

# Dell PowerMax: Reducción de datos

## Compresión y deduplicación en línea

Julio de 2022

H19254

## Informe técnico

### Resumen

Las plataformas de almacenamiento PowerMax cuentan con varias técnicas de reducción de datos, como la compresión y la deduplicación en línea. También incluyen la detección de patrones y la ubicación eficiente de los datos para ofrecer un excelente equilibrio entre rendimiento y eficiencia.

Dell Technologies

## Copyright

La información de esta publicación se proporciona tal cual. Dell Inc. no se hace responsable ni ofrece garantía de ningún tipo con respecto a la información de esta publicación y específicamente renuncia a toda garantía implícita de comerciabilidad o capacidad para un propósito determinado.

El uso, la copia y la distribución de cualquier software descrito en esta publicación requieren una licencia de software correspondiente.

Copyright © 2022 Dell Inc. o sus subsidiarias. Todos los derechos reservados. Dell Technologies, Dell, EMC, Dell EMC y otras marcas comerciales son marcas comerciales de Dell Inc. o sus filiales. Intel, el logotipo de Intel, el logotipo de Intel Inside y Xeon son marcas comerciales de Intel Corporation en los Estados Unidos o en otros países. Las demás marcas comerciales pueden ser marcas comerciales de sus respectivos dueños.

Publicado en México, julio de 2022 H19254.

Dell Inc. considera que la información de este documento es precisa en el momento de su publicación. La información está sujeta a cambios sin previo aviso.

# Contenido

Resumen ejecutivo .....	4
Reducción de datos .....	5
Uso de recursos del sistema.....	9
Administración y monitoreo.....	9
Servicios de datos admitidos.....	17
Conclusión.....	18
Referencias .....	19

## Resumen ejecutivo

### Visión general

La reducción de datos con Dell PowerMax aumenta la eficiencia del sistema porque combina la compresión en línea, la deduplicación en línea y la detección de patrones. El uso de estas técnicas de reducción de datos permite a los usuarios lograr grandes ahorros de capacidad. La reducción de datos comprime los datos y elimina las copias redundantes de datos. En esta documentación técnica, se explica cómo funciona la reducción de datos en los sistemas PowerMax y se describen los informes mediante aplicaciones de administración de Dell, como Unisphere for PowerMax, Solutions Enabler y el software Mainframe Enabler.

### Revisiones

Fecha	Descripción
Julio de 2022	Versión inicial

### Valoramos sus comentarios

Dell Technologies y los autores del presente documento recibirán con mucho agrado sus comentarios sobre este documento. Comuníquese con el equipo de Dell Technologies por [correo electrónico](#).

**Autor:** Robert Tasker

---

**Nota:** para obtener enlaces a otra documentación de este tema, consulte el [Centro de información de PowerMax y VMAX](#).

---

## Reducción de datos

### Visión general

La reducción de datos combina la compresión en línea, la deduplicación en línea, la detección de patrones, la ubicación eficiente de los datos y el aprendizaje automático (ML). Esta combinación crea un sistema en el que los usuarios pueden escribir más datos de host que la cantidad total de capacidad física disponible y, al mismo tiempo, lograr el rendimiento esperado de un sistema de almacenamiento empresarial. Esta función está activada de manera predeterminada y puede habilitarla o deshabilitarla en el nivel del grupo de almacenamiento. Además, todos los servicios de datos disponibles en los sistemas PowerMax 2500 y 8500 son compatibles. Esta compatibilidad también se aplica a la emulación de CKD, pero no incluye la deduplicación para CKD.

La compresión reduce el tamaño de los datos, mientras que la deduplicación (dedupe) almacena los datos como una sola instancia. La detección de patrones incluye una función de asignación distinta de cero que excluye que las cadenas de ceros consecutivos se almacenen como parte de los datos comprimidos.

La compresión, la deduplicación y la detección de patrones se realizan mediante la asistencia de hardware integrada dentro del sistema para reducir la sobrecarga de realizar estas funciones.

El aprendizaje automático identifica los datos más activos almacenados en el disco y garantiza que permanezcan inducidos para obtener un rendimiento óptimo. La ubicación eficiente de los datos utiliza una función denominada **compactación** que almacena estratégicamente los datos para minimizar el espacio desperdiciado y reduce la necesidad de funciones de recolección o desfragmentación (defrag) de elementos no utilizados.

### Compresión basada en la actividad

La Reducción basada en actividades (ABR) reduce el costo de rendimiento incurrido al descomprimir los datos a los que se accede con frecuencia. Esta función permite que hasta un 20 % de los datos más activos se almacenen en el sistema sin comprimir. Esta capacidad beneficia al sistema, ya que minimiza la latencia de rendimiento que resulta de la descompresión constante de los datos a los que se accede con frecuencia. Para determinar qué datos son los más utilizados, el sistema utiliza algoritmos de ML que procesan estadísticas de I/O. Realizar esta tarea mantiene un entorno equilibrado y óptimo para el ahorro de reducción de datos y el rendimiento.

### Compresión

La compresión reduce las cargas de trabajo de escritura entrantes al tamaño más pequeño posible para consumir la menor cantidad de capacidad. Los datos se comprimen cuando se transfieren a través del hardware de reducción de datos que utiliza el algoritmo de compresión de GZIP. Cuando se transfieren a través del hardware de reducción de datos, los datos se dividen en cuatro secciones que se comprimen en paralelo para maximizar la eficiencia. La suma de las cuatro secciones es el tamaño reducido final de los datos que se almacenan en el disco. Esta capacidad proporciona acceso granular a los datos reducidos. Solo se procesan las secciones que contienen los datos solicitados para las solicitudes parciales de lectura o escritura, ya que cada sección se puede manejar de manera independiente.

### Desduplicación

La desduplicación es un método de ahorro de capacidad que identifica copias idénticas de datos y almacena una sola instancia de cada copia. Hay algunos componentes de la desduplicación que son necesarios para que pueda proporcionar ahorros de capacidad eficientes.

- **ID de hash:** el ID de hash es un identificador único para los datos entrantes que se utiliza para determinar si se necesita una relación de desduplicación. El sistema utiliza un algoritmo SHA-256 para generar el ID de hash.
- **Tabla de ID de hash:** las tablas hash corresponden a una asignación de la memoria del sistema que se distribuye entre los directores del sistema. Estas tablas catalogan los ID de hash que utiliza el proceso de desduplicación. Las entradas de la tabla se utilizan para determinar si existe una relación de desduplicación o si se requiere una nueva entrada y los datos se pueden almacenar en el disco.
- **Objeto de administración de desduplicación (DMO):** el DMO es un objeto de 64 bytes dentro de la memoria del sistema que solo existe cuando existe una relación de desduplicación. Estos objetos almacenan y administran los punteros entre los dispositivos front-end y los datos desduplicados que

consumen la capacidad del back-end en el arreglo. Los DMO administran los punteros para los datos deduplicados entre los dispositivos front-end y los datos almacenados en el disco. También administra la tabla hash en la que se almacenan los ID de hash cuando existen relaciones de deduplicación.

