sistema si sono evoluti e si sono rese necessarie interfacce più veloci e nuove funzioni, le interfacce SATA e SAS sono state oggetto di numerose revisioni.

Le unità con memoria a stato solido (SSD) hanno presto preteso nuove prestazioni significative da queste interfacce. Le velocità di trasferimento dei dati delle unità SSD sono passate da decine di MB/s a centinaia e ora migliaia di MB/s. Oltre all'incremento delle velocità di trasferimento dei dati, a causa della mancanza di movimenti meccanici nelle unità SSD è anche aumentato il numero di operazioni di I/O al secondo (IOPS) che è possibile eseguire con questi dispositivi di memorizzazione.

In seguito a questo sviluppo è stato necessario implementare meglio gli standard esistenti e ottimizzare gli standard di interfaccia disponibili per gestire i nuovi requisiti di prestazione e, allo stesso tempo, mantenere la compatibilità con l'architettura di sistema esistente.

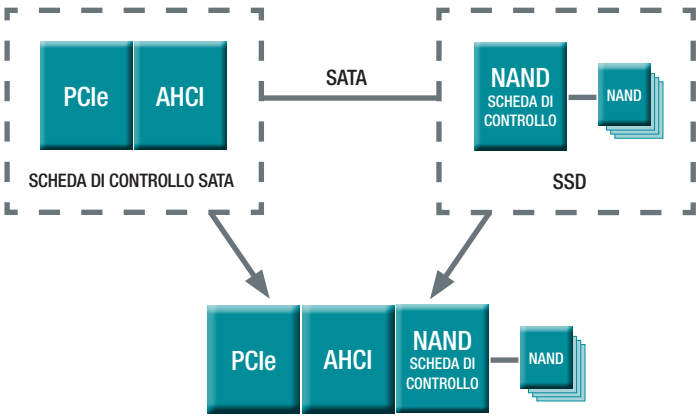
In questo documento vengono illustrati le diverse interfacce e i contrasti che caratterizzano i vari compromessi a livello di prestazioni e compatibilità.

Interfacce SSD classe Enterprise a confronto

Interfaccia SATA

SATA è un'interfaccia a basso costo progettata per il collegamento point-to-point tramite una traccia di cavo o circuito stampato. Il collegamento host viene stabilito con un'interfaccia AHCI (Advanced Host Controller Interface) che solitamente risiede nel chipset host come adattatore host sul bus PCIe. Questa interfaccia presenta alcune problematiche di progettazione che possono creare un carico di lavoro del bus di 1 µs (o più) per ogni comando. Questo non è un problema serio per le unità disco che trasferiscono 4 KB in 10 µs. Le unità SSD, invece, possono trasferire 4 KB di dati in 2 µs (o meno). Il carico di lavoro è quindi significativo, ma l'interfaccia SATA è meno interessante come interfaccia di memorizzazione di massa dalle prestazioni elevate.

SATA è un'interfaccia SSD a basso costo. Scegliere questa soluzione non dipende dalle prestazioni, ma dai costi. L'architettura SATA può inoltre essere consolidata in un adattatore host che gestisce il set di comandi SATA senza effettivamente includere l'interfaccia SATA fisica (PHY) (Figura 1).



Fonte: Seagate Technology, 2011

Figura 1. Consolidamento delle architetture

Interfaccia SAS

SAS è anche un'interfaccia seriale che si collega all'host tramite un adattatore host. Tuttavia, vi sono differenze significative che ne fanno la soluzione ideale come interfaccia SSD.

- Carico di lavoro hardware ridotto
- Velocità di trasferimento migliori
- Porte ampie
- Interfacce driver-scheda di controllo efficienti

Inoltre l'interfaccia SAS include funzioni, non disponibili in SATA, che migliorano l'affidabilità e la disponibilità dei dispositivi collegati all'interfaccia.

- Protocollo seriale robusto
- Supporto di più host
- Integrità dei dati end-to-end
- Funzionalità con due porte
- Livelli elevati di concomitanza e aggregazione

Carico di lavoro hardware ridotto

Non esiste un'interfaccia host universale per SAS equivalente alla scheda di controllo AHCI SATA. Invece più venditori competono nel mercato degli adattatori host SAS in cui le prestazioni rappresentano un fattore importante, non solo per interfacciare singole unità disco, ma anche vari sistemi RAID in cui le velocità di trasferimento di più rotazioni delle unità disco vengono aggregate per migliorare la velocità di trasferimento. Inoltre gli adattatori host SAS sono stati progettati per gestire unità SSD e unità disco dalle prestazioni più elevate (ad esempio, *unità da 15.000 giri/min* con distanza dell'attuatore ridotta). Poiché l'adattatore host hardware e il driver del dispositivo che gestisce l'adattatore host sono stati progettati come sistema, sono ora disponibili nuovi design ottimizzati per le unità SSD che migliorano ulteriormente non solo le velocità di trasferimento, ma anche le IOPS.

