

# Dell PowerMax: redução de dados

## Compactação e deduplicação em linha

Julho de 2022

H19254

### White paper

#### Resumo

As plataformas de armazenamento PowerMax apresentam várias técnicas de redução de dados, como compactação e deduplicação em linha. Eles também incluem detecção de padrões e posicionamento de dados eficiente para oferecer um excelente equilíbrio entre desempenho e eficiência.

Dell Technologies

## Direitos autorais

As informações nesta publicação são fornecidas no estado em que se encontram. A Dell Inc. não garante nenhum tipo de informação contida nesta publicação, assim como se isenta de garantias de comercialização ou adequação de um produto a um propósito específico.

O uso, a cópia e a distribuição de qualquer software descrito nesta publicação requerem uma licença de software aplicável.

Copyright © 2022 Dell Inc. ou suas subsidiárias. Todos os direitos reservados. Dell Technologies, Dell, EMC, Dell EMC e outras marcas comerciais são marcas comerciais da Dell Inc. ou de suas subsidiárias. Intel, o logotipo da Intel, o logotipo Intel Inside e Xeon são marcas comerciais da Intel Corporation nos EUA e/ou em outros países. As demais marcas comerciais podem pertencer a seus respectivos proprietários.

Publicado no Brasil em julho de 2022 H19254.

A Dell Inc. assegura que as informações apresentadas neste documento estão corretas na data da publicação. As informações estão sujeitas a alterações sem prévio aviso.

# Índice

Sumário executivo .....	4
Redução de dados .....	5
Uso de recursos do sistema.....	9
Gerenciamento e monitoramento .....	9
Serviços de dados compatíveis .....	17
Conclusão.....	18
Referências.....	19

## Sumário executivo

### Visão geral

A redução de dados com o Dell PowerMax aumenta a eficiência do sistema combinando a compactação em linha, a deduplicação em linha e a detecção de padrões. O uso dessas técnicas de redução de dados permite que os usuários obtenham grande economia de capacidade. A redução de dados compacta dados e elimina cópias redundantes. Este white paper explica como funciona a redução de dados nos sistemas PowerMax e descreve a geração de relatórios usando aplicativos de gerenciamento da Dell, como Unisphere for PowerMax, Solutions Enabler e software Mainframe Enabler.

### Revisões

Data	Descrição
Julho de 2022	Versão inicial

### Nós valorizamos seu feedback

A Dell Technologies e os autores deste documento agradecem seus comentários sobre este documento. Entre em contato com a equipe da Dell Technologies por [e-mail](#).

**Autor:** Robert Tasker

---

**Nota:** Para obter links para outras documentações deste tópico, consulte o [Hub de informações do PowerMax e VMAX](#).

---

# Redução de dados

## Visão geral

A redução de dados combina compactação em linha, desduplicação em linha, detecção de padrões, posicionamento de dados eficiente e aprendizado de máquina (ML). Essa combinação cria um sistema no qual os usuários podem gravar mais dados do host do que a quantidade total de capacidade física disponível, ao mesmo tempo que alcança o desempenho esperado de um sistema de armazenamento empresarial. Esse recurso está ativado por padrão e você pode habilitá-lo ou desabilitá-lo no nível do grupo de armazenamento. Além disso, todos os serviços de dados disponíveis nos sistemas PowerMax 2500 e 8500 são compatíveis. Esse suporte também se aplica à emulação do CKD, mas não inclui desduplicação para CKD.

A compactação reduz o tamanho dos dados, enquanto a desduplicação armazena dados como uma instância única. A detecção de padrões inclui uma função de alocação diferente de zero que impede strings de zeros consecutivos de serem armazenadas como parte dos dados compactados. A compactação, a desduplicação e a detecção de padrões são realizadas usando assistência de hardware integrada ao sistema para reduzir a sobrecarga de execução dessas funções. O aprendizado de máquina identifica os dados mais movimentados armazenados no disco e garante que eles permaneçam sem redução para obter o desempenho ideal. O posicionamento de dados eficiente usa uma função chamada **compactação**, que armazena dados estrategicamente para minimizar o espaço desperdiçado e reduz a necessidade de funções de coleta de lixo ou desfragmentação.

## Redução baseada em atividade

A redução baseada em atividade (ABR) reduz o custo de desempenho incorrido com a descompactação dos dados acessados com frequência. Essa função permite que até 20% dos dados mais usados sejam armazenados no sistema sem compactação. Essa capacidade beneficia o sistema, pois minimiza a latência de desempenho que resulta da descompactação constante dos dados acessados com frequência. Para determinar quais dados são os mais usados, o sistema usa algoritmos de ML que processam estatísticas de E/S. Realizar essa tarefa mantém um ambiente equilibrado e ideal para a economia e o desempenho da redução de dados.

## Compression

A compactação reduz para o menor tamanho possível as cargas de trabalho de gravação recebidas, com o objetivo de consumir a menor quantidade de capacidade. Os dados são compactados quando transmitidos pelo hardware de redução de dados que usa o algoritmo de compactação GZIP. Quando transmitidos pelo hardware de redução de dados, os dados são divididos em quatro seções que são compactadas em paralelo para maximizar a eficiência. A soma das quatro seções é o tamanho reduzido final dos dados armazenados no disco. Essa capacidade dá acesso granular a dados reduzidos. Somente as seções que contêm os dados para solicitações de leitura ou gravação parcial são processadas, pois cada seção pode ser tratada de modo independente.

## Desduplicação

A desduplicação é um método de economia da capacidade que identifica cópias idênticas de dados e armazena uma instância única de cada cópia. Há alguns componentes da desduplicação que são obrigatórios para que a equipe de TI forneça economia de capacidade eficiente.