La deduplicación se realiza mediante el mismo hardware de reducción de datos que la compresión, y se genera un ID de hash único cuando el hardware procesa los datos. A continuación, el ID de hash se compara con la tabla de ID de hash que busca el mismo ID. Cuando se encuentra una coincidencia, los datos no se almacenan en el disco y se crea un recurso compartido de deduplicación. Los punteros se establecen entre el volumen de front-end y el ID único en la tabla de ID de hash. Los punteros vinculan la instancia única de datos almacenados en el disco al volumen, lo que proporciona acceso futuro a los datos. El DMO administra los punteros entre los datos, los volúmenes de front-end que acceden a los datos y la tabla de ID de hash. Cuando no hay ninguna coincidencia en la tabla de ID de hash, se agrega una nueva entrada a para futuras comparaciones de ID de hash.

### Algoritmo de deduplicación

Los sistemas PowerMax utilizan el algoritmo de hash SHA-256 implementado en el hardware de reducción de datos para encontrar datos duplicados. A continuación, los datos se almacenan como una instancia única para que se compartan en varias fuentes. Este proceso proporciona una eficiencia de datos mejorada y, al mismo tiempo, mantiene la integridad de los datos.

El algoritmo SHA-256 genera un código de 32 bytes para cada bloque de datos de 32 KB. Considerar un sistema con 1 PB de datos escritos con un 5 % de actualización por día. En un millón años de funcionamiento, existe una probabilidad del 20 % de una colisión de hash. Como cada pista de 128 KB se maneja como cuatro bloques de 32 KB, tendría que haber una colisión de hash en los cuatro bloques de la misma pista de 128 KB para tener una colisión de hash real. Las probabilidades de que los cuatro bloques colisionen hacen que esto sea solo teórico (menos de un 1 % de probabilidades en un billón de años de funcionamiento). Además, cuando se encuentra una coincidencia durante la fase de comparación de la deduplicación, se realiza una comparación de byte por byte. Esta comparación se realiza para confirmar que hay una coincidencia antes de actualizar las tablas y configurar los punteros para permitir el acceso a los datos.

### Compactación

La ubicación de los datos se realiza mediante un proceso llamado compactación. La compactación coloca de manera intuitiva los datos reducidos o no provocados en el disco en la mejor ubicación posible disponible. La operación de almacenamiento de datos en disco utiliza objetos de escritura. Cada objeto tiene 6 MB de capacidad contigua del dispositivo de datos de back-end en todas las unidades configuradas en el sistema. Los objetos de escritura se alinean en los límites de 1 K y se consumen secuencialmente en un solo uso. Los objetos de escritura se distribuyen entre fracciones completas para todos los tipos de RAID compatibles a fin de optimizar las escrituras. Cada objeto admite datos reducidos o no reducidos para la emulación de FBA y CKD.

- **Objeto de escritura de FBA:** un objeto de escritura no provocado consta de 48 segmentos de FBA. Un objeto de escritura reducido consta de 1000 segmentos reducidos. Las entradas reducidas para objetos de escritura varían de 1 KB a 96 KB.
- **Objeto de escritura de CKD:** un objeto de escritura no provocado consta de 108 segmentos de CKD. Un objeto de escritura reducido consta de 1000 segmentos reducidos. Las entradas reducidas para objetos de escritura varían de 1 KB a 52 KB.

## Compresión de datos extendida

Los sistemas PowerMax 2500 y 8500 incluyen una función adicional denominada compresión de datos extendida (EDC) que comprime los datos ya comprimidos para obtener mayores ahorros de capacidad. Esta tarea se logra mediante la identificación de los datos a los que no se ha accedido por un período de tiempo extendido. Los factores que hacen que los datos sean candidatos a la EDC son los siguientes:

- Los datos pertenecen a un grupo de almacenamiento habilitado para la reducción de datos.
- No se ha accedido a los datos durante 30 días.
- Los datos aún no están comprimidos por EDC.

Los datos que califican para EDC se comprimen mediante el algoritmo Def9\_128\_SW para reducir aún más la cantidad de capacidad utilizada para almacenar los datos. Este es un proceso en segundo plano automatizado dentro del sistema. Los ahorros adicionales se incluyen en el grupo de almacenamiento. tasa de compresión alcanzada en el nivel. La EDC solo está disponible con los arreglos de almacenamiento PowerMax.

## CKD Compresión

La Reducción basada en actividades (ABR) reduce el costo de rendimiento incurrido al descomprimir los datos a los que se accede con frecuencia. Esta función permite que hasta un 20 % de los datos más activos se almacenen en el sistema sin comprimir. Este resultado beneficia al sistema, ya que elimina el impacto negativo en el rendimiento que resulta de la descompresión constante de los datos a los que se accede con frecuencia. Para determinar el nivel ocupado de datos, el sistema utiliza algoritmos de ML que procesan estadísticas recopiladas de I/O entrantes a los dispositivos de front-end. Esta acción mantiene en equilibrio entre los recursos del sistema, lo que proporciona un entorno óptimo para el ahorro y el rendimiento de la reducción de datos.

La compresión reduce las cargas de trabajo de escritura entrantes al tamaño más pequeño posible para consumir la menor cantidad de capacidad posible. Los datos se comprimen cuando se transfieren a través del hardware de reducción de datos incorporado en el sistema que utiliza el algoritmo de compresión GZIP. Cuando se transfieren a través del hardware de reducción de datos, los datos se dividen en cuatro secciones que se comprimen en paralelo para maximizar la eficiencia del hardware. La suma de las cuatro secciones es el tamaño reducido final de los datos que se almacenan en el disco. Este resultado proporciona acceso granular para datos reducidos cuando hay una solicitud parcial de lectura o escritura. Solo se procesan las secciones que contienen los datos solicitados, ya que cada sección se puede manejar de manera independiente.

La ubicación de los datos se realiza mediante un proceso de ubicación de datos llamado compactación. La compactación coloca de manera intuitiva los datos reducidos o no provocados en el disco en la mejor ubicación posible disponible. La operación de almacenamiento de datos en disco utiliza objetos de escritura. Cada objeto tiene 6 MB de capacidad contigua del dispositivo de datos de back-end en todas las unidades configuradas en el sistema. Los objetos de escritura se alinean en los límites de 1 K y se consumen secuencialmente en un solo uso. Objetos de escritura distribuidos en fracciones completas para todos los tipos de RAID compatibles para optimizar las escrituras. Cada objeto admite datos reducidos o no reducidos. Un objeto de escritura no provocado consta de 108 segmentos de CKD. Un objeto de escritura reducido consta de 1000 segmentos reducidos. Las entradas reducidas para objetos de escritura varían de 1 KB a 52 KB.

