



Il tiering automatizzato dello storage

Come prendere da soli la giusta decisione

Nello storage il posizionamento e il prezzo contano davvero

Come si dice nel mercato immobiliare, i tre principali fattori che determinano il prezzo sono la posizione, la posizione e la posizione. Lo stesso vale per il collocamento dello storage. Eventuali scelte sbagliate nel posizionamento dei dati vi costeranno care, mentre scegliendo con oculatezza il risparmio sui costi può diventare rilevante.

Attenendosi a due principi guida non si può sbagliare: a) dedicate il vostro spazio di archiviazione più veloce ai carichi di lavoro che sono più sensibili al fattore tempo e b) non sprecate lo spazio di archiviazione più costoso per contenuti a cui dovete accedere solo raramente.

Purtroppo non tutto è così facile. La verità è che al contrario di quanto avviene in occasione dell'acquisto di un'abitazione, lo schema con cui gli utenti accedono ai dati cambia frequentemente. I contenuti che solo ieri venivano considerati interessanti, oggi possono essere obsoleti. E non c'è modo di avere il tempo di raccogliere informazioni utili a tenere il passo con questi volubili cambiamenti. In questo caso una buona conoscenza della tecnologia e un po' di automazione possono essere d'aiuto.

Il giusto compromesso tra costo e prestazioni

La scienza del tiering automatizzato dello storage arriva ad analizzare i comportamenti dell'I/O, determinando la frequenza d'uso e spostando poi dinamicamente i blocchi di informazioni sul dispositivo o sul livello di storage della categoria più adatta. Il software DataCore™ SANsymphony™-V "promuove" automaticamente i blocchi utilizzati più frequentemente al livello più veloce, mentre i blocchi utilizzati con frequenza minore vengono "retrocesi" al livello più lento. Tutto il resto sta nel mezzo.

Ovviamente possono esserci delle eccezioni, specialmente quando bisogna assegnare dello storage ad alte prestazioni a volumi utilizzati raramente, magari per le elaborazioni di fine trimestre. In questi casi, si possono assegnare volumi specifici (dischi virtuali) a un livello predefinito, oppure definire una "affinità" con un particolare livello. Solo quando quel livello è completamente esaurito verrà utilizzato un livello inferiore.

L'aspetto economico dei livelli nei pool di storage virtuale

Per poter apprezzare le ragioni economiche dell'auto-tiering, basti considerare le differenze nel rapporto prezzo/prestazioni tra le tre più popolari categorie di dischi oggi utilizzate. In testa, con le prestazioni e il prezzo più elevati, ci sono i dischi a stato solido (SSD), seguiti dalle unità Serial Attached SCSI (SAS) e infine dai più economici dischi Serial ATA (SATA).

Nelle operazioni di lettura casuale, gli SSD sono da 25 a 100 volte più veloci dei dischi fissi SAS (HDD), con un costo per gigabyte più elevato da 15 a 20 volte. In pratica, gli SSD riducono sostanzialmente il numero di HDD necessari per schemi di utilizzo I/O a casualità elevata. Una sola scheda di I/O SSD PCI equivale a grandi linee a 320 dischi fissi. Questo significa avere una quantità di hardware 300 volte inferiore da installare, mantenere, raffreddare e monitorare.

Le stime dicono che la capacità su disco dei data center si divide in tre categorie di dischi, come illustrato nella Figura 1:

- 5% SSD flash [più veloci, più costosi] >> Livello 1
- 35% HDD SAS [fascia media, prezzi abbordabili] >> Livello 2
- 60% HDD SATA [più economici, grande capienza] >> Livello 3

In generale, è inutile sprecare la capacità dei più costosi SSD per blocchi di dati a cui si accede solo di rado. Per questo, il software DataCore mantiene le informazioni meno utilizzate sui più economici dischi SATA, che offrono grande capienza e che corrispondono al Livello 3. Nel momento in cui gli algoritmi di auto-tiering verificano un uso intensivo di determinati blocchi, li spostano immediatamente sugli SSD di Livello 1.

- Gli SSD riducono sostanzialmente il numero di HDD necessari a gestire schemi di I/O a casualità elevata.
- Gli HDD sono più adatti a schemi di I/O seriali.