- **ID de hash:** é um identificador exclusivo para dados de entrada utilizado para determinar se uma relação de desduplicação é necessária. O sistema usa um algoritmo SHA-256 para gerar a ID de hash.
- **Tabela de ID de hash:** tabelas de hash são uma alocação da memória do sistema distribuída entre os directors do sistema. Essas tabelas catalogam os IDs de hash usados pelo processo de desduplicação. As entradas na tabela são usadas para determinar se existe uma relação de desduplicação ou se uma nova entrada é necessária e os dados podem ser armazenados no disco.
- **Objeto de gerenciamento de desduplicação (DMO):** o DMO é um objeto de 64 bytes na memória do sistema que só existe quando há uma relação de desduplicação. Esses objetos armazenam

e gerenciam os ponteiros entre os dispositivos de front-end e os dados deduplicados que consomem capacidade de back-end no array. Os DMOs gerenciam os ponteiros para dados deduplicados entre os dispositivos de front-end e os dados armazenados em disco. Ele também gerencia em qual tabela de hash os IDs de hash são armazenados quando há relações de deduplicação.

A deduplicação é realizada usando o mesmo hardware de redução de dados que a compactação, e um ID de hash exclusivo é gerado quando os dados são processados pelo hardware. Em seguida, o ID de hash é comparado à tabela de ID de hash, que procura o mesmo ID. Quando uma correspondência é encontrada, os dados não são armazenados no disco e um compartilhamento de deduplicação é criado. Os ponteiros são definidos entre o volume de front-end e o ID exclusivo na tabela de ID de hash. Os ponteiros vinculam a instância única dos dados armazenados no disco ao volume, fornecendo acesso futuro aos dados. O DMO gerencia os ponteiros entre os dados, os volumes de front-end que acessam os dados e a tabela de ID de hash. Quando não há correspondência na tabela de ID de hash, uma nova entrada é adicionada para futuras comparações de ID de hash.

### Algoritmo de deduplicação

Os sistemas PowerMax usam o algoritmo de hash SHA-256 implementado no hardware de redução de dados para encontrar dados duplicados. Em seguida, os dados são armazenados como uma instância única para que várias fontes a compartilhem. Esse processo oferece maior eficiência de dados e, ao mesmo tempo, mantém a integridade deles.

O algoritmo SHA-256 gera um código de 32 bytes para cada bloco de dados de 32 KB. Considere um sistema com 1 PB de dados gravados com 5% de atualização por dia. Em 1 milhão de anos de operação, há uma probabilidade de 20% de uma colisão de hash. Como cada faixa de 128 KB é tratada como 4 blocos de 32 KB, é necessário ter uma colisão de hash em todos os quatro blocos na mesma faixa de 128 KB para haver uma colisão de hash real. As probabilidades de que os quatro colidam faz com que isso seja apenas uma possibilidade teórica (menos de 1% de chance em um trilhão de anos de operação). Além disso, quando uma correspondência é encontrada durante a fase de comparação da deduplicação, uma comparação de byte por byte é realizada. Essa comparação é feita para confirmar se há uma correspondência antes de atualizar as tabelas e definir os ponteiros para permitir o acesso aos dados.

### Compactação

O posicionamento de dados é realizado usando um processo chamado compactação. A compactação coloca intuitivamente os dados reduzidos ou não reduzidos no disco no melhor local disponível possível. A operação de armazenamento de dados no disco usa objetos de gravação. Cada objeto tem 6 MB de capacidade contígua de dispositivos de dados de back-end em todas as unidades configuradas no sistema. Os objetos de gravação são alinhados em limites de 1 K e são consumidos sequencialmente em um só uso. Os objetos de gravação são distribuídos em faixas completas para todos os tipos de RAID compatíveis a fim de otimizar as gravações. Cada objeto dá suporte a dados reduzidos ou não reduzidos para emulação do FBA e CKD.

- **Objeto de gravação do FBA:** um objeto de gravação não reduzido consiste em 48 faixas do FBA. Um objeto de gravação reduzido consiste em 1.000 faixas reduzidas. As entradas reduzidas para objetos de gravação variam de 1 KB a 96 KB.
- **Objeto de gravação do CKD:** um objeto de gravação não reduzido consiste em 108 faixas do CKD. Um objeto de gravação reduzido consiste em 1.000 faixas reduzidas. As entradas reduzidas para objetos de gravação variam de 1 KB a 52 KB.

## Compactação de dados estendida

Os sistemas PowerMax 2500 e 8500 incluem uma função extra chamada compactação de dados estendida (EDC) que compacta os dados já compactados para obter mais economia de capacidade. Isso é feito por meio da identificação de dados que não foram acessados por um tempo estendido. Veja abaixo os fatores que fazem com que os dados sejam candidatos à EDC:

- Os dados pertencem a um grupo de armazenamento habilitado para redução de dados.
- Os dados não foram acessados por 30 dias.
- Os dados ainda não foram compactados por EDC.

Os dados que se qualificam para EDC são compactados usando o algoritmo Def9\_128\_SW para reduzir ainda mais a quantidade de capacidade usada para armazenar os dados. Esse é um processo de segundo plano automatizado no sistema. Há uma economia extra incluída na taxa de compactação, atingida no nível do grupo de armazenamento. A EDC está disponível somente com storage arrays PowerMax.

## CKD compactação

A redução baseada em atividade (ABR) reduz o custo de desempenho incorrido com a descompactação dos dados acessados com frequência. Essa função permite que até 20% dos dados mais usados sejam armazenados no sistema sem compactação. Esse resultado beneficia o sistema, pois elimina o impacto negativo de desempenho que resulta da descompactação constante dos dados acessados com frequência. Para determinar o nível de uso de dados, o sistema usa algoritmos de ML que processam estatísticas coletadas da E/S recebida para os dispositivos de front-end. Essa ação mantém equilíbrio entre os recursos do sistema, fornecendo um ambiente ideal para economia de redução de dados e desempenho.

A compactação reduz para o menor tamanho possível as cargas de trabalho de gravação recebidas, com o objetivo de consumir a menor quantidade de capacidade possível. Os dados são compactados quando transmitidos pelo hardware de redução de dados integrado ao sistema que usa o algoritmo de compactação GZIP. Quando transmitidos pelo hardware de redução de dados, os dados são divididos em quatro seções que são compactadas em paralelo para maximizar a eficiência do hardware. A soma das quatro seções é o tamanho reduzido final dos dados armazenados no disco. Esse resultado fornece acesso granular para dados reduzidos quando há uma solicitação de leitura ou gravação parcial. Somente as seções que contêm os dados solicitados são processadas, pois cada seção pode ser tratada de modo independente.