## Flujo de I/O de la reducción de datos

Todas las I/O se pasan a través de la memoria caché y luego son procesadas por el sistema. Las acciones de reducción de datos se realizan después de que el sistema recibe los datos, antes de que se coloquen en el disco. El uso de un proceso en línea requiere controles adicionales dentro del flujo de I/O donde se aplica la reducción de datos. El sistema utiliza estas comprobaciones para determinar si los datos entrantes necesitan pasar por el hardware de reducción de datos o no. Los datos entrantes para un grupo de almacenamiento con reducción de datos habilitada seguirán el flujo de la reducción de datos. Sin embargo, debido a la reducción basada en la actividad (ABR), los datos

activos de un grupo de almacenamiento con reducción de datos habilitada omitirán el flujo de reducción de datos para la optimización del rendimiento. Los datos que no se comprimen debido a ABR pueden comprimirse más adelante y trasladarse a un pool de compresión. Los datos para un grupo de almacenamiento con reducción de datos deshabilitada ignorarán el flujo de reducción de datos y se escribirán en el sistema sin reducir.

Existen algunos tipos de I/O diferentes a tener en cuenta: leer, escribir y escribir-actualizar.

- **Leer:** una solicitud para acceder a los datos que ya están completando el arreglo.
- **Escribir:** I/O entrantes que consumirán espacio de disco.
- **Escribir-actualizar:** I/O entrantes que pueden cambiar los datos que se asignan al espacio de disco en el arreglo.

La siguiente figura describe la ruta que seguirán las I/O que se determina según las características del conjunto de datos o del grupo de almacenamiento relacionado.

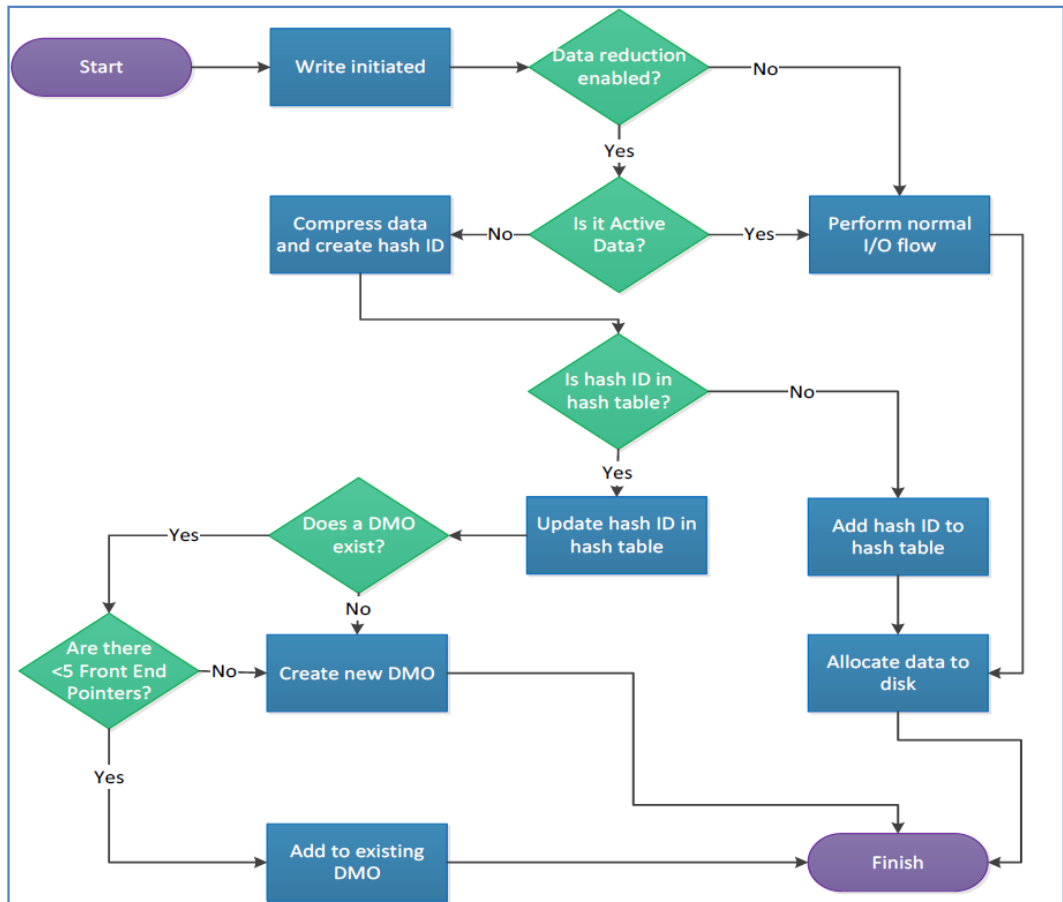


Figura 1. Flujo de I/O de reducción de datos para sistemas de almacenamiento empresarial PowerMax



## Uso de recursos del sistema

### Visión general

La capacidad y la memoria son los dos recursos principales configurados en cada sistema PowerMax. La capacidad se categoriza en física, real y aprovisionada. La memoria se categoriza como recursos del sistema.

Los recursos de memoria admiten las estructuras de metadatos para la capacidad aprovisionada, así como la capacidad física. La cantidad de capacidad real disponible se relaciona con la cantidad de capacidad física, la cantidad de recursos del sistema disponibles y la reducción de los datos escritos en el sistema. Los datos escritos que son altamente reducibles consumen menos capacidad física, lo que da como resultado más capacidad real. Lo contrario también es cierto: los datos escritos que no son reducibles pueden dar lugar a una menor capacidad real disponible. La información que se describe en la sección [Reducción de datos](#) (capacidad, recursos del sistema) está disponible dentro de las aplicaciones de administración utilizadas para los sistemas PowerMax 2500 y 8500, Unisphere for PowerMax, Solutions Enabler y Mainframe Enabler Software. Unisphere for PowerMax es una interfaz de usuario (UI) que proporciona datos en gráficos, gráficos y formularios de lista. Solutions Enabler es una interfaz de línea de comandos estándar que proporciona los mismos datos, pero no en forma de tablas ni gráficos. Mainframe Enablers es un conjunto de componentes que monitorea y administra los sistemas Dell Storage en un entorno de mainframe. En las imágenes que se muestran en las siguientes secciones de este informe, se muestra Unisphere para PowerMax que administra un sistema PowerMax 2500 u 8500.

### Capacidad física

La capacidad física es la cantidad de espacio de disco configurada en el sistema en función de los discos instalados y la protección RAID aplicada. En una configuración en la que la reducción de datos no está en uso, la capacidad física es la cantidad total de capacidad disponible para los datos del host. Por ejemplo, un sistema que muestra 100 TB de capacidad física indica que puede alojar 100 TB de datos de host que no utilizan la reducción de datos.