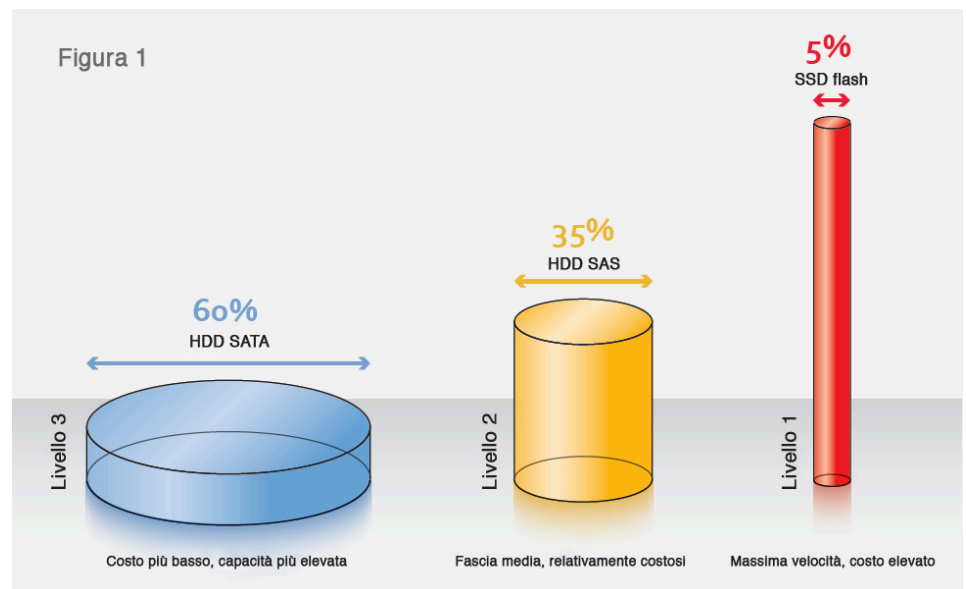
Al contrario delle limitate e spesso rigide classificazioni definite all'interno delle enclosure per i dischi, il software per la virtualizzazione dello storage di DataCore consente di definire i livelli in base all'indice prezzo/prestazioni definito dall'utente. A volte i vari pool di storage sono composti da batterie di costosi dischi di una particolare marca, da sottosistemi RAID di fascia media di un altro fornitore e da una grande quantità di dischi non "intelligenti" e relativamente economici (JBOD). Con DataCore è possibile definire da che cosa è costituito ciascun livello in base agli altri componenti del

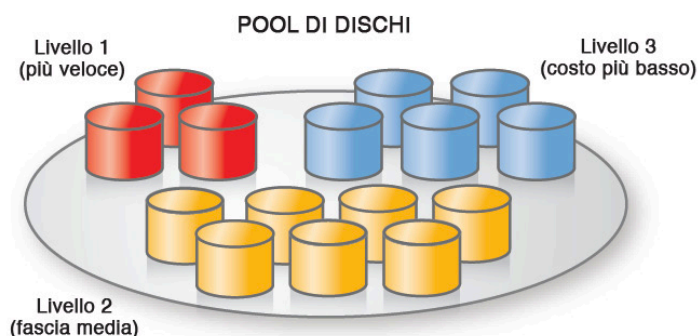
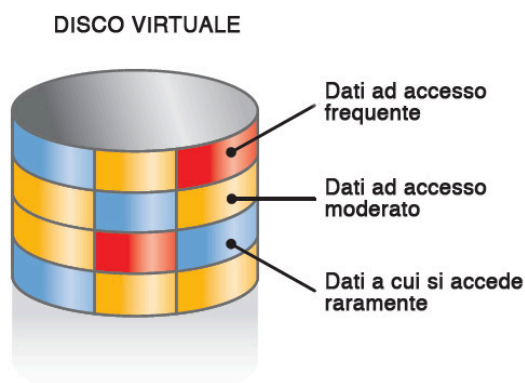
pool. Si possono impostare fino a 16 livelli differenti, anche se di solito ne bastano tre o quattro per raggiungere ottimi risultati.

Con l'arrivo sul mercato di nuovi prodotti di storage, quelli che oggi appartengono alla categoria più elevata possono essere riassegnati a un livello inferiore per far posto a tecnologie di disco con prestazioni ancora più elevate, anche se più costose.