Velocità di trasferimento migliori

Attualmente le porte SAS supportano velocità di dati fino a 6 Gbit/s. Società come LSI e PMC-Sierra stanno provando i design sviluppati per supportare velocità di trasferimento dei dati pari a 12 Gbit/s e più di 2 milioni di IOPS, con la possibilità di 24 Gbit/s in futuro.

Porte ampie

Inerente all'architettura SAS è il concetto di porte ampie, in cui più collegamenti possono essere aggregati per consentire più percorsi simultanei tra uno o più host e un dispositivo. Il connettore dell'unità SAS corrente definisce due porte per l'unità. Secondo la progettazione, le unità disco correnti non supportano le porte ampie, ma solo due porte, in cui ogni porta ha un indirizzo SAS diverso che impedisce la configurazione come porta ampia.

Le proposte accettate per *SAS-3* (12 Gbit/s) prevedono un aumento del numero di porte sull'unità a quattro. Tutte le porte si collegano allo stesso dominio o in coppie a domini diversi. Un numero molto limitato di unità SSD può supportare le porte ampie oltre alle due porte su un dispositivo a due porte.

Interfacce SSD classe Enterprise a confronto

Protocollo seriale robusto

Il protocollo seriale SAS fornisce la formazione per i trasmettitori e i ricevitori seriali. Così si migliora la qualità del segnale sul cavo o sulla scheda madre grazie alla compensazione della lunghezza del canale, della differenza dell'impedenza e dell'interferenza tra simboli. Il protocollo seriale SAS gestisce inoltre il rilevamento degli errori e la ritrasmissione a livello hardware. Ciò significa che il recupero in seguito a problemi di segnalazione intermittente è più veloce.

Supporto di più host

L'interfaccia SAS e lo switching fabric consentono a più host di accedere allo stesso dispositivo. Questa funzione può essere utilizzata per gestire gli errori dell'host, nonché gli errori dei percorsi di dati a favore di una migliore disponibilità delle informazioni.

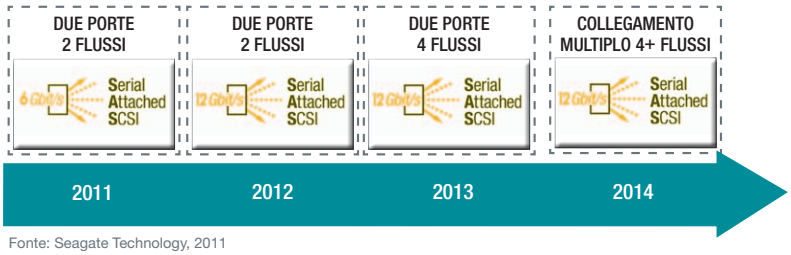
Integrità dei dati end-to-end

L'interfaccia SAS può verificare l'integrità dei dati tramite controlli CRC (Cyclic Redundancy Check) dei dati dal momento in cui vengono creati nel buffer dei dati host, tramite il trasferimento nell'interfaccia PCIe e nell'interfaccia SAS, fino a quando non vengono memorizzati sul dispositivo e di nuovo letti e trasferiti sul buffer. In questo modo è possibile utilizzare più punti di controllo lungo il percorso dalle applicazioni attraverso le schede di controllo RAID e sui dispositivi. A volta questa funzione viene detta protezione delle informazioni (PI).

Funzionalità con due porte

I dispositivi di destinazione SAS supportano il funzionamento con due porte. Ciò consente di creare due domini di errore e garantisce una disponibilità maggiore. Anche se si verifica un errore in uno dei percorsi di una porta che impedisce l'accesso lungo il percorso, è possibile accedere al dispositivo dalla seconda porta.

Storicamente Seagate ha promosso l'adozione dell'interfaccia nel mercato. Seagate collabora con SCSI Trade Association (STA) e altri leader dell'industria per sfruttare l'infrastruttura SAS esistente, distribuita su larga scala (Figura 2). Nella Tabella 1 viene illustrato come l'interfaccia SAS da 12 Gbit/s e i collegamenti multipli avvantaggiano i costruttori di sistemi e le organizzazioni degli utenti finali.