### Capacidad real

La capacidad real es la cantidad de espacio disponible cuando la reducción de datos está en uso. La cantidad total en la instalación inicial depende de la cantidad de memoria configurada en el sistema y se basa en un ahorro de reducción de datos predeterminado de 4:1 (3:1 para la emulación de CKD). Por ejemplo, ese mismo sistema con 100 TB de capacidad física mostrará 400 TB de capacidad real. Este valor de 400 TB es un punto de partida de capacidad real y cambiará a medida que los datos se escriban en el sistema y se aplique la reducción de datos.

### Capacidad aprovisionada

La capacidad aprovisionada es la representación de la capacidad disponible en forma de dispositivos creados y presentados a hosts y aplicaciones que pretenden consumir capacidad física o real en el sistema.

## Administración y monitoreo

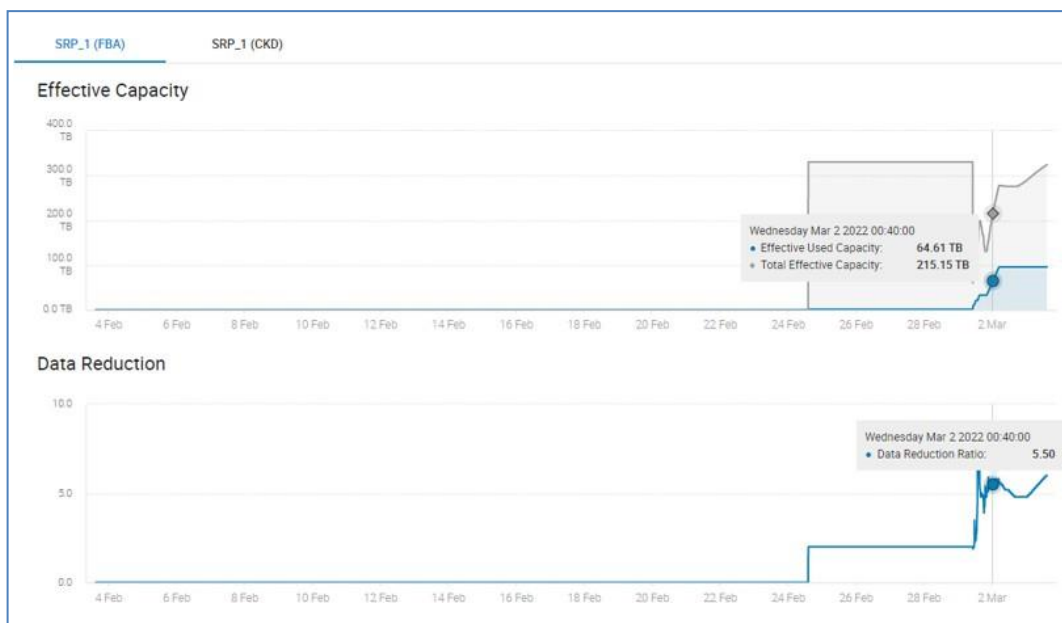
### Visión general

Unisphere for PowerMax es una interfaz del usuario que se utiliza para administrar y monitorear la capacidad y el uso de recursos del sistema. En el nivel del sistema, la información sobre el uso de la capacidad, la reducción de datos y los recursos del sistema se muestra en el tablero de capacidad. Desde el tablero de capacidad, los usuarios pueden ir a pantallas que muestran información sobre la capacidad real y aprovisionada, la capacidad de instantáneas y la reducción de datos, así como los recursos del sistema.

### Tablero de capacidad

En Unisphere for PowerMax, hay varias pantallas que proporcionan información relacionada con el uso de la capacidad.

El tablero principal muestra un gráfico interactivo que muestra el uso eficaz de la capacidad y la reducción de datos con el tiempo. En esta pantalla, se muestra el historial del uso eficaz de la capacidad y cómo se relaciona la relación de reducción de datos con la capacidad real. Esta información se puede utilizar para monitorear y rastrear las tendencias de uso eficaz de la capacidad en relación con la tasa de reducción de datos que se muestra. Los sistemas PowerMax 2500 u 8500 se pueden configurar con emulación de FBA y CKD dentro del mismo pool de recursos de almacenamiento, pero el gráfico histórico es específico de la vista de emulación seleccionada.



**Figura 2. Gráfico histórico del tablero de capacidad que muestra la capacidad real y la reducción de datos para la emulación de FBA**

El tablero principal también ofrece datos en forma de gráficos de barras para la capacidad aprovisionada, la capacidad real, el uso de instantáneas y la reducción de datos. Cada sección se puede expandir a una pantalla más detallada que muestra datos más granulares para cada elemento.