Struttura gerarchica addio

È importante chiarire una cosa: al contrario della classica vecchia gestione gerarchica dello storage (HSM, Hierarchical Storage Management), l'auto-tiering non si occupa di suddividere i documenti per la memorizzazione nearline o offline, per il backup a lungo termine e per la definitiva archiviazione. Questa tecnologia opera esclusivamente sullo storage





su disco attivo e online, operando a livello di blocco o di gruppi di blocchi (chunk): non ha nulla delle strutture dei file system o dei database, né sarebbe utile che le avesse.

Una mappa per capire

Un modo per visualizzare il funzionamento dell'auto-tiering è quello di immaginare lo spazio di archiviazione del data center come una serie di piccoli elementi disposti in una griglia. Invece di avere una corrispondenza tra un quadrato e un'unità a disco, in questo esempio utilizzeremo 128 MB per ciascun elemento. Questa tecnica è conosciuta come auto-tiering "sub LUN" (Logical Unit Number). Se colorassimo questa mappa in base agli elementi più attivi, vedremmo certe aree evidenziate in rosso incandescente (a indicare un uso intenso), altre in un freddo blu (quelle meno utilizzate) e in giallo tutto quello che sta in mezzo. SANsymphony-V lavora silenziosamente dietro le quinte per spostare al livello più alto i blocchi che si trovano nell'area più "calda", facendo retrocedere gli elementi blu ai livelli inferiori.

La gestione adattativa per prestazioni ottimali

È importante notare che anche se è possibile avere un database di grandi dimensioni mappato su uno o più dischi

fissi, l'auto-tiering opera a un livello più granulare. Il software promuove a un livello superiore solamente quegli elementi del database che meritano dischi più veloci. L'amministratore, di norma, non può realizzare una gestione a questo livello di dettaglio.

Questo comportamento adattativo si traduce in migliori risposte applicative per quei carichi di lavoro che sono più sensibili al fattore tempo.

Ancora più veloci grazie al "caching"

Nei dibattiti che riguardano l'auto-tiering si sente spesso fare una netta distinzione tra migrazione e caching, come se una cosa escludesse l'altra. DataCore offre entrambe le possibilità, permettendo eventualmente di ignorare queste tecniche, quando le condizioni lo richiedono, per specifici dischi virtuali.

Grazie anche a questo punto di forza, la tecnica di caching adattativo di SANsymphony-V permette di migliorare le prestazioni dell'intero pool di storage, con una sorta di effetto turbo sull'I/O dei dischi. Le cache velocizzano le richieste di lettura e scrittura. I blocchi che vengono letti più di frequente (con una scala ridottissima: solo 4 KB) rimangono nella cache per ridurre il carico di lavoro dei dischi sottostanti e ridurre i tempi di

latenza dell'I/O. Le scritture casuali traggono beneficio dalla fusione di più operazioni per ottenere un I/O su disco più sequenziale.

La tecnica del caching si dimostra preziosa indipendentemente dal livello nel quale si trovano i blocchi, dato che tende a reagire molto più velocemente ai picchi a breve termine.

Non servono interventi manuali

Per gli amministratori di sistema, la parola operativa nel tiering automatizzato dello storage è "auto". Policy di semplice impostazione permettono di non effettuare modifiche all'infrastruttura, che spesso si rivelano fonte di stress e di perdite di tempo, e di evitare così futuri tentativi di migliorare la distribuzione dei carichi di lavoro sui dischi. Tutto avviene mantenendo comunque il controllo attraverso la definizione dei livelli e dei profili. Questi determinano il modo in cui il software gestisce le preferenze nel rapporto prezzo/prestazioni scelte dall'utente, che possono in ogni caso essere adattate in funzione delle variazioni operative.

Da ambienti IT tradizionali a quelli completamente virtualizzati

I progetti per la virtualizzazione di server e desktop (VDI) hanno spinto fortemente verso l'automazione del tiering dello storage, soprattutto perché le caratteristiche dei carichi di lavoro sono così differenziate da mettere in crisi i tradizionali sistemi di ottimizzazione nell'uso dello spazio su disco basati sul concetto dei silos. In passato, quando ogni server era dedicato a una specifica applicazione, era ragionevolmente semplice prevedere e modellare il mix di storage più adeguato alle necessità. Era infatti possibile decidere in modo intelligente e semi-permanente a quali carichi di lavoro riservare le tecnologie di disco più costose. Farlo con 1.200 macchine virtuali che condividono una SAN strutturata su livelli diversi è praticamente impossibile.

