

# Dell PowerMax: riduzione dei dati

## Compressione e deduplica in linea

Luglio 2022

H19254

## White paper

### Abstract

Le piattaforme di storage PowerMax offrono numerose tecniche di riduzione dei dati, come la compressione e la deduplica in linea. Tra di esse, vi sono il rilevamento dei modelli e il posizionamento efficiente dei dati per offrire un ottimo equilibrio tra prestazioni ed efficienza.

Dell Technologies

## Copyright

Le informazioni contenute nella presente documentazione vengono fornite "così come sono". Dell Inc. non fornisce alcuna dichiarazione o garanzia in relazione alle informazioni contenute nel presente documento, in modo specifico per quanto attiene alle garanzie di commerciabilità o idoneità per uno scopo specifico.

L'utilizzo, la copia e la distribuzione dei prodotti software descritti in questo documento richiedono una licenza d'uso valida per ciascun software.

Copyright © 2022 Dell Inc. o sue società controllate. Tutti i diritti riservati. Dell Technologies, Dell, EMC, Dell EMC e altri marchi sono marchi di Dell Inc. o delle sue società controllate. Intel, il logo Intel, il logo Intel Inside e Xeon sono marchi di Intel Corporation negli Stati Uniti e/o in altri Paesi. Gli altri marchi sono di proprietà dei rispettivi titolari.

Pubblicato negli Stati Uniti, luglio 2022, H19254.

Dell Inc. ritiene che le informazioni contenute nel presente documento siano esatte alla data di pubblicazione. Le informazioni sono soggette a modifiche senza preavviso.

## Sommario

<b>Executive Summary .....</b>	<b>4</b>
<b>Riduzione dei dati .....</b>	<b>5</b>
<b>Uso delle risorse di sistema.....</b>	<b>9</b>
<b>Gestione e monitoraggio.....</b>	<b>9</b>
<b>Data service supportati.....</b>	<b>17</b>
<b>Conclu-sioni .....</b>	<b>18</b>
<b>Riferimenti .....</b>	<b>19</b>

## Executive Summary

### Panoramica

La riduzione dei dati con Dell PowerMax offre un incremento dell'efficienza del sistema grazie alla combinazione di compressione in linea, deduplica in linea e rilevamento dei modelli. L'utilizzo di queste tecniche di riduzione dei dati consente agli utenti di ottenere un notevole risparmio in termini di capacità. La riduzione dei dati comprime i dati e ne elimina le copie ridondanti. Il presente white paper illustra il funzionamento della riduzione dei dati nei sistemi PowerMax e descrive i processi di reporting tramite applicazioni di gestione Dell come Unisphere for PowerMax, Solutions Enabler e il software Mainframe Enabler.

### Revisioni

Data	Descrizione
Luglio 2022	Versione iniziale

### La tua opinione è importante

Dell Technologies e gli autori del presente documento sono sempre lieti di conoscere l'opinione dei propri clienti su questo documento. È possibile contattare il team Dell Technologies via [e-mail](#).

**Autore:** Robert Tasker

---

**Nota:** per i link ad altri documenti su questo argomento, consultare [PowerMax and VMAX Info Hub](#).

---

## Riduzione dei dati

### Panoramica

La riduzione dei dati combina compressione in linea, deduplica in linea, rilevamento dei modelli, posizionamento efficiente dei dati e apprendimento automatico (ML). Questa combinazione crea un sistema in cui gli utenti possono scrivere più dati host rispetto alla quantità totale di capacità fisica disponibile, pur senza rinunciare alle prestazioni di un enterprise storage system. È possibile abilitare o disabilitare questa funzione, attivata per impostazione predefinita, a livello di storage group. Inoltre, sono supportati tutti i data service disponibili nei sistemi PowerMax 2500 e 8500. Il supporto si applica anche all'emulazione CKD, ma non include la deduplica per CKD.

La compressione riduce le dimensioni dei dati, mentre la deduplica archivia i dati come singola istanza. Il rilevamento dei modelli include una funzione di allocazione diversa da zero, che esclude l'archiviazione di stringhe di zeri consecutivi come parte dei dati compressi. La compressione, la deduplica e il rilevamento dei modelli vengono eseguiti tramite l'assistenza hardware integrata all'interno del sistema per ridurre l'overhead associato all'esecuzione di queste funzioni. L'apprendimento automatico identifica i dati soggetti a maggiore traffico archiviati su disco e garantisce che rimangano invariati per ottenere prestazioni ottimali. Il posizionamento efficiente dei dati utilizza una funzione denominata **compattazione**, che permette di archiviare strategicamente i dati per ridurre al minimo lo spazio inutilizzato e la necessità di funzioni di garbage collection o deframmentazione.

### Riduzione basata sull'attività

La riduzione basata sull'attività (ABR) riduce i costi in termini di prestazioni sostenuti dalla decompressione dei dati a cui si accede di frequente. Questa funzione consente di archiviare fino al 20% dei dati soggetti a maggiore traffico nel sistema senza compressione. Una simile capacità offre numerosi vantaggi al sistema, in quanto riduce al minimo la latenza delle prestazioni derivante dalla decompressione costante dei dati a cui si accede di frequente. Per determinare quali sono i dati soggetti a maggiore traffico, il sistema utilizza algoritmi ML che elaborano le statistiche di I/O. Questa attività mantiene un ambiente bilanciato e ottimale per il risparmio e le prestazioni di riduzione dei dati.

### Compressione

La compressione riduce i carichi di lavoro di scrittura in ingresso alla minima dimensione possibile per consumare la minore quantità di capacità. I dati vengono compressi durante il trasferimento attraverso l'hardware per la riduzione dei dati, che utilizza l'algoritmo di compressione GZIP. Quando vengono trasferiti attraverso l'hardware per la riduzione dei dati, i dati vengono suddivisi in quattro sezioni, che vengono compresse in parallelo per ottimizzare l'efficienza. La somma delle quattro sezioni rappresenta la dimensione finale ridotta dei dati archiviati su disco. Questa capacità fornisce un accesso granulare ai dati ridotti. Solo le sezioni che contengono i dati necessari per le richieste parziali di lettura o scrittura vengono elaborate, in quanto ogni sezione può essere gestita in modo indipendente.

### Deduplica

La deduplica è un metodo di risparmio della capacità che identifica copie identiche di dati e archivia una singola istanza di ciascuna copia. La deduplica include alcuni componenti necessari per risparmiare capacità in modo efficiente.

- **ID hash:** l'ID hash è un ID univoco per i dati in ingresso che viene utilizzato per determinare se è necessaria una relazione di deduplica. Il sistema utilizza un algoritmo SHA-256 per generare l'ID hash.
- **Tabella degli ID hash:** le tabelle hash costituiscono un'allocazione della memoria di sistema distribuita tra System Director. Queste tabelle catalogano gli ID hash utilizzati dal processo di deduplica. Le voci della tabella vengono utilizzate per determinare se esiste una relazione di deduplica o se è necessaria una nuova voce e i dati possono essere archiviati su disco.
- **Object di gestione della deduplica (DMO):** il DMO è un object a 64 byte all'interno della memoria di sistema, presente solo quando esiste una relazione di deduplica. Questi object memorizzano

## Riduzione dei dati

e gestiscono i puntatori tra i dispositivi front-end e i dati deduplicati che utilizzano capacità di back-end nell'array. I DMO gestiscono i puntatori per i dati deduplicati tra i dispositivi front-end e i dati memorizzati su disco. Ciò consente anche di stabilire in quale tabella hash vengono archiviati gli ID hash in presenza di relazioni di deduplicazione.