O posicionamento de dados é realizado usando um processo chamado compactação. A compactação coloca intuitivamente os dados reduzidos ou não reduzidos no disco no melhor local disponível possível. A operação de armazenamento de dados no disco usa objetos de gravação. Cada objeto tem 6 MB de capacidade contígua de dispositivos de dados de back-end em todas as unidades configuradas no sistema. Os objetos de gravação são alinhados em limites de 1 K e são consumidos sequencialmente em um só uso. Os objetos de gravação são distribuídos em faixas completas para todos os tipos de RAID compatíveis a fim de otimizar as gravações. Cada objeto dá suporte a dados reduzidos ou não reduzidos. Um objeto de gravação não reduzido consiste em 108 faixas do CKD. Um objeto de gravação reduzido consiste em 1.000 faixas reduzidas. As entradas reduzidas para objetos de gravação variam de 1 KB a 52 KB.

## Fluxo de E/S de redução de dados

Todos os E/S passam pelo cache e são processados pelo sistema. As ações de redução de dados são executadas depois que os dados são recebidos pelo sistema, mas antes de serem colocados em disco. O uso de um processo em linha requer verificações adicionais no fluxo de E/S, no qual a redução de dados é aplicada. O sistema usa essas verificações para determinar se os dados de entrada precisam passar pelo hardware de redução de dados. Os dados recebidos de um grupo de armazenamento com redução de dados ativada seguirão o fluxo de redução de dados. No entanto, devido à função de compactação baseada em atividade

(ABR), os dados ativos de um grupo de armazenamento com redução de dados ativada ignorarão o fluxo de redução de dados para otimização do desempenho. Os dados não compactados devido à ABR podem ser compactados posteriormente e movidos para um pool de compactação. Os dados de um grupo de armazenamento com redução de dados desativada seguirão o fluxo de redução de dados e serão gravados no sistema sem redução.

Há alguns tipos diferentes de E/S para considerar: leitura, gravação e gravação-atualização.

- **Leitura:** uma solicitação para acessar os dados que já estão preenchendo o array.
- **Gravação:** E/S de entrada que consumirá espaço em disco.
- **Gravação-atualização:** E/S de entrada que pode alterar os dados alocados para o espaço em disco no array.

A figura a seguir descreve o caminho que a E/S fará, definido pelas características do conjunto de dados ou do grupo de armazenamento relacionado.

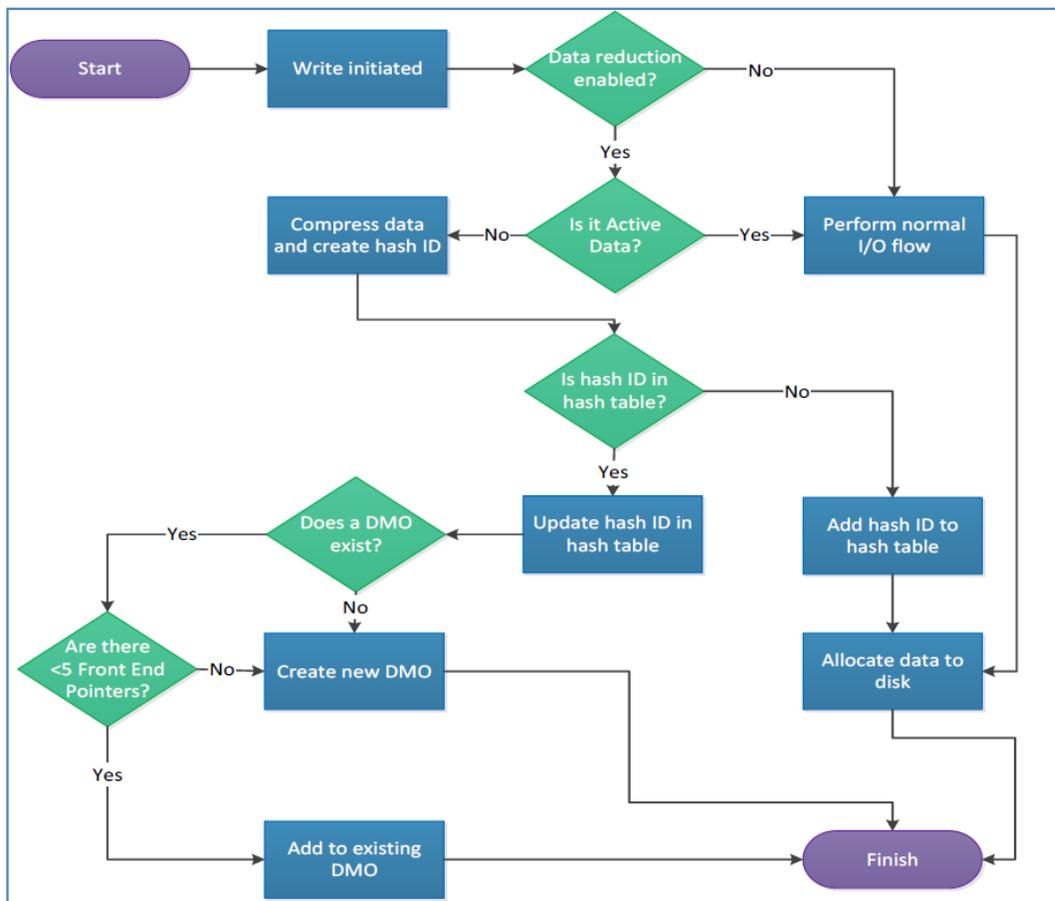


Figura 1. Fluxo de E/S de redução de dados para sistemas de armazenamento empresarial de dados do PowerMax

## Uso de recursos do sistema

### Visão geral

Capacidade e memória são os dois principais recursos configurados em cada sistema PowerMax. A capacidade é categorizada em física, efetiva e provisionada. A memória é categorizada como recursos do sistema.