Fonte: Seagate Technology, 2011

Figura 2. Evoluzione dell'interfaccia SAS

Interfaccia PCI Express

PCI Express (PCIe) è l'interfaccia fondamentale che collega i dispositivi periferici al processore host e tramite una scheda di controllo della memoria all'architettura di memoria nel sistema. Le interfacce SATA e SAS esaminate in precedenza si collegano tramite un'interfaccia PCIe (o adattatore host) al processore host e alla memoria.

Tabella 1. Vantaggi offerti da SAS e dall'utilizzo di più collegamenti SAS

Collegamenti multipli (larghezza di banda)	X4 (4x600 MB/s)
Energia a disposizione	25 W (2,5")
Latenza totale	Molto bassa
Protocollo con più host	**Sì**
Disponibilità elevata	**Sì (due porte)**
Scalabilità	**Eccellente**
Stack di protocolli robusto e consolidato	**Sì**
Manutenzione hot swap	**Sì**
Compatibile con software di gestione esistente	**Sì**

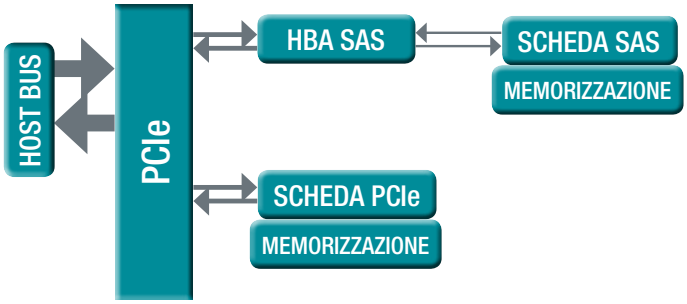
Fonte: Seagate Technology, 2011

L'interfaccia PCIe è un'implementazione seriale dell'interfaccia PCI originale che forniva il collegamento indirizzo parallelo/dati tra periferiche e processore host/memoria. L'interfaccia PCIe comunica su una o più *corsie* che consistono di un'interfaccia seriale di trasmissione e una di ricezione per ogni corsia. Per collegare un host a un dispositivo, è possibile utilizzare fino a 32 corsie. La velocità dei dati seriali in ogni corsia dipende dalla versione dello standard PCIe implementato. La versione corrente è 3.0 e la velocità di trasferimento dei dati è circa 1 GB/s.

Per un server 1U, l'interfaccia PCIe è stata progettata in modo da utilizzare un solo connettore su una scheda madre (client) o un adattatore ad angolo retto a due connettori su una scheda madre (server). È disponibile anche un sistema di cablaggio (anche se viene utilizzato raramente). Un server 2U, 4U o 7U è dotato di molti più slot PCIe, simili a implementazioni client. La specifica PCIe utilizza anche la formazione per il trasmettitore (e il ricevitore) per adeguare le variazioni di impedenza di una configurazione, ma è la soluzione ideale in quanto i canali di trasmissione sono di lunghezza inferiore rispetto a SAS.

Gli switch PCIe sono compatibili con la virtualizzazione I/O a radice singola (SR-IOV) e la virtualizzazione I/O multiradice (MR-IOV), metodi utilizzati per migliorare le prestazioni della scheda di controllo nei sistemi virtuali (ipervisori) con uno o più host. Solo adesso SR-IOV è generalmente disponibile negli adattatori. Tuttavia VMware potrebbe non sfruttarlo ancora. Solitamente MR-IOV non è supportato sugli adattatori.

Interfacce SSD classe Enterprise a confronto



Fonte: Seagate Technology, 2011
Figura 3. Evoluzione dell'interfaccia SAS

I dispositivi di memorizzazione che si collegano tramite l'interfaccia PCIe lo fanno tramite un collegamento di registro diretto o tramite un adattatore host che quindi si collega al dispositivo tramite un cablaggio aggiuntivo o un'interfaccia tipo scheda madre.

Attualmente vi sono numerose implementazioni diverse di entrambe le architetture. SATA utilizza un'implementazione HBA (Host Bus Adapter) nel chipset del sistema *(southbridge)*, AHCI Intel o AMD, che richiede driver AHCI diversi, ma si mappa a implementazioni legacy IDE compatibili. Queste interfacce implementano inoltre varie funzioni di gestione RAID.