La deduplicazione viene eseguita utilizzando lo stesso hardware di compressione anche per la riduzione dei dati, generando un ID hash univoco quando i dati vengono elaborati dall'hardware. L'ID hash viene quindi confrontato con la tabella degli ID hash alla ricerca dello stesso ID. Quando viene trovata una corrispondenza, i dati, che non vengono archiviati su disco, generano una condivisione di deduplicazione. I puntatori sono impostati tra il volume front-end e l'ID univoco (UID) nella tabella degli ID hash. I puntatori collegano la singola istanza dei dati archiviati su disco al volume, fornendo l'accesso futuro ai dati. DMO gestisce i puntatori tra i dati, i volumi front-end che accedono ai dati e la tabella degli ID hash. Quando non vi è alcuna corrispondenza nella tabella degli ID hash, viene aggiunta una nuova voce per il futuro confronto degli ID hash.

## Algoritmo di deduplicazione

I sistemi PowerMax utilizzano l'algoritmo di hashing SHA-256 implementato nell'hardware per la riduzione dei dati per trovare i dati duplicati. I dati vengono quindi archiviati come singola istanza condivisa da più fonti. Questo processo fornisce una maggiore efficienza dei dati, mantenendo al contempo l'integrità degli stessi.

L'algoritmo SHA-256 genera un codice a 32 byte per ciascun blocco di dati da 32 KB. Si consideri un sistema con 1 PB di dati scritti con una percentuale di aggiornamento del 5% al giorno. In un milione anni di operatività, sussiste il 20% di probabilità di collisione hash. Poiché ogni traccia da 128 KB viene gestita come quattro blocchi da 32 KB, si dovrebbe verificare una collisione hash su tutti e quattro i blocchi simultaneamente nella stessa traccia da 128 KB per un'effettiva collisione hash. Le probabilità di avere una collisione in tutti e quattro i blocchi rende questo scenario solo teorico (meno dell'1% di possibilità in un trilione di anni di operatività). Inoltre, quando viene trovata una corrispondenza durante la fase di confronto della deduplicazione, viene eseguito un confronto byte per byte. Ciò serve a confermare la corrispondenza prima di aggiornare le tabelle e impostare i puntatori per consentire l'accesso ai dati.

## Compattazione

Il posizionamento dei dati viene eseguito utilizzando un processo denominato compattazione. La compattazione posiziona in modo intuitivo i dati ridotti o non ridotti su disco nella migliore posizione disponibile. L'operazione di archiviazione dei dati su disco utilizza object di scrittura. Ogni object ha una capacità di dispositivi dati back-end contigua di 6 MB tra le unità configurate nel sistema. Gli object di scrittura sono allineati su limiti da 1 K e vengono utilizzati in sequenza in una singola volta. Gli object di scrittura vengono distribuiti su stripe complete per tutti i RAID type supportati al fine di ottimizzare le scritture. Ogni object supporta dati ridotti o non ridotti per l'emulazione FBA e CKD.

- **Object di scrittura FBA:** un object di scrittura non ridotto è costituito da 48 percorsi FBA. Un object di scrittura ridotto è costituito da 1000 percorsi ridotti. La riduzione delle voci per gli object di scrittura va da 1 KB a 96 KB.
- **Object di scrittura CKD:** un object di scrittura non ridotto è costituito da 108 percorsi CKD. Un object di scrittura ridotto è costituito da 1000 percorsi ridotti. La riduzione delle voci per gli object di scrittura va da 1 KB a 52 KB.

## Compressione dei dati estesa

I sistemi PowerMax 2500 e 8500 includono una funzione aggiuntiva denominata compressione dei dati estesa (EDC), che comprime i dati già compressi per ottenere un ulteriore risparmio di capacità. Questa operazione viene eseguita identificando i dati che non hanno registrato accessi per un periodo di tempo prolungato. I fattori che rendono i dati idonei per l'EDC sono elencati di seguito:

- I dati devono appartenere a uno storage group abilitato per la riduzione dei dati.
- L'accesso ai dati non deve essere stato effettuato per 30 giorni.
- I dati non devono essere già compressi mediante EDC.

I dati idonei per l'EDC vengono compressi utilizzando l'algoritmo Def9\_128\_SW al fine di ridurre ulteriormente la quantità di capacità utilizzata per archiviare i dati. Si tratta di un processo in background automatico all'interno del sistema. Un ulteriore risparmio è incluso nel rapporto di compressione ottenuto a livello di storage group. L'EDC è disponibile solo con gli array di storage PowerMax.

## CKD compressione

La riduzione basata sull'attività (ABR) riduce i costi in termini di prestazioni sostenuti dalla decompressione dei dati a cui si accede di frequente. Questa funzione consente di archiviare fino al 20% dei dati soggetti a maggiore traffico nel sistema senza compressione. Questo risultato è vantaggioso per il sistema, in quanto elimina l'impatto negativo sulle prestazioni derivante dalla decompressione costante dei dati a cui si accede di frequente. Per determinare il livello di traffico dei dati, il sistema utilizza algoritmi ML che elaborano le statistiche raccolte dall'I/O in ingresso nei dispositivi front-end. Questa azione consente di mantenere il giusto equilibrio tra le risorse di sistema, garantendo un ambiente ottimale in termini sia di risparmi sia di prestazioni della riduzione dei dati.

La compressione riduce i carichi di lavoro di scrittura in ingresso alla minima dimensione disponibile per consumare la minore quantità di capacità possibile. I dati vengono compressi durante il trasferimento attraverso l'hardware per la riduzione dei dati integrato nel sistema, che utilizza l'algoritmo di compressione GZIP. Quando vengono trasferiti attraverso l'hardware per la riduzione dei dati, i dati vengono suddivisi in quattro sezioni, che vengono compresse in parallelo per ottimizzare l'efficienza dell'hardware. La somma delle quattro sezioni rappresenta la dimensione finale ridotta dei dati archiviati su disco. Questo risultato fornisce un accesso granulare ai dati ridotti quando è presente una richiesta di lettura o scrittura parziale. Solo le sezioni che contengono i dati richiesti vengono elaborate, in quanto ogni sezione può essere gestita in modo indipendente.

Il posizionamento dei dati viene eseguito utilizzando un relativo processo denominato compattazione. La compattazione posiziona in modo intuitivo i dati ridotti o non ridotti su disco nella migliore posizione disponibile. L'operazione di archiviazione dei dati su disco utilizza object di scrittura. Ogni object ha una capacità di dispositivi dati back-end contigua di 6 MB tra le unità configurate nel sistema. Gli object di scrittura sono allineati su limiti da 1 K e vengono utilizzati in sequenza in una singola volta. Gli object di scrittura vengono distribuiti su stripe complete per tutti i RAID type supportati per ottimizzare le scritture. Ogni object supporta dati ridotti o non ridotti. Un object di scrittura non ridotto è costituito da 108 percorsi CKD. Un object di scrittura ridotto è costituito da 1000 percorsi ridotti. La riduzione delle voci per gli object di scrittura va da 1 KB a 52 KB.