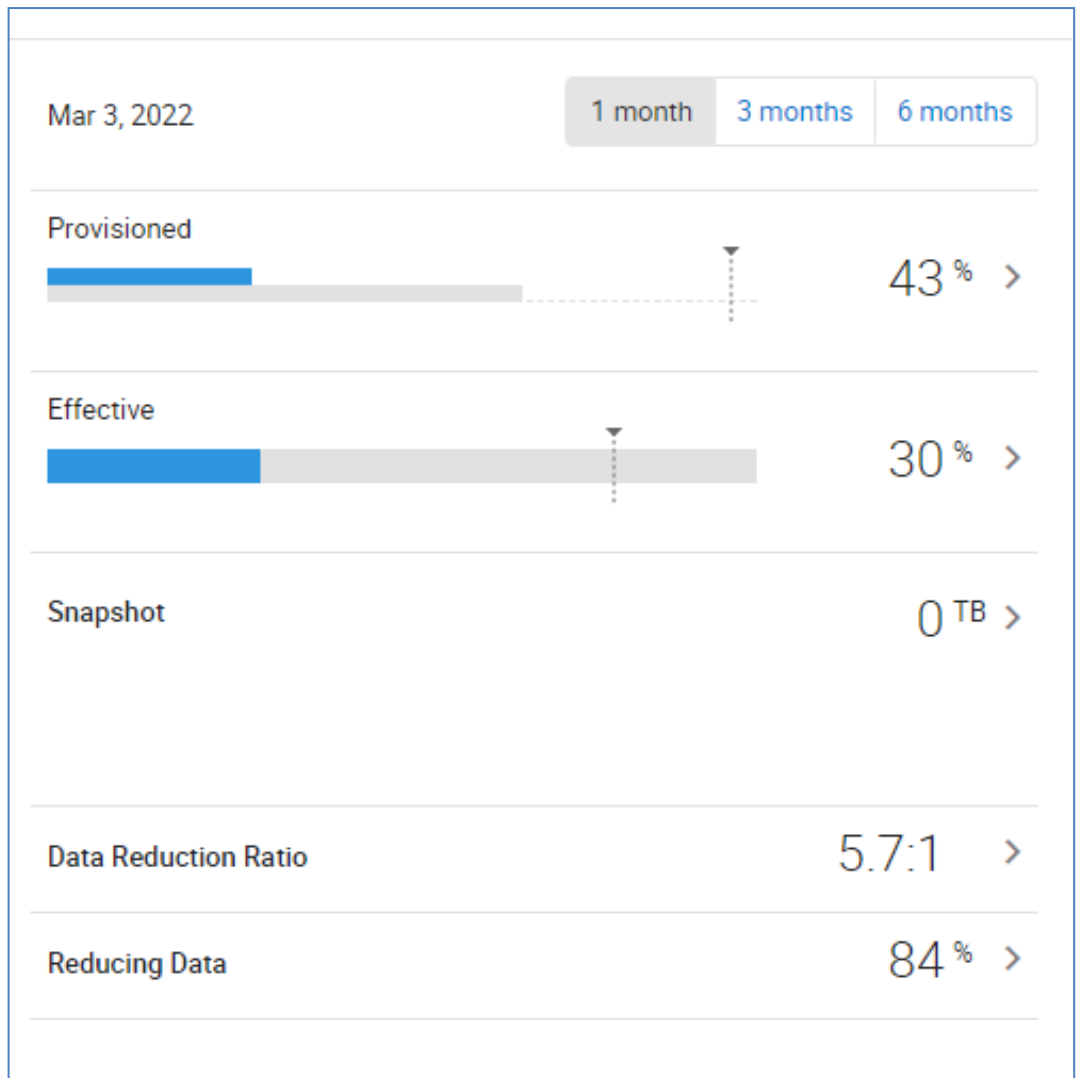


Figura 3. Gráficos de barras de tableros de capacidad para la capacidad aprovisionada, la capacidad real, el uso de instantáneas y la reducción de datos

### Aprovisionado

La capacidad aprovisionada es la cantidad de capacidad que se aprovisiona en forma de dispositivos que se presentan a los hosts y la aplicación como capacidad disponible. La capacidad aprovisionada se rastrea mediante dos métricas: capacidad de SRP y recursos del sistema.

- Capacidad de SRP** muestra la cantidad de capacidad aprovisionada como un valor de TB aprovisionado y la cantidad disponible de capacidad real. La cantidad de capacidad de SRP real se basa inicialmente en un ahorro de reducción de datos predeterminado de 4:1 (3:1 para la emulación de CKD). Esta cantidad se calcula utilizando la cantidad de capacidad física configurada en el sistema. A medida que los usuarios crean dispositivos, aumenta la capacidad aprovisionada. El valor porcentual que se muestra es el porcentaje de suscripción y se calcula utilizando la cantidad aprovisionada y la cantidad de capacidad real.

- **Los recursos del sistema** representan la cantidad de memoria disponible para admitir la capacidad aprovisionada en forma de metadatos. El valor total presentado solo cambiará si se agrega caché al sistema. La cantidad utilizada representa solo eso. La cantidad libre es una indicación de cuánta capacidad aprovisionada adicional puede admitir el sistema. A medida que los usuarios crean dispositivos, la cantidad utilizada aumentará. Cuando hay una variación entre los dos valores, los recursos de memoria se utilizan para admitir la reducción de datos u otras funciones que utilizan memoria, como tomar instantáneas de dispositivos existentes.

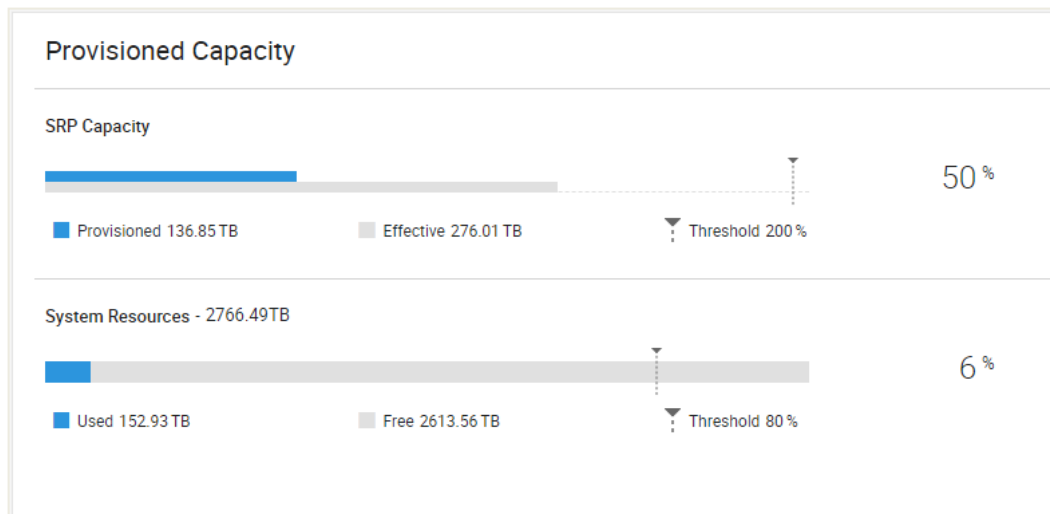


Figura 4. Uso de recursos del sistema dentro de la visualización de capacidad aprovisionada

## Real

La capacidad real representa la cantidad de capacidad disponible para el usuario en función de una expectativa de ahorros por el uso de la reducción de datos. La pantalla de capacidad real proporciona una vista detallada de los recursos físicos y reales disponibles. Esto se muestra en tres secciones: capacidad física, recursos de capacidad real y uso de la capacidad real.

- **Capacidad física** muestra la cantidad de capacidad física disponible en los discos duros configurados en el sistema. Las cantidades que se muestran son los valores después de aplicar el formato y la protección RAID. El valor que se muestra es la cantidad de capacidad que el sistema puede admitir para los datos del host cuando no se utiliza la reducción de datos.
- **Los recursos de capacidad real** indican los valores alcanzables en función del uso actual de recursos del sistema. El valor de recursos de capacidad real que se muestra se ajustará en relación con los ahorros de reducción de datos actuales y el uso físico y real de la capacidad.
- **Uso de capacidad real** muestra la cantidad actual de capacidad real disponible en función del uso de recursos del sistema y los ahorros actuales de reducción de datos. El valor que se muestra dentro del gráfico circular es la capacidad real actual disponible. Los valores presentados a la derecha desglosan el uso en tres categorías: uso de instantáneas, uso de usuario y libre.