SAS ha più venditori di HBA, con expander aggiuntivi e schede di controllo RAID disponibili, che utilizzano driver di dispositivi proprietari e il BIOS per soddisfare varie esigenze in termini di prestazioni e configurabilità.

L'interfaccia driver-scheda di controllo PCIe è implementata nella specifica NVM Express e nella specifica SCSI su PCIe (SOP) proposta.

L'architettura consolidata SATA descritta sopra è un altro esempio del collegamento di registro diretto PCIe.

Unità SSD PCIe oggi

Oggi sul mercato sono disponibili due tipi principali di unità SSD PCIe: nativa e aggregatrice. La scheda di controllo nativa si collega al bus PCIe host e controlla direttamente più bus di memoria Flash. Solitamente questi utilizzano un'interfaccia software proprietaria del produttore, impiegata solo per il dispositivo specifico. Alcune di queste implementazioni affidano la conversione degli indirizzi e altre funzioni alla CPU host e alla memoria. Ne consegue una riduzione delle risorse di sistema per le applicazioni quando i dispositivi vengono utilizzati con carichi di lavoro pesanti. Inoltre, essendo relativamente nuove sul mercato, a volte queste combinazioni uniche di unità e hardware sono soggette a instabilità in quanto gli ecosistemi sono ancora in evoluzione.

Il modello aggregatore adotta un approccio diverso al design. Utilizza una scheda di controllo RAID SAS o SATA esistente, a cui sono collegate più unita SSD SAS o SATA. Queste sono combinate insieme su un'unica scheda PCIe. La scheda di controllo RAID aggrega le prestazioni di più dispositivi per offrire prestazioni di alto livello. Basati su interfacce software e hardware classe Enterprise consolidati esistenti, questi design sono molto stabili e maturi. Inoltre questi design utilizzano schede di controllo intelligenti che eseguono conversioni di indirizzi e altre funzioni, consentendo l'uso completo di cicli di CPU del sistema e della memoria da parte delle applicazioni, anche con carichi di lavoro I/O pesanti.

Futuro delle unità SSD PCIe

Gli approcci SOP e NVMe sono simili da un punto di vista dell'architettura. Tuttavia NVMe viene sviluppato in un gruppo di lavoro industriale, mentre SOP viene sviluppato in un forum di standard aperti riconosciuto. NVMe è la soluzione ideale solo per l'utilizzo con i dispositivi di memoria non volatile, mentre SOP è la soluzione ideale anche per l'uso con gli adattatori HBA e le schede di controllo RAID con funzioni di potenziamento tra vari dispositivi SOP. Inoltre SOP sfrutta molto le architetture e le funzioni industriali esistenti, mentre NVMe usa un nuovo set di istruzioni molto limitato e un'interfaccia di coda.

Vantaggi e problemi dell'interfaccia

Ciascuna delle architetture di memorizzazione descritte offre dei vantaggi, ma presenta anche alcuni problemi. A seconda del design complessivo del sistema, i vantaggi derivanti dall'uso di un'architettura specifica possono essere superiori alle problematiche associate all'architettura. Per prendere la decisione appropriata, è necessaria un'analisi attenta. Nel prendere una decisione si deve inoltre considerare la compatibilità con un design di sistema esistente.

Ad esempio, se si aggiorna un computer portatile dotato di un'unità disco SATA da 2,5" con un'unità SSD, l'aggiornamento funziona solo con un'unità SSD delle stesse dimensioni fisiche e con la stessa interfaccia SATA (o una più recente). In questo caso vi sarà un limite sulla velocità dell'unità SSD. Se si supera la velocità dell'interfaccia SATA host esistente, le prestazioni del sistema non risultano migliori.

In una situazione simile, un server classe Enterprise che utilizza un'unità disco SAS da 15.000 giri/min con distanza dell'attuatore ridotta per memorizzare un indice di database può essere aggiornato con un'unità SSD SAS. Le prestazioni complessive del sistema aumentano, ma solo nella misura in cui un altro fattore di sistema diventa il nuovo collo di bottiglia (CPU, memoria, rete, adattatori, ecc.).

In una nuova architettura di sistema, l'aggiunta di memoria a stato solido può migliorare notevolmente le prestazioni del sistema, ma solo nella misura in cui il resto dell'architettura del sistema può supportare l'aumento della velocità di trasferimento e della larghezza di banda dei dati. Le velocità di trasferimento dei dati migliori delle unità SSD richiedono un'alimentazione del dispositivo superiore e una dissipazione del calore maggiore ovunque venga montata l'unità SSD.