## Flusso di I/O della riduzione dei dati

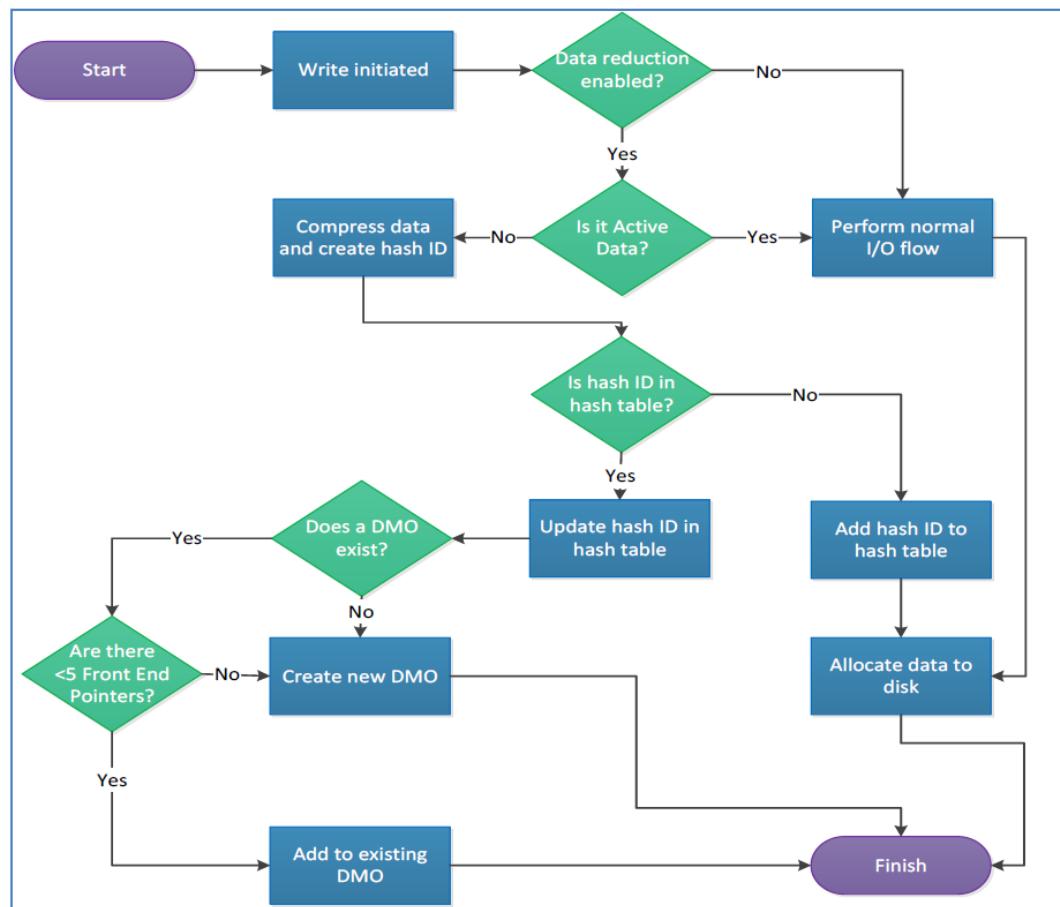
Tutti gli I/O passano attraverso la memoria cache e vengono quindi elaborati dal sistema. Le azioni di riduzione dei dati vengono eseguite dopo la ricezione dei dati da parte del sistema, ma prima che vengano posizionati sul disco. L'utilizzo di un processo in linea richiede controlli aggiuntivi all'interno del flusso di I/O in cui si applica la riduzione dei dati. Il sistema utilizza questi controlli per determinare se i dati in ingresso devono passare o meno attraverso l'hardware per la riduzione dei dati. I dati in ingresso per uno storage group con la riduzione dei dati abilitata seguiranno il flusso di riduzione dei dati. Tuttavia, a causa della funzione di riduzione basata sull'attività (ABR), i dati attivi per uno

storage group con la riduzione dei dati abilitata ignoreranno il flusso di riduzione dei dati per l'ottimizzazione delle prestazioni. I dati non compressi a causa di ABR possono essere compressi in un secondo momento e spostati in un pool di compressione. I dati per uno storage group con la riduzione dei dati disabilitata ignoreranno il flusso di riduzione dei dati e verranno scritti nel sistema senza alcuna riduzione.

Esistono diversi tipi di I/O da considerare: lettura, scrittura e scrittura-aggiornamento.

- **Lettura:** richiesta di accesso ai dati che già popolano l'array.
- **Scrittura:** I/O in ingresso che utilizzerà spazio su disco.
- **Scrittura-aggiornamento:** I/O in ingresso che può modificare i dati allocati allo spazio su disco nell'array.

Nella figura riportata di seguito viene descritto il percorso che verrà seguito dall'I/O, determinato dalle caratteristiche del dataset o dello storage group correlato.



**Figura 1. Flusso di I/O di riduzione dei dati per gli enterprise storage system PowerMax**

## Uso delle risorse di sistema

### Panoramica

Capacità e memoria sono le due risorse principali configurate in ogni sistema PowerMax. La capacità è suddivisa nelle seguenti categorie: fisica, effettiva e di provisioning. Le memorie sono classificate come risorse di sistema.

Le risorse di memoria supportano le strutture di metadati per la capacità di provisioning e per la capacità fisica. La quantità di capacità effettiva disponibile è correlata alla quantità di capacità fisica, alla quantità di risorse di sistema disponibili e alla riducibilità dei dati scritti nel sistema. I dati scritti altamente riducibili consumano meno capacità fisica, il che determina una maggiore capacità effettiva. È vero anche il contrario: i dati scritti che non sono riducibili possono comportare una minore capacità effettiva disponibile. Le informazioni descritte nella sezione [Riduzione dei dati](#) (capacità, risorse di sistema) sono disponibili all'interno delle applicazioni di gestione utilizzate per i sistemi PowerMax 2500 e 8500, Unisphere for PowerMax, Solutions Enabler e il software Mainframe Enabler. Unisphere for PowerMax è un'interfaccia utente (UI) che fornisce dati sotto forma di grafici, diagrammi ed elenchi. Solutions Enabler è un'interfaccia standard a riga di comando che fornisce gli stessi dati, ma non sotto forma di grafici. Mainframe Enablers è una suite di componenti che monitorano e gestiscono i sistemi di storage Dell in un ambiente mainframe. Le immagini mostrate nelle sezioni successive del presente documento illustrano il funzionamento di Unisphere for PowerMax per la gestione di un sistema PowerMax 2500 o 8500.

### Capacità fisica

La capacità fisica è la quantità di spazio su disco configurata nel sistema in base ai dischi installati e alla protezione RAID applicata. In una configurazione in cui la riduzione dei dati non è in uso, la capacità fisica rappresenta la quantità totale di capacità disponibile per i dati host. Ad esempio, se un sistema ha 100 TB di capacità fisica, ciò significa che può ospitare 100 TB di dati host che non utilizzano la riduzione dei dati.