Os recursos de memória são compatíveis com as estruturas de metadados para capacidade provisionada, bem como a capacidade física. A quantidade de capacidade efetiva disponível está relacionada à quantidade de capacidade física, à quantidade de recursos do sistema disponíveis e à redundância dos dados gravados no sistema. Os dados gravados que são altamente redutíveis consomem menos capacidade física, resultando em capacidade mais efetiva. O contrário também se aplica: os dados gravados que não são redutíveis podem resultar em uma capacidade efetiva menos disponível. As informações descritas na seção [Redução de dados](#) (capacidade, recursos do sistema) estão disponíveis nos aplicativos de gerenciamento usados para sistemas PowerMax 2500 e 8500, Unisphere for PowerMax, Solutions Enabler e software Mainframe Enabler. O Unisphere for PowerMax é uma interface do usuário (UI) que fornece dados em gráficos, tabelas e formato de lista. O Solutions Enabler é uma interface de linha de comando padrão que fornece os mesmos dados, mas não na forma de gráficos e tabelas. Os Mainframe Enablers são uma suíte de componentes que monitoram e gerenciam os sistemas Dell Storage em um ambiente de mainframe. As imagens mostradas nas próximas seções deste artigo descrevem o Unisphere for PowerMax gerenciando um sistema PowerMax 2500 ou 8500.

### Capacidade física

A capacidade física é a quantidade de espaço em disco configurada no sistema com base nos discos instalados e na proteção RAID aplicada. Em uma configuração em que a redução de dados não está em uso, a capacidade física é a quantidade total de capacidade disponível para dados do host. Por exemplo, um sistema que mostra 100 TB de capacidade física indica que pode acomodar 100 TB de dados do host que não estão usando a redução de dados.

### Capacidade efetiva

A capacidade efetiva é a quantidade de espaço disponível quando a redução de dados está em uso. A quantidade total na instalação inicial depende da quantidade de memória configurada no sistema e é baseada em uma economia de redução de dados padrão de 4:1 (3:1 para emulação do CKD). Por exemplo, esse mesmo sistema com 100 TB de capacidade física mostrará 400 TB de capacidade efetiva. Esse valor de 400 TB é um ponto inicial de capacidade efetiva e mudará à medida que os dados forem gravados no sistema e a redução de dados será aplicada.

### Capacidade provisionada

A capacidade provisionada é a representação da capacidade disponível na forma de dispositivos criados e apresentados a hosts e aplicativos que pretendem consumir capacidade física ou efetiva no sistema.

## Gerenciamento e monitoramento

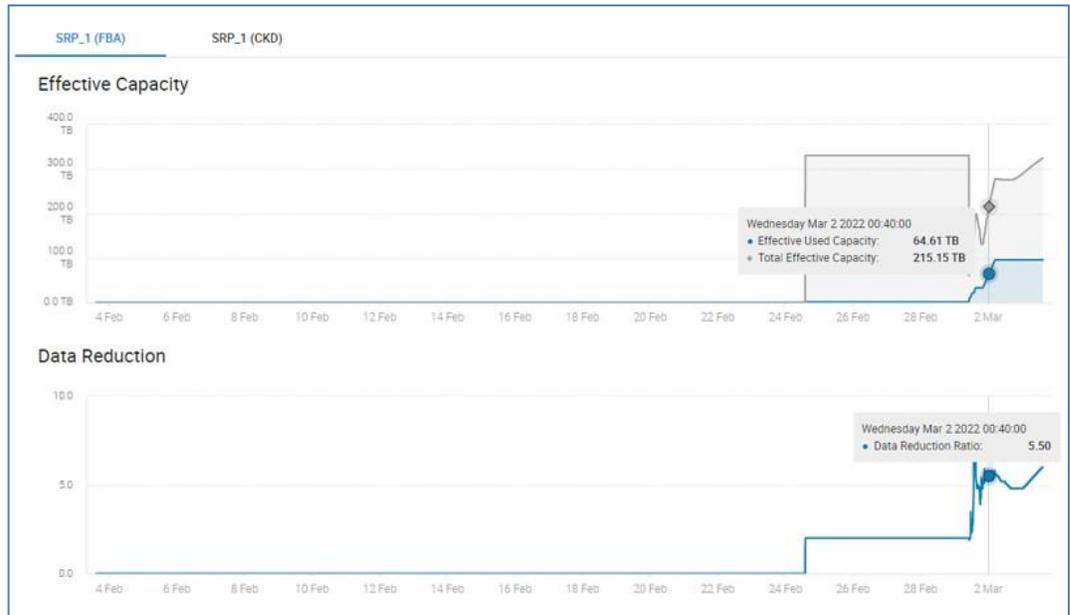
### Visão geral

O Unisphere for PowerMax é uma interface do usuário usada para gerenciar e monitorar a capacidade e o uso de recursos do sistema. No nível do sistema, as informações sobre o uso da capacidade, a redução de dados e os recursos do sistema são exibidas no painel de indicadores de capacidade. No painel de indicadores de capacidade, os usuários podem acessar telas que exibem informações sobre capacidade efetiva e provisionada, capacidade de snapshot e redução de dados, bem como recursos do sistema.

### Painel de indicadores de capacidade

No Unisphere for PowerMax, há vários monitores que fornecem informações relacionadas ao uso da capacidade.

O painel de indicadores principal exibe um gráfico interativo que mostra o uso efetivo da capacidade e a redução de dados ao longo do tempo. Esta tela mostra o histórico de uso efetivo da capacidade e como a taxa de redução de dados está relacionada à capacidade efetiva. Essas informações podem ser usadas para monitorar e rastrear tendências de uso efetivo da capacidade em relação à taxa de redução de dados que está sendo mostrada. Os sistemas PowerMax 2500 ou 8500 podem ser configurados com emulação do CKD e FBA no mesmo pool de recursos de armazenamento, mas o gráfico histórico é específico para a exibição de emulação selecionada.



**Figura 2. Gráfico histórico do painel de indicadores de capacidade mostrando a capacidade efetiva e a redução de dados para emulação de FBA**

O painel de indicadores principal também oferece dados na forma de gráficos de barras para capacidade provisionada, capacidade efetiva, uso de snapshots e redução de dados. Cada seção pode ser expandida em uma exibição mais detalhada mostrando dados mais granulares para cada item.