Interfacce SSD classe Enterprise a confronto

Tabella 2. Unità SSD PCIe native e aggregatrici a confronto		
	Nativa	**Aggregatrice**
Comandi/Trasporto	Proprietario (FTL[1] in host/memoria principale) 	SCSI o SATA (unità SSD multiple, scheda di controllo) 
Comitato	Nessuno	Nessuno
Base su standard	No	**Sì**
Prestazioni con Flash	Alto	Alto
Carico di lavoro CPU	Alto	**Basso**
Latenza con coda corta	Molto bassa	Bassa
Latenza con coda lunga	Moderata	**Bassa**
Supporto per estensione	No	**Sì** (RAID, HBA, ecc.)
Maturità	In evoluzione	**Basata su architetture industriali consolidate**
Set di funzioni classe Enterprise (PI, sicurezza, gestione, ecc.)	No	**In base all'implementazione**

[1] FTL: Flash Translation Layer
Fonte: Seagate Technology, 2011

Tabella 3. Unità SSD PCIe SOP e NVMe a confronto		
	SOP[1]	**NVMe[2]**
Comandi/Trasporto	SOP/PQI[3] (FTL in scheda di controllo) 	NVMe/NVMe (FTL in scheda di controllo) 
Comitato	**T10/INCITS[4]**	Gruppo di lavoro industriale
Base su standard	**Sì (ANSI/ISO)**	No
Prestazioni con Flash	Molto elevate	Molto elevate
Carico di lavoro CPU	Basso	Basso
Latenza con coda corta	Molto bassa	Molto bassa
Latenza con coda lunga	Bassa	Bassa
Supporto per estensione	**Sì** (RAID, HBA, ecc.)	No (solo NVM)
Maturità	**Basata su architetture industriali consolidate**	Da determinare
Set di funzioni classe Enterprise (PI, sicurezza, gestione, ecc.)	**Supporto completo**	Limitato

[1] SOP: SCSI su PCI Express
[2] NVMe: Memory Express non volatile
[3] Interfaccia di coda PCIe
[4] INCITS: International Committee for Information Technology Standards
Fonte: Seagate Technology, 2011

Altri fattori sono la tempistica della disponibilità dei driver dei dispositivi del sistema operativo, il supporto del BIOS di queste nuove interfacce SSD e l'affidabilità iniziale del software.

Interfacce e latenza delle unità SSD Flash

Vi sono molte idee errate riguardo a quali fattori aggiungono latenza e quanto questi incidono effettivamente sulle prestazioni dell'applicazione. Quando si esamina questo aspetto, è importante concentrarsi sulla visione totale, non solo su una parte.

I fattori principali che contribuiscono alla latenza nelle unità SSD sono le parti Flash. I tempi di accesso SLC sono 25 µs+, mentre i tempi di accesso MLC sono 50 µs+. Nessuno dei due comporta un conflitto di accesso. Man mano che la lunghezza della coda aumenta, il conflitto di accesso alle parti Flash può incrementare notevolmente la latenza.

Una volta che la parte Flash avvia l'accesso, le altre richieste alla stessa parte devono attendere. Un massimo di otto die Flash condividono un bus comune. Ciò significa che i die devono attendere il loro turno per usare il bus. Le attività di gestione aggiungono ulteriore latenza (conversione di indirizzi, raccolta rifiuti, livello di usura, ecc.).

Un altro aspetto è il sistema operativo, che aggiunge latenza indipendentemente dal protocollo di accesso e dall'intercollegamento. Sono inclusi il file system, l'archiviazione di volumi, i driver di classe e i carichi di lavoro di scambio di contesto.

Le differenze nei protocolli e negli intercollegamenti hanno effetti irrilevanti sulla latenza così come vista da un'applicazione (frazioni di un microsecondo).

www.seagate.com

NORD E SUD AMERICA Seagate Technology LLC 10200 South De Anza Boulevard, Cupertino, California 95014, Stati Uniti, +1 408 6581000
ASIA/AREA DEL PACIFICO Seagate Singapore International Headquarters Pte. Ltd. 7000 Ang Mo Kio Avenue 5, Singapore 569877, +65 64853888
EUROPA, MEDIO ORIENTE E AFRICA Seagate Technology SAS 16–18, rue du Dôme, 92100 Boulogne-Billancourt, Francia, +33 1 41861000