### Capacità effettiva

La capacità effettiva è la quantità di spazio disponibile quando la riduzione dei dati è in uso. La quantità totale all'installazione iniziale dipende dalla quantità di memoria configurata nel sistema e si basa su un risparmio di riduzione dei dati predefinito di 4:1 (3:1 per l'emulazione CKD).

Ad esempio, lo stesso sistema da 100 TB di capacità fisica indicherà 400 TB di capacità effettiva. Il valore di 400 TB è un punto di partenza di capacità effettiva e cambierà man mano che i dati vengono scritti nel sistema e la riduzione dei dati viene applicata.

### Capacità di provisioning

La capacità di provisioning è la rappresentazione della capacità disponibile sotto forma di dispositivi creati e presentati agli host e alle applicazioni destinati a consumare capacità fisica o effettiva nel sistema.

## Gestione e monitoraggio

### Panoramica

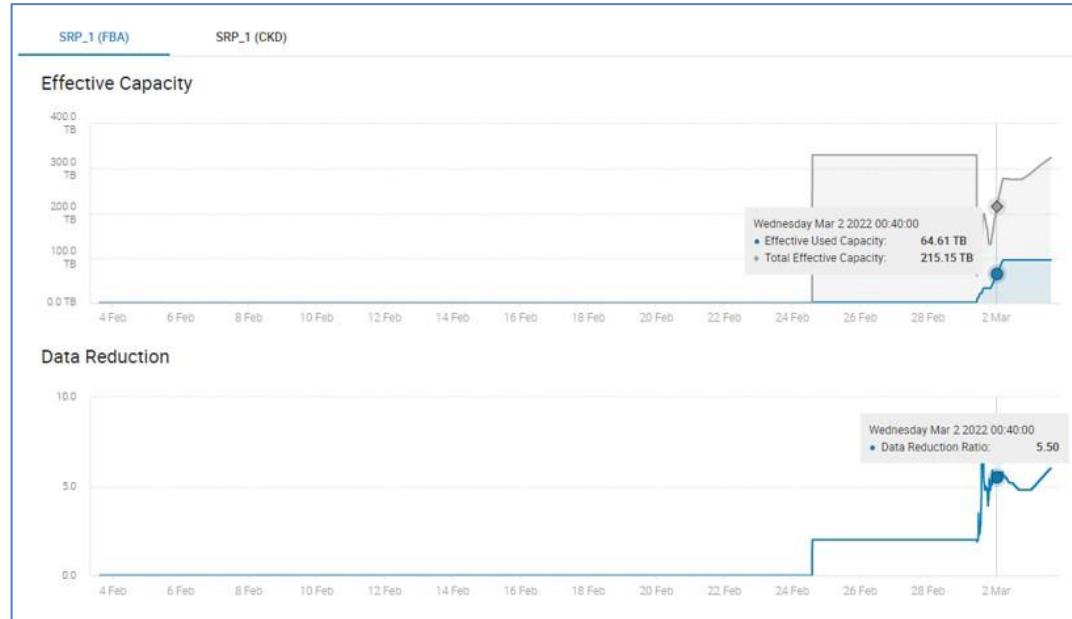
Unisphere for PowerMax è un'interfaccia utente utilizzata per gestire e monitorare l'uso di capacità e risorse del sistema. A livello di sistema, le informazioni relative a uso della capacità, riduzione dei dati e risorse di sistema vengono visualizzate nel dashboard di capacità. Dal dashboard di capacità, gli utenti possono accedere a schermate che visualizzano informazioni sulla capacità effettiva e di provisioning, sulla capacità delle istantanee, sulla riduzione dei dati e sulle risorse di sistema.

### Dashboard di capacità

All'interno di Unisphere for PowerMax, sono disponibili più schermate che forniscono informazioni relative all'uso della capacità.