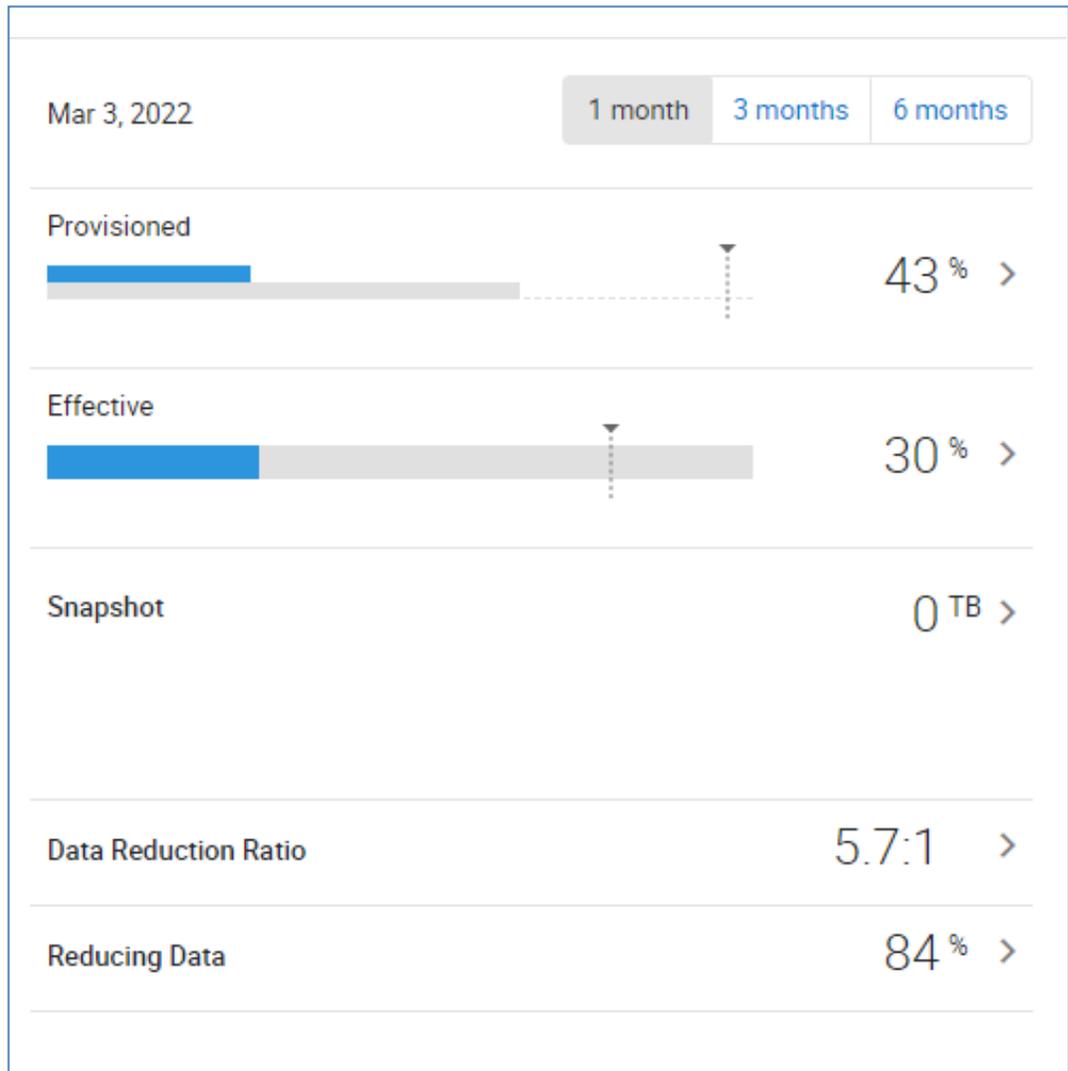


Figura 3. Gráficos de barras do painel de indicadores de capacidade referentes a capacidade provisionada, capacidade efetiva, uso de snapshots e redução de dados

### Provisioned

Capacidade provisionada é a quantidade de capacidade que é provisionada na forma de dispositivos e apresentada aos hosts e aplicativos como capacidade disponível. A capacidade provisionada é rastreada usando duas métricas: capacidade do SRP e recursos do sistema.

- **A capacidade do SRP** mostra a quantidade de capacidade provisionada como um valor de TB provisionado e a quantidade disponível de capacidade efetiva. A quantidade de capacidade efetiva do SRP é inicialmente baseada em uma economia padrão de redução de dados de 4:1 (3:1 para emulação do CKD). Esse valor é calculado usando a quantidade de capacidade física configurada no sistema. À medida que os usuários criam dispositivos, a capacidade provisionada aumenta. O valor percentual mostrado é o percentual de assinatura e é calculado usando o valor provisionado e a quantidade de capacidade efetiva.

- **Os recursos do sistema** representam a quantidade de memória disponível para dar suporte à capacidade provisionada na forma de metadados. O valor total apresentado só será alterado se o cache for adicionado ao sistema. A quantidade usada representa exatamente isso. A quantidade livre é uma indicação de quanto de capacidade extra provisionada o sistema pode aceitar. À medida que os usuários criam dispositivos, a quantidade usada aumenta. Quando há uma variação entre os dois valores, os recursos de memória são usados para dar suporte à redução de dados ou outros recursos que usam memória, como capturar snapshots de dispositivos existentes.

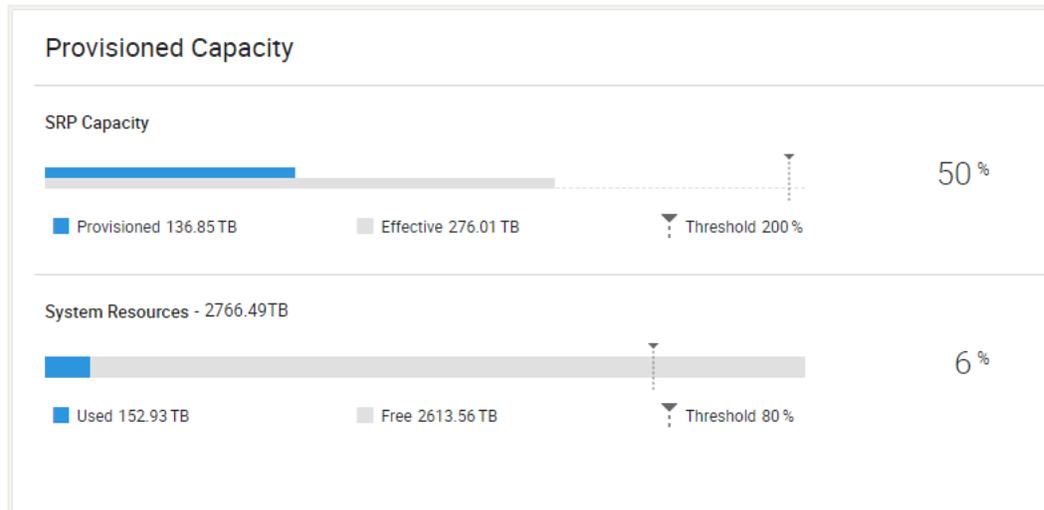


Figura 4. Uso de recursos do sistema dentro da exibição de capacidade provisionada