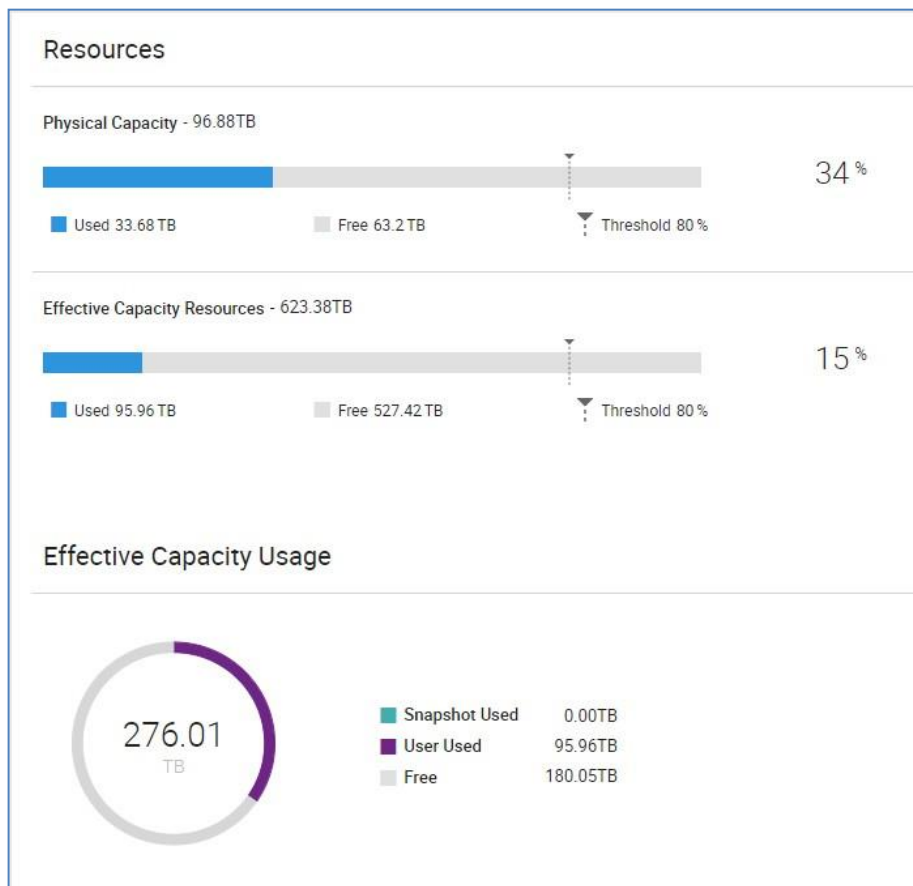


Figura 5. Uso de recursos que muestra la pantalla Capacidad real

### Instantánea

La capacidad de instantánea de back-end puede ser significativamente menor que la deltas por instantánea debido a la eficiencia de funciones como asignaciones compartidas y reducción de datos.

Pase el cursor sobre el gráfico de la barra de instantáneas en el tablero de capacidad para obtener detalles generales. Los valores de instantánea se definen de la siguiente manera:

- **Utilizada:** capacidad real utilizada para los datos de cambio de instantáneas
- **Libre:** cantidad restante de datos de cambio de instantánea según la cantidad utilizada y los metadatos restantes
- **Total:** utilizada + libre
- **Umbral:** umbral para alertas sobre datos de cambio de instantáneas

Haga clic en los gráficos de barras para ir al tablero Capacidad real y al tablero Capacidad de instantáneas.

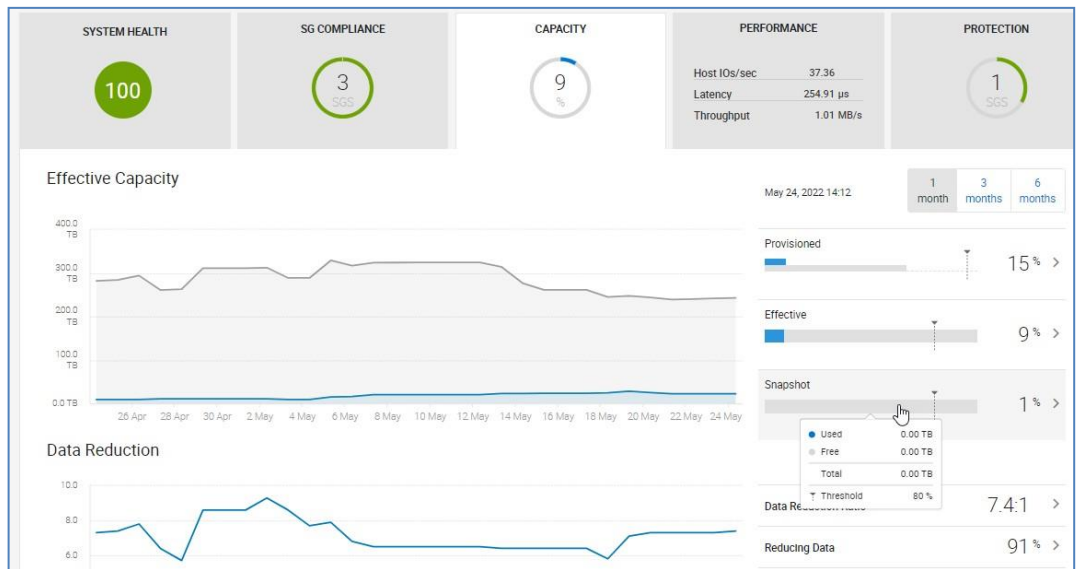


Figura 6. Tablero de capacidad

En el tablero Capacidad real:

- **Instantánea utilizada:** cantidad de capacidad real utilizada por los datos de cambio de instantánea

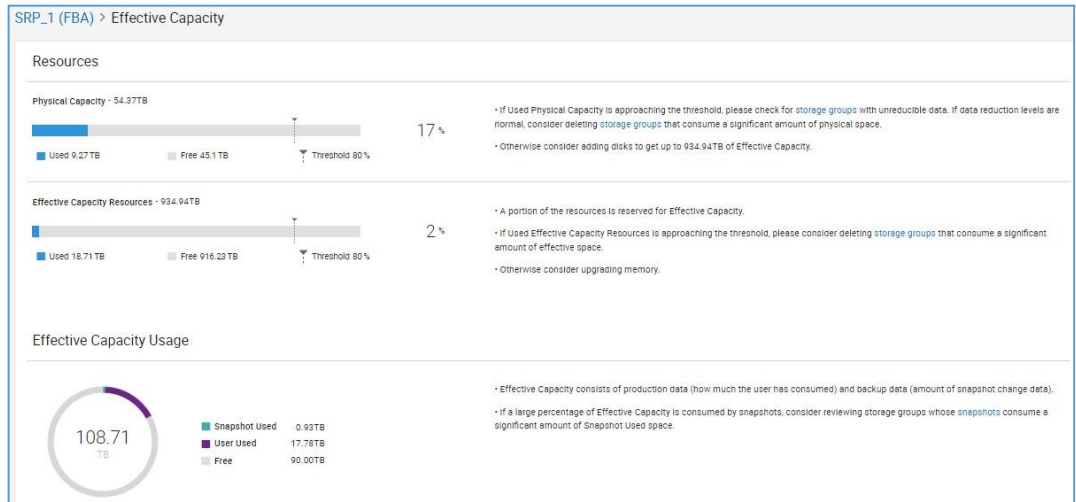


Figura 7. Tablero de capacidad real

Los valores de la página Capacidad de instantáneas se definen de la siguiente manera:

- **Instantánea real utilizada:** porcentaje de la capacidad real utilizada por los datos de instantáneas para el SRP
- **Instantánea física utilizada:** porcentaje de capacidad útil consumida por los datos de instantáneas para el SRP
- **Recursos de instantáneas:** porcentaje de metadatos de instantánea consumidos para todo el sistema
- **Recursos de instantánea utilizados:** capacidad utilizada por las instantáneas como parte de la capacidad de metadatos de instantáneas

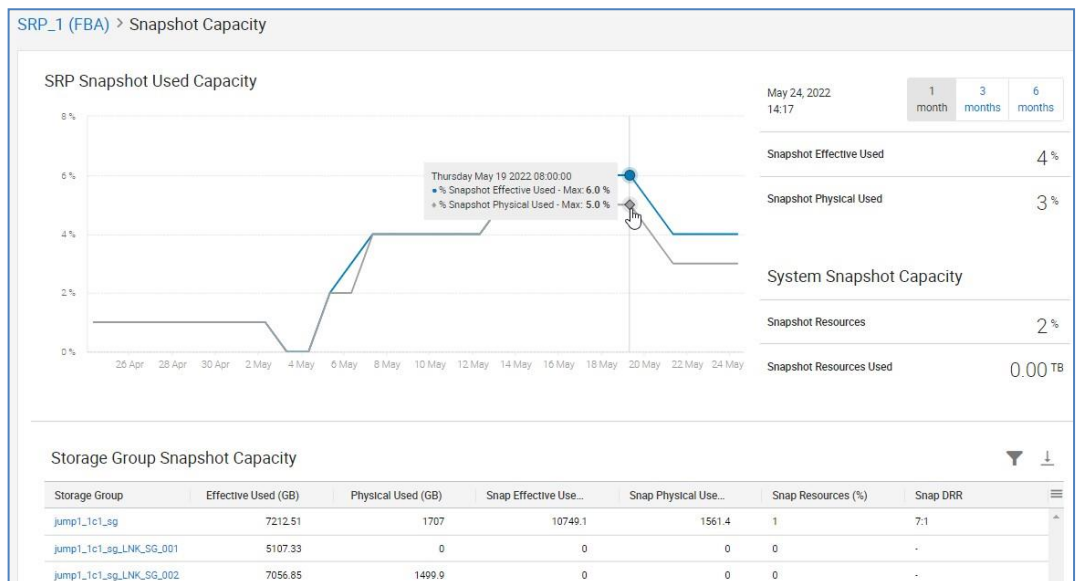


Figura 8. Uso de capacidad de la instantánea

## Tasa de reducción de datos

La visualización de reducción de datos proporciona una única ubicación para que los usuarios vean la eficiencia de la reducción de datos. Hay tres secciones: tasa de reducción de datos, un gráfico interactivo histórico y una tabla de todos los grupos de almacenamiento. La relación de reducción de datos que se muestra solo representa la habilitación y la reducción de los datos escritos en el sistema.

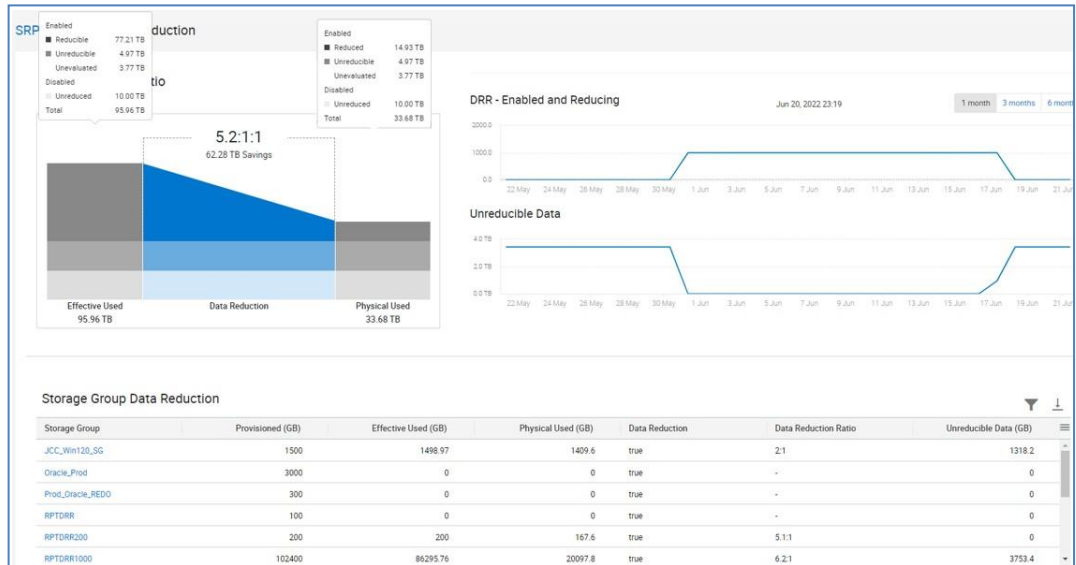


Figura 9. Visualización de reducción de datos

La **relación de reducción** de datos se muestra como un gráfico que presenta la capacidad utilizada real, la relación de reducción de datos y la capacidad física utilizada. Física utilizada se refiere a la cantidad real de capacidad física que se está utilizando. La reducción de datos presenta los ahorros como una relación.

**Real utilizada** representa los datos escritos en el sistema antes de que se logre cualquier ahorro cuando se aplica la reducción de datos. Todos los valores mostrados representan el tamaño completo tal como lo escribió el host o la aplicación. Hay dos categorías en las que se colocan los datos escritos: Habilitados y Deshabilitados.

- **Habilitados** indica que los datos que se contabilizan están habilitados para la reducción de datos y están sujetos al proceso de reducción de datos y a la función de reducción basada en la actividad. Hay tres categorías adicionales en las que los datos pueden incluirse cuando se habilita la reducción de datos, Reducibles, Irreducibles y No evaluados.
  - **Los datos reducibles** son la cantidad de datos escritos que el proceso de reducción de datos identificó como datos que se pueden reducir para utilizar menos capacidad física de la que se escribió en el sistema.
  - **Los datos irreducibles** son datos que no se pueden reducir.
  - **Los datos no evaluados** no han sido evaluados todavía por el proceso de reducción de datos. Aún no se ha determinado si los datos son reducibles o irreducibles.
- **Deshabilitados** indica que los datos escritos en el sistema no están sujetos a ningún ahorro de reducción de datos. Todos los datos identificados como deshabilitados se mostrarán como no provocados.

**Física utilizada** representa los datos escritos en el sistema después de que se hayan almacenado en el disco. Esto representa todos los datos habilitados y deshabilitados, así como todos los datos reducidos y no provocados. Hay dos categorías que representan los datos almacenados en el disco: Habilitados y Deshabilitados.