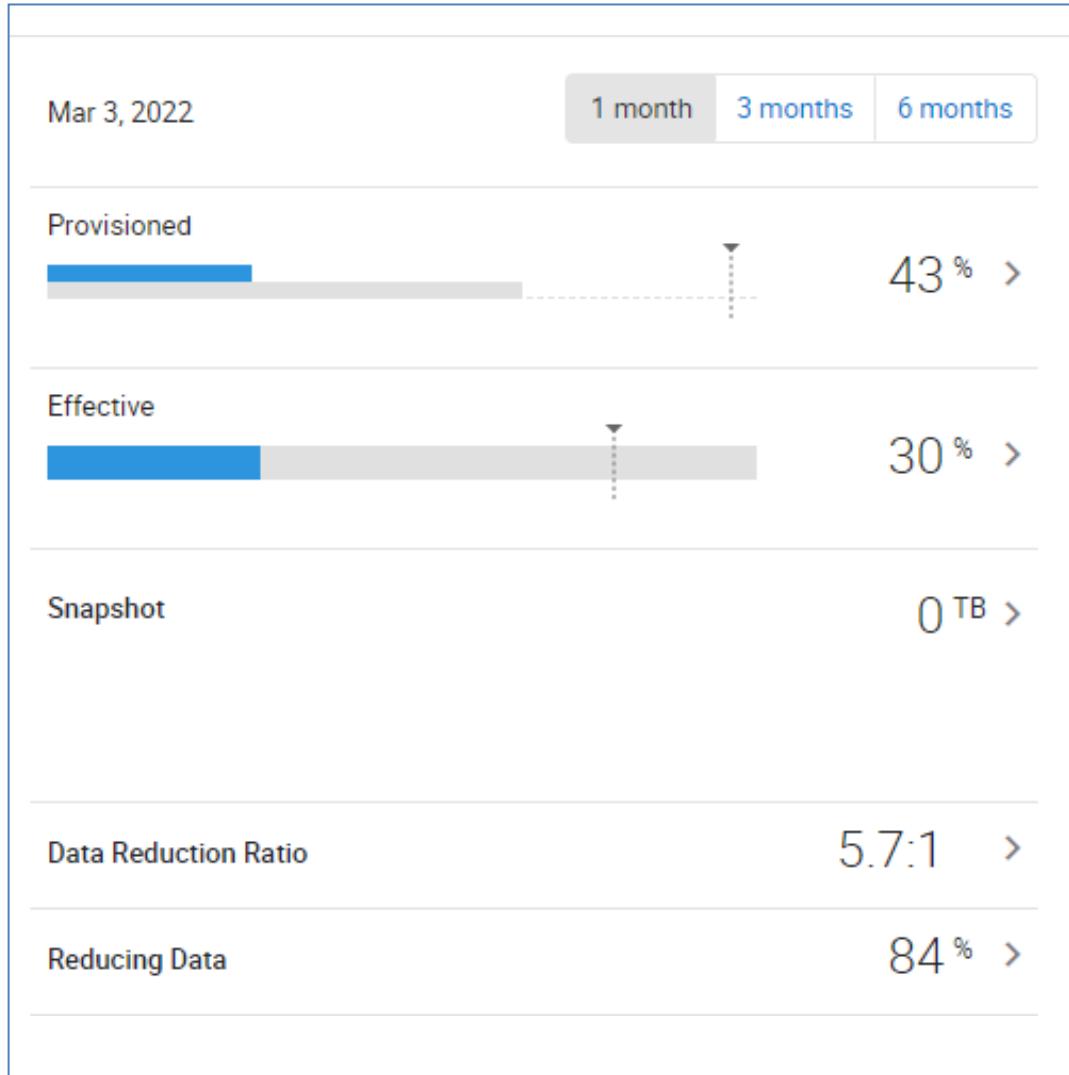
## Gestione e monitoraggio

La dashboard principale visualizza un grafico interattivo che mostra l'uso effettivo della capacità e la riduzione dei dati nel tempo. Questa schermata mostra la cronologia di uso della capacità effettiva e il rapporto di riduzione dei dati correlato alla capacità effettiva stessa. Queste informazioni possono essere utilizzate per monitorare e tenere traccia delle tendenze di uso della capacità effettiva in relazione al rapporto di riduzione dei dati mostrato. I sistemi PowerMax 2500 o 8500 possono essere configurati con l'emulazione FBA e CKD all'interno dello stesso pool di risorse di storage, sebbene il grafico con i dati cronologici sia specifico della vista di emulazione selezionata.



**Figura 2. Grafico con i dati cronologici del dashboard di capacità che mostra la capacità effettiva e la riduzione dei dati per l'emulazione FBA**

Il dashboard principale offre inoltre dati sotto forma di grafici a barre per capacità di provisioning, capacità effettiva, uso delle istantanee e riduzione dei dati. Ogni sezione può essere estesa a una schermata più dettagliata, che mostra i dati in modo granulare per ogni elemento.



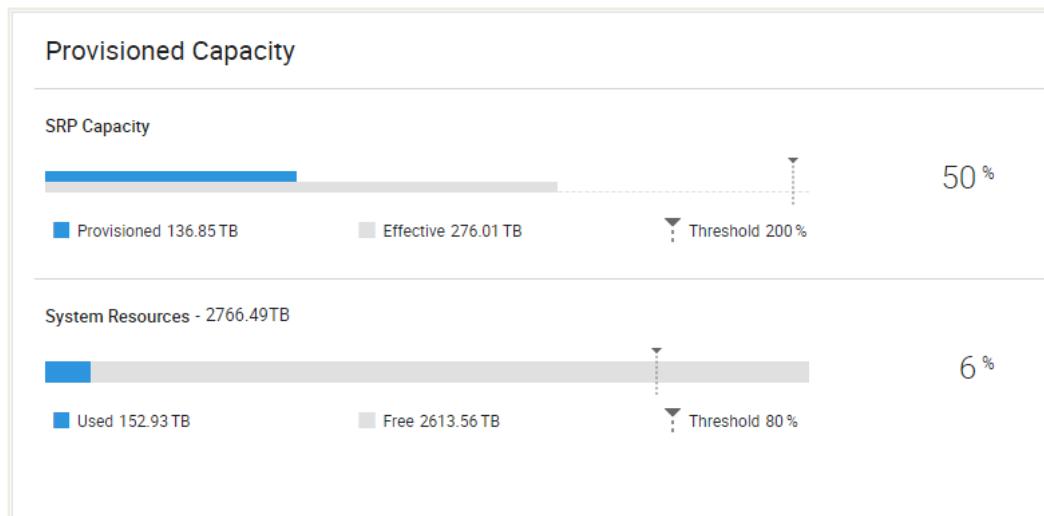
**Figura 3. Grafici a barre del dashboard di capacità per capacità di provisioning, capacità effettiva, uso delle istantanee e riduzione dei dati**

### Sottoposto a provisioning.

La capacità di provisioning è la quantità di capacità sottoposta a provisioning sotto forma di dispositivi presentati agli host e all'applicazione come capacità disponibile. La capacità di provisioning viene monitorata utilizzando due metriche: capacità SRP e risorse di sistema.

- **La capacità SRP** mostra la quantità di capacità di provisioning sotto forma di valore di TB sottoposti a provisioning e la quantità disponibile di capacità effettiva. La quantità di capacità SRP effettiva si basa inizialmente su un risparmio predefinito di riduzione dei dati di 4:1 (3:1 per l'emulazione CKD). Questa quantità viene calcolata utilizzando la quantità di capacità fisica configurata nel sistema. Man mano che gli utenti creano dispositivi sottoposti a provisioning, aumenta la capacità. Il valore mostrato rappresenta la percentuale di abbonamento, che viene calcolata utilizzando la quantità di capacità di provisioning e la quantità di capacità effettiva.

- **Le risorse di sistema** rappresentano la quantità di memoria disponibile per supportare la capacità di provisioning sotto forma di metadati. Il valore totale presentato cambierà solo se viene aggiunta memoria cache al sistema. La quantità utilizzata rappresenta questo valore. La quantità libera è un'indicazione della quantità di capacità aggiuntiva fornita che il sistema è in grado di supportare. Quando gli utenti creano dispositivi, la quantità di utilizzo aumenta. In caso di variazione tra i due valori, le risorse di memoria vengono utilizzate per supportare la riduzione dei dati o altre funzionalità che utilizzano memoria, ad esempio l'acquisizione di istantanee dei dispositivi esistenti.



**Figura 4. Uso delle risorse di sistema all'interno della schermata della capacità di provisioning**

## Capacità effettiva

La capacità effettiva rappresenta la quantità di capacità disponibile per l'utente sulla base del risparmio previsto derivante dall'utilizzo della riduzione dei dati. La schermata della capacità effettiva fornisce una vista dettagliata delle risorse fisiche ed effettive disponibili. Ciò è illustrato nelle tre sezioni seguenti: Capacità fisica, Risorse di capacità effettiva e Uso della capacità effettiva.

- La **capacità fisica** mostra la relativa quantità disponibile nei dischi rigidi configurati nel sistema. Le quantità mostrate rappresentano i valori dopo l'applicazione della formattazione e della protezione RAID. Il valore mostrato è la quantità di capacità supportata dal sistema per i dati host quando la riduzione dei dati non viene utilizzata.
- Le **risorse di capacità effettiva** indicano i valori ottenibili in base all'uso corrente delle risorse di sistema. Il valore delle risorse di capacità effettiva mostrato si adatterà al risparmio derivante dalla riduzione dei dati corrente e all'uso della capacità fisica ed effettiva.
- L'**uso della capacità effettiva** visualizza la quantità corrente di capacità effettiva disponibile in base all'uso delle risorse di sistema e al risparmio derivante dalla riduzione dei dati corrente. Il valore visualizzato all'interno del grafico circolare è la capacità effettiva attualmente disponibile. I valori riportati sulla destra suddividono l'uso in tre categorie: uso delle istantanee, utilizzo dell'utente e quantità libera.