## Efetiva

A capacidade efetiva representa a quantidade de capacidade disponível para o usuário com base na expectativa de economia proveniente da redução de dados. A exibição de capacidade efetiva fornece uma visão detalhada dos recursos físicos e efetivos disponíveis. Isso é mostrado em três seções: Capacidade física, Recursos de capacidade efetiva e Uso de capacidade efetiva.

- **Capacidade física** mostra a quantidade de capacidade física disponível nos discos rígidos configurados no sistema. Os valores exibidos são os valores resultantes depois que a formatação e a proteção RAID são aplicadas. O valor exibido é a quantidade de capacidade aceita pelo sistema para os dados do host quando a redução de dados não está sendo usada.
- **Recursos de capacidade efetiva** indicam os valores atingíveis com base no uso atual de recursos do sistema. O valor dos recursos de capacidade efetiva exibido será ajustado em relação à economia gerada pela redução de dados atual e ao uso de capacidade física e efetiva.
- **Uso de capacidade efetiva** exibe a quantidade atual de capacidade efetiva que está disponível com base no uso de recursos do sistema e na economia atual da redução de dados. O valor exibido no gráfico circular é a capacidade efetiva atual disponível. Os valores apresentados à direita dividem o uso em três categorias: Uso de snapshots, Usuário usado e Livre.

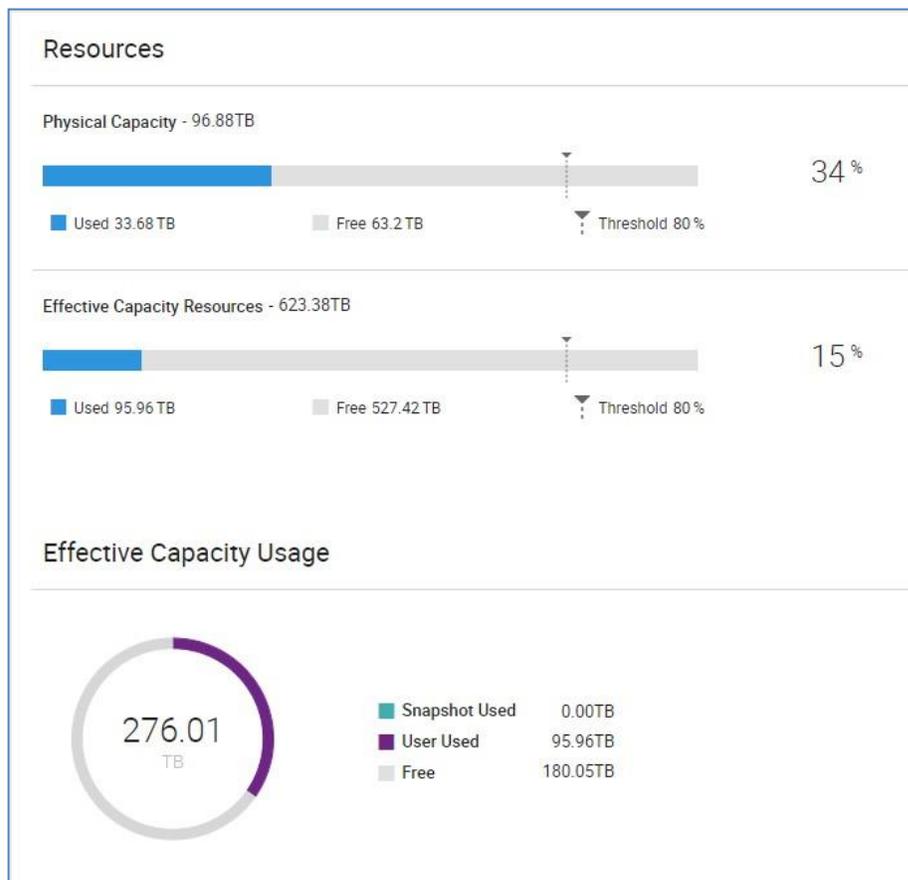


Figura 5. Uso de recursos mostrando a exibição de capacidade efetiva

## Snapshot

A capacidade de snapshot de back-end pode ser significativamente menor do que os deltas por snapshot devido à eficiência de recursos, como alocações compartilhadas e redução de dados.

Passa o mouse sobre o gráfico de barras de snapshots no Painel de indicadores de capacidade para ver detalhes de alto nível. Os valores de snapshot são definidos como:

- **Usada:** capacidade efetiva utilizada para dados de alteração de snapshot
- **Livre:** quantidade restante de dados de alteração de snapshot com base no volume usado e nos metadados restantes
- **Total:** usada + livre
- **Limite:** limite para alertas sobre dados de alteração de snapshot

Clique nos gráficos de barras para acessar o painel de indicadores Capacidade efetiva e o painel de indicadores Capacidade de snapshot.