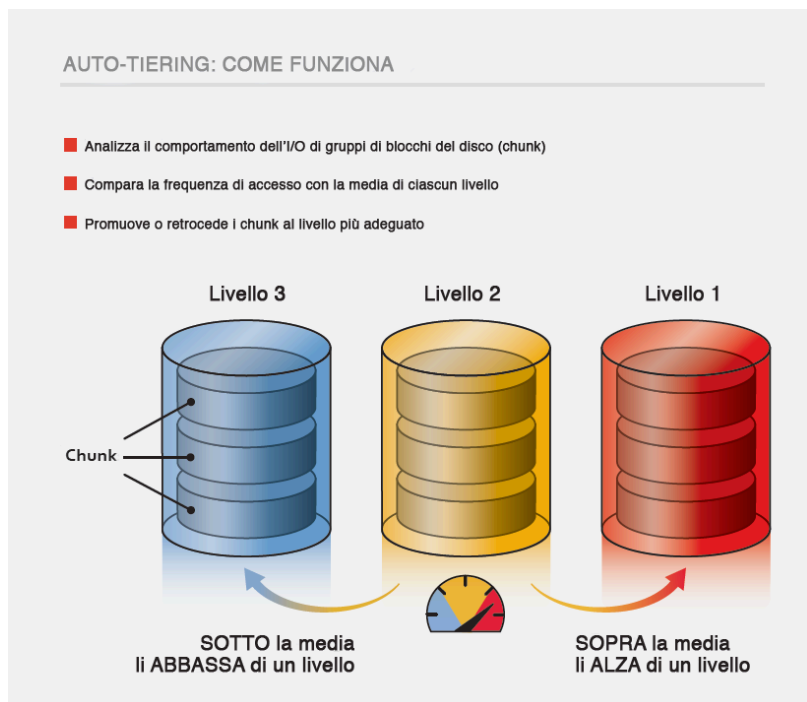
Detto questo, i clienti DataCore possono trarre vantaggio dall'auto-tiering in qualunque scenario possibile di data center, dagli ambienti IT completamente virtualizzati alle cloud pubbliche o private, fino alle più tradizionali configurazioni fisiche. Inoltre, è importante notare che SANSymphony-V automatizza la gestione dei livelli non solo sfruttando differenti tecnologie di disco, ma anche attraverso dispositivi di storage diversi, spesso incompatibili tra loro e indipendentemente dalla marca.

Niente hardware specifico

Invece di confinarlo all'interno di un sistema di storage, DataCore implementa il suo software di auto-tiering esternamente, senza la necessità di hardware specifico. Le più recenti tecnologie SSD possono coesistere accanto ai dischi fissi tradizionali. SANSymphony-V ne prende il controllo con uguale semplicità.

Mai più senza

Oggi lo storage rappresenta il singolo costo più elevato dell'intera infrastruttura IT ed è l'elemento più critico in assoluto per le prestazioni complessive degli ambienti virtuali. Per questo una gestione intelligente dell'allocatione dello spazio su disco è fondamentale. La funzionalità di tiering automatico dello storage integrata in DataCore SANSymphony-V, abbinata a una completa suite di servizi per la virtualizzazione dello storage indipendenti dall'hardware, aiuta a massimizzare l'utilizzo, le prestazioni e la disponibilità di questi preziosi asset.



0611

Informazioni su DataCore Software

DataCore Software sviluppa il software per la virtualizzazione dello storage necessario per ottenere il più alto livello di disponibilità, le prestazioni più elevate e il massimo utilizzo delle risorse di storage negli ambienti IT fisici e virtuali. La società offre la fondamentale terza dimensione su cui si basa il successo dei progetti di virtualizzazione di server e desktop, indipendentemente dai modelli e dalle marche dei dispositivi di storage utilizzati.

Per maggiori informazioni,
visitate il sito: www.datacore.com
o inviate un'e-mail a: info@datacore.com