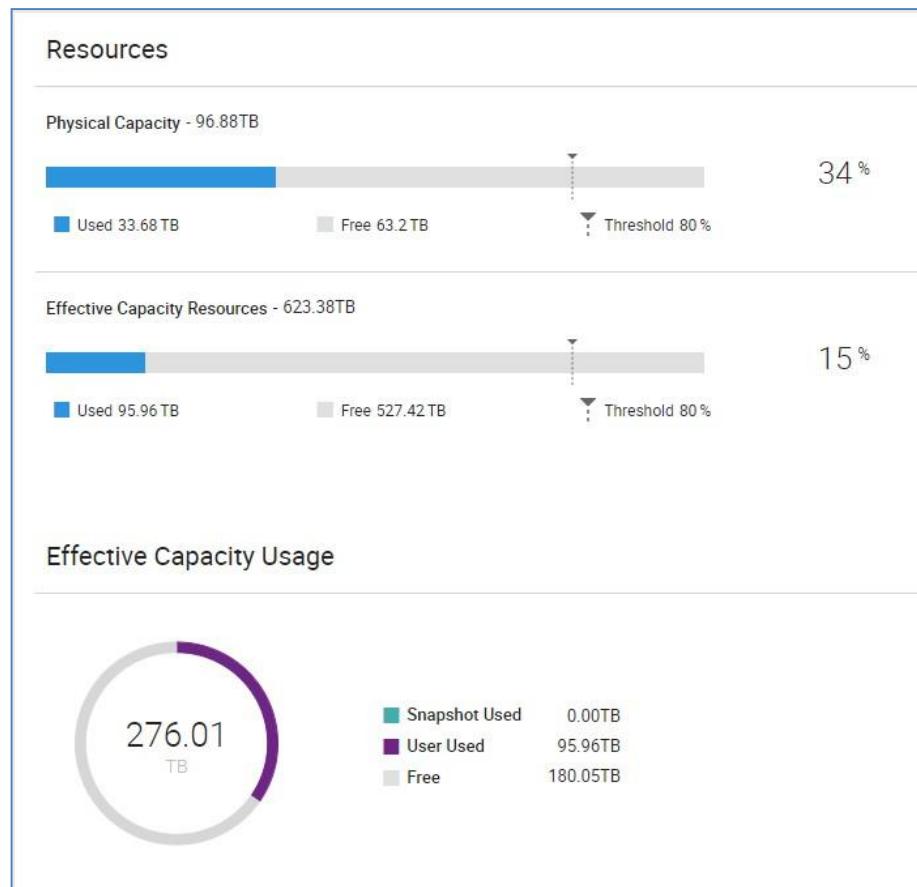


Figura 5. Uso delle risorse mostrato dalla schermata della capacità effettiva

## Snapshot

La capacità delle istantanee in back-end può essere significativamente inferiore rispetto ai delta per istantanea grazie all'efficienza di alcune funzionalità, come la condivisione delle allocazioni e la riduzione dei dati.

Per visualizzare i dettagli di alto livello, è possibile posizionare il puntatore del mouse sul grafico a barre delle istantanee nel dashboard di capacità. I valori delle istantanee sono definiti come:

- **Quantità utilizzata:** capacità effettiva utilizzata per i dati di modifica delle istantanee
- **Quantità libera:** quantità rimanente dei dati di modifica delle istantanee in base alla quantità utilizzata e ai metadati rimanenti
- **Quantità totale:** quantità utilizzata + quantità libera
- **Soglia:** soglia per l'invio di avvisi sui dati di modifica delle istantanee

Cliccare sui grafici a barre per passare al dashboard di capacità effettiva e al dashboard di capacità delle istantanee.

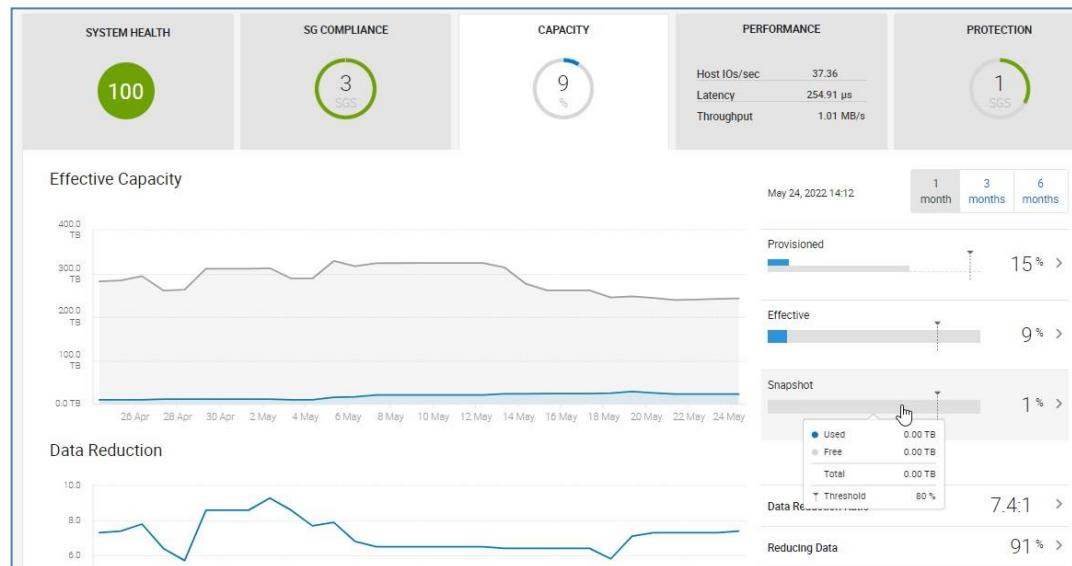
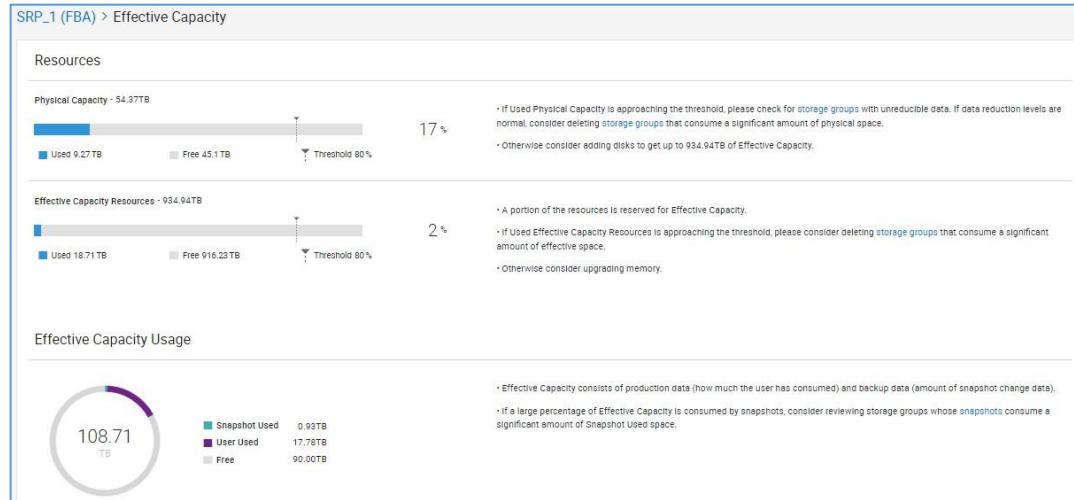


Figura 6. Dashboard di capacità

Nel dashboard di capacità effettiva:

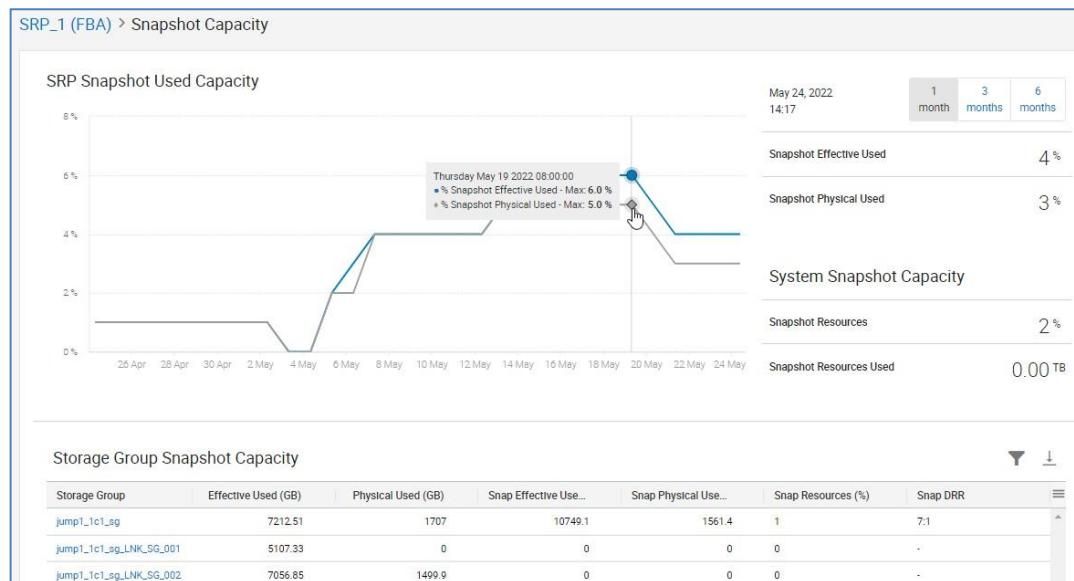
- **Utilizzo delle istantanee:** quantità di capacità effettiva utilizzata dai dati di modifica delle istantanee



**Figura 7. Dashboard di capacità effettiva**

I valori della pagina di capacità delle istantanee sono definiti come:

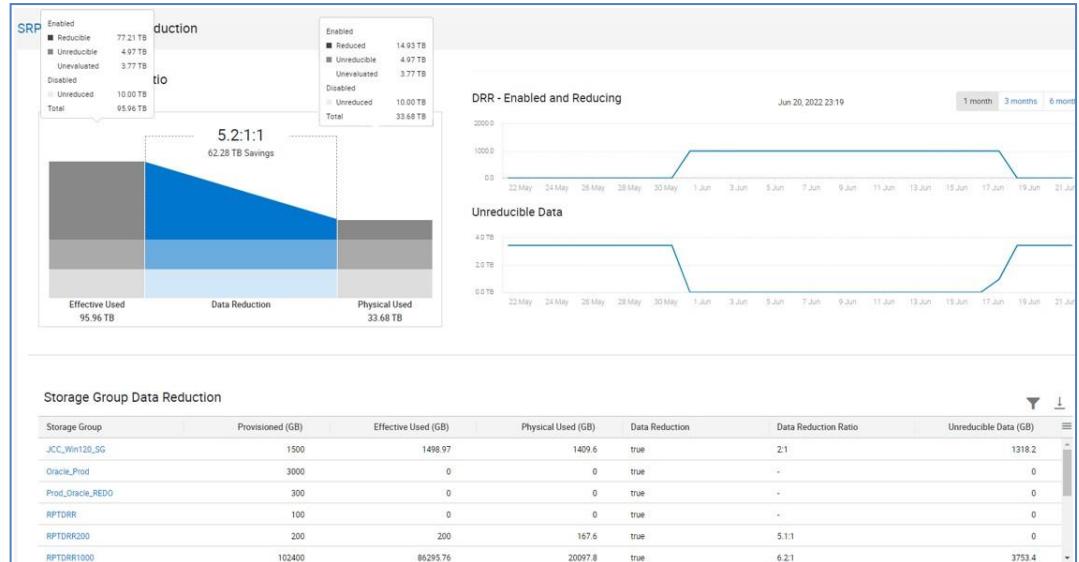
- **Utilizzo effettivo delle istantanee:** percentuale di capacità effettiva utilizzata dai dati delle istantanee per SRP
- **Utilizzo fisico delle istantanee:** percentuale di capacità utilizzabile coperta dai dati delle istantanee per SRP
- **Risorse delle istantanee:** percentuale di metadati delle istantanee utilizzati per l'intero sistema
- **Risorse delle istantanee utilizzate:** capacità utilizzata dalle istantanee come parte della capacità dei metadati delle istantanee



**Figura 8. Uso della capacità delle istantanee**

## Rapporto di riduzione dei dati

La schermata della riduzione dei dati fornisce un'unica posizione in cui gli utenti possono visualizzare l'efficienza della riduzione dei dati. Sono disponibili tre sezioni: rapporto di riduzione dei dati, un grafico interattivo con i dati cronologici e una tabella di tutti gli storage group. Il rapporto di riduzione dei dati visualizzato riguarda solo l'abilitazione e la riduzione dei dati scritti nel sistema.



**Figura 9. Schermata di riduzione dei dati**

**Il rapporto di riduzione dei dati** viene visualizzato sotto forma di grafico che presenta la capacità effettiva utilizzata, il rapporto di riduzione dei dati e la capacità fisica utilizzata. La capacità fisica utilizzata fa riferimento alla quantità effettiva di capacità fisica in uso. La riduzione dei dati presenta il risparmio sotto forma di rapporto.

**L'utilizzo effettivo** rappresenta i dati scritti nel sistema prima di ottenere un risparmio applicando la riduzione dei dati. Tutti i valori visualizzati rappresentano la dimensione completa così com'è stata scritta dall'host o dall'applicazione. Esistono due categorie in cui si collocano i dati scritti: Abilitato e Disabilitato.

- **Abilitato** indica che i dati presi in considerazione per la riduzione dei dati sono soggetti al processo di riduzione dei dati e abilitati per la funzione di riduzione basata sull'attività. Esistono tre categorie aggiuntive in cui possono rientrare i dati quando la riduzione dei dati ha uno stato Abilitato, Riducibile, Non riducibile e Non valutato.
  - I dati **riducibili** rappresentano la quantità di dati scritti identificata dal processo di riduzione dei dati come dati che possono essere ridotti per utilizzare una capacità fisica inferiore rispetto a quella scritta nel sistema.
  - I dati **non riducibili** non possono essere ridotti.
  - I dati **non valutati** non vengono presi in considerazione durante il processo di riduzione dei dati. In tal caso, non è ancora stato stabilito se i dati sono riducibili o non riducibili.
- **Disabilitato** indica che i dati scritti nel sistema non sono soggetti ad alcun risparmio derivante dalla riduzione dei dati. Tutti i dati identificati come disabilitati verranno visualizzati come non ridotti.

**L'utilizzo fisico** rappresenta i dati scritti nel sistema dopo l'archiviazione su disco. Ciò riguarda tutti i dati abilitati e disabilitati, nonché tutti i dati ridotti e non ridotti. Esistono due categorie che rappresentano i dati archiviati su disco: Abilitato e Disabilitato.