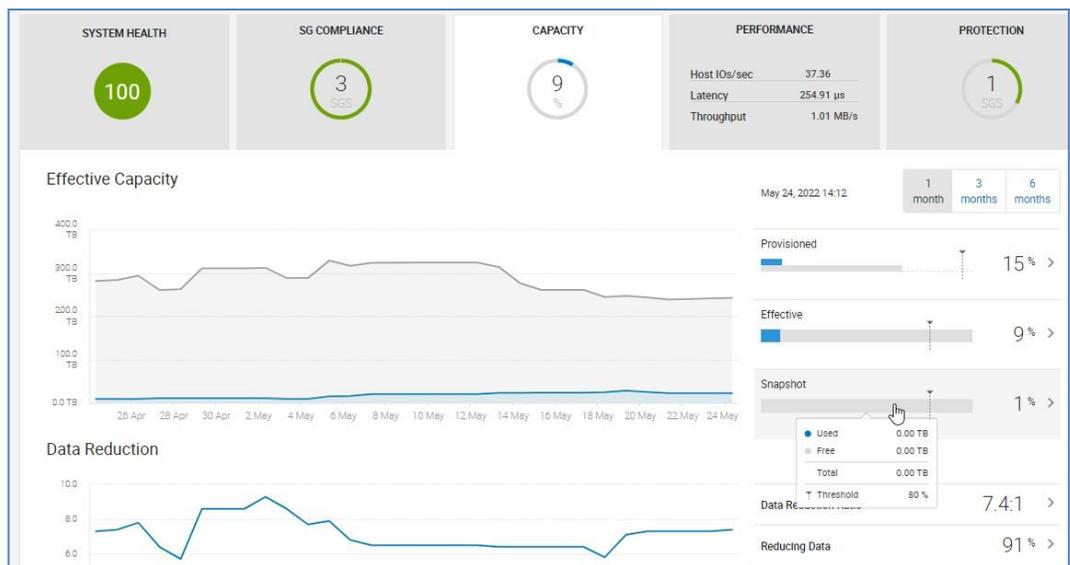


Figura 6. Painel de indicadores de capacidade

No painel de indicadores Capacidade efetiva:

- **Snapshot usado:** quantidade de capacidade efetiva usada pelos dados de alteração de snapshot

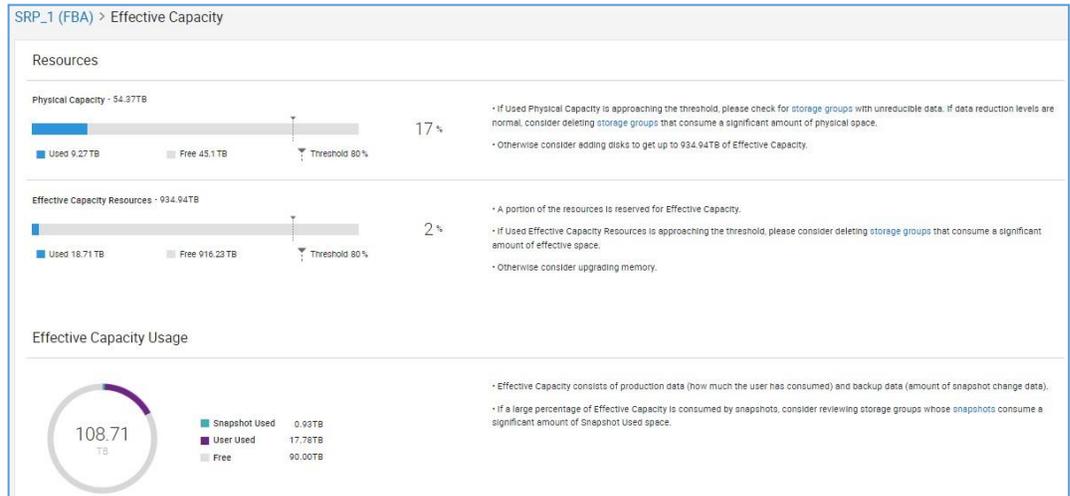


Figura 7. Painel de indicadores de capacidade efetiva

Os valores na página Capacidade de snapshot são definidos como:

- **Snapshot efetivo usado:** porcentagem de capacidade efetiva usada pelos dados de snapshot para o SRP
- **Snapshot físico usado:** porcentagem de capacidade útil consumida pelos dados de snapshot para o SRP
- **Recursos de snapshot:** porcentagem de metadados de snapshot consumidos para todo o sistema
- **Recursos de snapshot usados:** capacidade usada por snapshots como parte da capacidade de metadados de snapshot

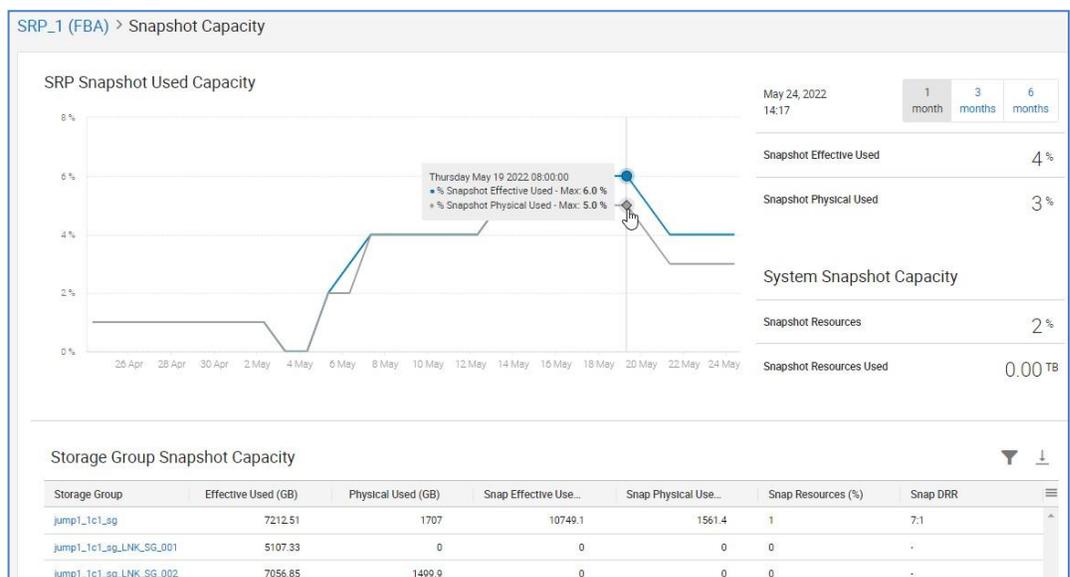


Figura 8. Uso da capacidade de snapshot

## Razão de redução de dados

A exibição de redução de dados fornece um local único para que os usuários visualizem a eficiência da redução de dados. Há três seções: Taxa de redução de dados, um gráfico histórico interativo e uma tabela de todos os grupos de armazenamento. A taxa de redução de dados exibida representa apenas os dados habilitados e de redução gravados no sistema.