- **Habilitados** indica los datos que pasaron por el proceso de reducción de datos. Hay tres subcategorías de estos datos: Reducidos, Irreducibles y No evaluados.
  - **Los datos reducidos** recorrieron el proceso de reducción de datos. Este proceso incluye el paso a través del hardware de reducción de datos y su almacenamiento en disco. La reducción de los datos almacenados en el disco consume menos espacio de disco que lo escrito por el host o la aplicación.
  - **Irreducibles** indica que los datos se enviaron a través del proceso de reducción de datos, incluido el hardware de reducción de datos, pero no fue posible reducirlos. Algunos datos irreducibles contabilizados en la sección Física utilizada pueden contribuir al ahorro de reducción de datos, ya que los datos se comparten debido a la deduplicación.
  - **Los datos no evaluados** aún no han sido evaluados por el proceso de reducción de datos. Por lo tanto, aún no se ha determinado si los datos son reducibles o irreducibles.
- **Deshabilitados** indica que los datos escritos en el sistema no están sujetos a ningún ahorro de reducción de datos. Todos los datos identificados como deshabilitados se mostrarán como no provocados.

El gráfico interactivo que muestra la **DRR habilitada, la reducción** y los **datos irreducibles** proporciona datos históricos. Esto muestra el efecto de los datos irreducibles en la relación de reducción de datos. Este gráfico permite que el usuario rastree y monitoree los cambios en la relación de reducción de datos que pueden ser causados por datos irreducibles.

La lista de grupos de almacenamiento proporciona información sobre el uso de la capacidad y la reducción de datos específica para cada grupo de almacenamiento del sistema. Cuando se utiliza el gráfico interactivo para rastrear los cambios en la relación de reducción de datos, la lista de grupos de almacenamiento se puede utilizar para identificar grupos de almacenamiento que tienen grandes cantidades de datos irreducibles que afectan la relación de reducción de datos.

**Cálculo de las tasas de eficiencia:** los datos necesarios para calcular la tasa de reducción de datos están disponibles en una ventana emergente en el gráfico reducción de datos.

- **Relación de reducción de datos:** la relación de reducción de datos se calcula mediante el uso habilitado y reducible de real utilizado y habilitado y reducido desde el uso físico.

$$\text{Habilitado reducible} \div \text{Habilitado reducido}$$

- **Tasa de reducción de datos general:** la tasa general de reducción de datos del sistema se calcula utilizando los valores totales de Real Used y Physical Used

$$\text{Total real utilizado} \div \text{Total físico utilizado}$$

## Servicios de datos admitidos

### Visión general

La reducción de datos se admite para el almacenamiento FBA. Los sistemas combinados de FBA/CKD son compatibles con el mismo pool de recursos de almacenamiento; sin embargo, la reducción de datos para la emulación de CKD solo aprovechará la compresión y la reducción basada en la actividad. Todos los demás servicios de datos ofrecidos en los sistemas PowerMax y VMAX All Flash son compatibles. Estos servicios incluyen replicación local (SnapVX), replicación remota (SRDF), D@RE y VMware vSphere Virtual Volumes (vVols).

### Replicación local (SnapVX)

Las instantáneas de SnapVX protegen las aplicaciones sin el uso de volúmenes de destino para capturar datos modificados conocidos como deltas. Las deltas de instantáneas se mantienen automáticamente en el back-end de almacenamiento mediante punteros a las imágenes de un punto en el tiempo pertinentes. El uso compartido de recursos y la deduplicación de datos aprovechan automáticamente este diseño para proporcionar beneficios de caché, capacidad y rendimiento.

Los datos de origen comprimidos permanecen comprimidos al convertirse en deltas de instantáneas. Los datos de origen sin comprimir se pueden comprimir a medida que se convierten en delta de instantáneas o a medida que se vuelven menos activos. La actividad de lectura a través de destinos vinculados puede impedir que una delta sin comprimir se comprima o puede provocar que las deltas de instantáneas comprimidas se descompriman. Las deltas de instantáneas están disponibles para la deduplicación.

La habilitación de la reducción de datos en un destino vinculado solo afectará a los datos que pertenecen al destino vinculado. Los datos de los destinos y clones vinculados están disponibles para la deduplicación.

### Replicación remota (SRDF)

La compresión para SRDF es compatible y se conoce como compresión SRDF. La compresión SRDF es una función diseñada para reducir el consumo de ancho de banda y, al mismo tiempo, enviar datos desde y hacia sistemas conectados mediante la replicación remota. La compresión de SRDF y la reducción de datos utilizan el mismo hardware; sin embargo, tienen distintos propósitos. Los datos que se han comprimido mediante la reducción de datos se descomprimen antes de ser enviados a través del enlace de SRDF. Si se aplica la compresión SRDF y la compresión en línea, los datos se descomprimen, se comprimen mediante la función de compresión de SRDF y se envían al sitio remoto.

### Cifrado de datos en reposo (D@RE)

D@RE proporciona cifrado basado en hardware, en el arreglo y en back-end. La reducción de datos se realiza como un proceso en línea. Los datos se pasan a través del hardware de reducción de datos antes de ser enviados a través del hardware de cifrado. Por lo tanto, los datos se comprimen, deduplican o ambas cosas antes de ser cifrados. En un sistema habilitado por D@RE, los datos cifrados en el disco ya se han comprimido, deduplicado o ambas cosas.

### Volúmenes virtuales

La reducción de datos es compatible con la asignación de datos a vVols y sigue la misma ruta de I/O que todos los demás datos. La ruta de I/O se puede ver en la [Figura 1](#). La reducción de datos está habilitada en el nivel de recursos de almacenamiento en un contenedor de almacenamiento de vVols, ya que no hay grupos de almacenamiento para vVols.

## Conclusión

### Resumen

El uso de la capacidad de almacenamiento físico es una preocupación común de los administradores de almacenamiento en toda la industria de almacenamiento. Las cantidades de datos constantes y cada vez mayores han creado la necesidad de mayor eficiencia en el uso de la capacidad física. Los sistemas de almacenamiento de datos Dell PowerMax 2500 y 8500 llevan esta eficiencia al siguiente nivel. La reducción de datos proporciona ahorros de capacidad excepcionales y, al mismo tiempo, ofrece un rendimiento óptimo. Este resultado conduce a un espacio físico del centro de datos más pequeño y a una reducción general en el TCO. Además de los ahorros, el uso de la reducción de datos es tan simple como un solo clic para habilitar o deshabilitar. El sistema maneja todo el trabajo.

## Referencias

### Documen- tación de Dell Technologies

La siguiente documentación de Dell Technologies proporciona otra información relacionada con este documento. El acceso a estos documentos depende de sus credenciales de inicio de sesión. Si no tiene acceso a un documento, comuníquese con su representante de Dell Technologies.

- [Centro de información de PowerMax y VMAX](#)