- **Abilitato** indica i dati sottoposti al processo di riduzione dei dati. Lo stato di questi dati può essere classificato in tre sottocategorie: Ridotto, Non riducibile e Non valutato.
  - I dati **ridotti** sono dati inviati attraverso il processo di riduzione dei dati. Questo processo include il passaggio tramite l'hardware per la riduzione dei dati e l'archiviazione su disco. La riduzione dei dati archiviati su disco consuma meno spazio su disco rispetto a quanto scritto dall'host o dall'applicazione.
  - I dati **non ridotti** sono dati inviati tramite il processo di riduzione dei dati, incluso l'hardware per la riduzione dei dati, ma per i quali non è stato possibile effettuare la riduzione. Alcuni dati non riducibili presi in considerazione all'interno della sezione relativa all'utilizzo fisico possono contribuire al risparmio derivante dalla riduzione dei dati, in quanto si tratta di dati condivisi a causa della deduplica.
  - I dati **non valutati** non vengono presi in considerazione durante il processo di riduzione dei dati. In tal caso, pertanto, non è ancora stato stabilito se i dati sono riducibili o non riducibili.
- **Disabilitato** indica che i dati scritti nel sistema non sono soggetti ad alcun risparmio derivante dalla riduzione dei dati. Tutti i dati identificati come disabilitati verranno visualizzati come non ridotti.

Il grafico interattivo **Abilitazione e riduzione DRR** e i **dati non riducibili** forniscono dati cronologici. Mostra l'effetto dei dati non riducibili sul rapporto di riduzione dei dati. Questo grafico consente all'utente di tenere traccia e monitorare le modifiche del rapporto di riduzione dei dati potenzialmente causate da dati non riducibili.

L'elenco degli storage group fornisce informazioni sull'uso della capacità e sulla riduzione dei dati specifiche per ogni storage group nel sistema. Quando si utilizza il grafico interattivo per tenere traccia delle modifiche apportate al rapporto di riduzione dei dati, è possibile utilizzare l'elenco degli storage group per identificare quelli contenenti grandi quantità di dati non riducibili che influiscono sul rapporto di riduzione dei dati.

**Calcolo dei rapporti di efficienza:** i dati necessari per calcolare il rapporto di riduzione dei dati sono disponibili in una finestra pop-up nel grafico di riduzione dei dati.

- **Rapporto di riduzione dei dati:** il rapporto di riduzione dei dati viene calcolato utilizzando i dati abilitati e riducibili dall'utilizzo effettivo e i dati abilitati e ridotti dall'utilizzo fisico.  

$$\text{Dati abilitati riducibili} \div \text{Dati abilitati ridotti}$$
- **Rapporto complessivo di riduzione dei dati:** il rapporto complessivo di riduzione dei dati del sistema viene calcolato utilizzando i valori totali ottenuti dall'utilizzo effettivo e dall'utilizzo fisico  

$$\text{Utilizzo effettivo totale} \div \text{Utilizzo fisico totale}$$

## Data service supportati

### Panoramica

La riduzione dei dati è supportata per lo storage FBA. I sistemi FBA/CKD misti sono supportati all'interno dello stesso pool di risorse di storage, sebbene la riduzione dei dati per l'emulazione CKD sfrutti solo la riduzione basata su attività e compressione. Sono supportati tutti gli altri data service offerti in entrambi i sistemi PowerMax e VMAX All-Flash. Questi servizi includono inoltre replica locale (SnapVX), replica remota (SRDF), D@RE e VMware vSphere Virtual Volumes (vVols).

### Replica locale (SnapVX)

Le istantanee SnapVX proteggono le applicazioni senza l'utilizzo di volumi di destinazione per acquisire i dati di modifica, noti come delta. I delta delle istantanee vengono gestiti automaticamente sul back-end dello storage utilizzando puntatori alle immagini point-in-time pertinenti. La condivisione delle risorse

e la deduplica dei dati sfruttano automaticamente questa progettazione per fornire vantaggi in termini di memoria cache, capacità e prestazioni.

I dati di origine compressi rimangono tali quando diventano delta delle istantanee. I dati di origine non compressi possono essere compressi man mano che diventano delta delle istantanee oppure meno attivi. L'attività di lettura tramite destinazioni collegate può impedire la compressione di un delta non compresso o la decompressione dei delta delle istantanee compressi. I delta delle istantanee sono disponibili per la deduplica.

L'abilitazione della riduzione dei dati su una destinazione collegata avrà effetto solo sui dati di proprietà della destinazione collegata. I dati su destinazioni e cloni collegati sono disponibili per la deduplica.

#### **Replica remota (SRDF)**

La compressione per SRDF è supportata e nota come compressione SRDF. La compressione SRDF è una funzione progettata per ridurre l'utilizzo della larghezza di banda durante l'invio dei dati da e verso sistemi connessi tramite la replica remota. La compressione SRDF e la riduzione dei dati utilizzano entrambi lo stesso hardware; tuttavia, hanno scopi diversi. I dati compressi tramite riduzione dei dati vengono decompressi prima di essere inviati attraverso il link SRDF. Nel caso in cui si applichino la compressione SRDF e la compressione in linea, i dati vengono decompressi, compressi utilizzando la funzione di compressione SRDF e inviati al sito remoto.

#### **Data at Rest Encryption (D@RE)**

D@RE fornisce una crittografia back-end su array basata sull'hardware; la riduzione dei dati viene eseguita come processo in linea. I dati vengono fatti passare attraverso l'hardware per la riduzione dei dati prima di essere inviati tramite l'hardware di crittografia. Di conseguenza, i dati vengono compressi, deduplicati o entrambi prima di essere crittografati. In un sistema abilitato per D@RE, i dati crittografati su disco sono già stati compressi, deduplicati o entrambe le cose.

#### **Virtual Volumes (vVol)**

La riduzione dei dati è supportata per l'allocazione dei dati ai vVol e segue lo stesso percorso I/O di tutti gli altri dati. Il percorso I/O è riportato nella [figura 1](#). La riduzione dei dati è abilitata a livello di risorse di storage in un container di storage vVol in quanto non sono presenti storage group per i vVols.

## **Conclu- sioni**

#### **Riepilogo**

L'utilizzo della capacità di storage fisico è un problema comune agli Storage Administrator dell'intero settore dello storage. La costante crescita delle quantità di dati ha fatto emergere la necessità di una maggiore efficienza nell'uso della capacità fisica. I sistemi di storage dei dati Dell PowerMax 2500 e 8500 portano l'efficienza a un livello superiore. La riduzione dei dati offre un eccezionale risparmio in termini di capacità e prestazioni ottimali. Ciò determina un minore ingombro del data center e una riduzione complessiva del TCO. Oltre a offrire risparmi, la riduzione dei dati può essere abilitata o disabilitata con un semplice clic. Tutte le attività vengono gestite dal sistema.

## Riferimenti

### Documentazione di Dell Technologies

La seguente documentazione Dell Technologies fornisce ulteriori informazioni correlate. L'accesso a questi documenti dipende dalle credenziali di login di cui si dispone. Se non è possibile accedere a un documento, contattare il proprio responsabile Dell Technologies.

- [PowerMax and VMAX Info Hub](#)