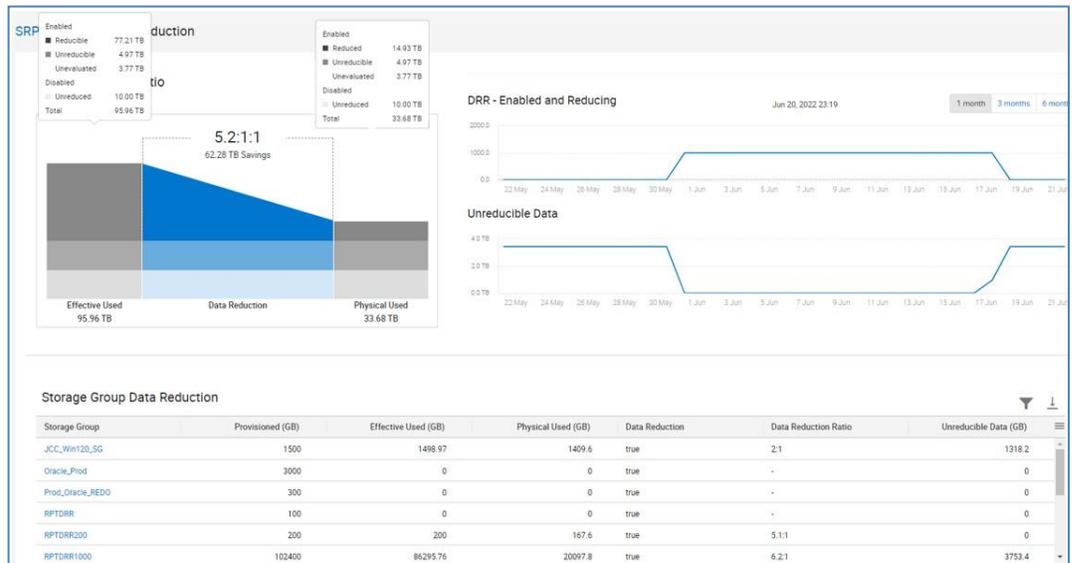


Figura 9. Exibição de redução de dados

A **taxa de redução de dados** é exibida como um gráfico que apresenta a capacidade efetiva utilizada, a taxa de redução de dados e a capacidade física utilizada. Física utilizada refere-se à quantidade real de capacidade física que está sendo usada. A redução de dados apresenta a economia como um índice.

**Efetiva utilizada** representa os dados gravados no sistema antes da obtenção de qualquer economia quando a redução de dados é aplicada. Todos os valores exibidos representam o tamanho total conforme foi gravado pelo host ou pelo aplicativo. Há duas categorias em que os dados gravados são distribuídos: Habilitados e Desabilitados.

- **Habilitados** indica que os dados que estão sendo contabilizados são habilitados para redução e estão sujeitos ao processo de redução de dados e à função de redução baseada em atividade. Há três categorias adicionais em que os dados podem se enquadrar quando a redução de dados está habilitada, Redutíveis, Não redutíveis e Não avaliados.
  - **Dados redutíveis** é o volume de dados gravados identificados pelo processo de redução como dados que podem ser reduzidos para usar menos capacidade física do que foi gravado no sistema.
  - **Dados não redutíveis** são dados que não podem ser reduzidos.
  - **Dados não avaliados** são aqueles que ainda não foram avaliados pelo processo de redução de dados. Ainda não foi determinado se os dados são redutíveis ou não.
- **Desabilitados** indica que os dados gravados no sistema não estão sujeitos a nenhuma economia pela redução de dados. Todos os dados identificados como desabilitados serão mostrados como não redutíveis.

**Física utilizada** representa os dados gravados no sistema depois de serem armazenados no disco. Isso representa todos os dados habilitados e desabilitados, bem como todos os dados redutíveis e não redutíveis. Há duas categorias que representam os dados armazenados no disco: Habilitados e Desabilitados.

- **Habilitados** indica dados que passaram pelo processo de redução. Há três subcategorias desses dados: Reduzidos, Não redutíveis e Não avaliados.
  - Dados **reduzidos** foram enviados por meio do processo de redução de dados. Esse processo inclui a passagem pelo hardware de redução de dados e o armazenamento em disco. Os dados reduzidos armazenados no disco consomem menos espaço em disco do que o que foi gravado pelo host ou pelo aplicativo.
  - **Não redutíveis** indica que os dados foram enviados por meio do processo de redução, inclusive o hardware de redução de dados, mas não foi possível reduzi-los. Alguns dados não redutíveis considerados na seção Física usada podem estar contribuindo para a economia da redução de dados como dados compartilhados devido à desduplicação.
  - Dados **não avaliados** são dados ainda não avaliados pelo processo de redução de dados. Portanto, ainda não foi determinado se os dados são redutíveis ou não.
- **Desabilitados** indica que os dados gravados no sistema não estão sujeitos a nenhuma economia pela redução de dados. Todos os dados identificados como desabilitados serão mostrados como não redutíveis.

O gráfico interativo que mapeia **dados habilitados, de redução e não redutíveis de DRR** apresenta dados históricos. Isso mostra o efeito dos dados não redutíveis na taxa de redução de dados. Esse gráfico permite que o usuário rastreie e monitore as alterações na taxa de redução de dados que podem ser causadas por dados não redutíveis.

A lista de grupos de armazenamento apresenta informações de uso de capacidade e redução de dados específicas para cada grupo de armazenamento no sistema. Ao usar o gráfico interativo para rastrear as alterações na taxa de redução de dados, a lista de grupos de armazenamento pode ser usada para identificar grupos de armazenamento que têm grandes volumes de dados não redutíveis que estão afetando a taxa de redução de dados.

**Calculando taxas de eficiência:** os dados necessários para calcular a taxa de redução de dados estão disponíveis em uma janela pop-up no gráfico de redução de dados.

- **Taxa de redução de dados:** a taxa de redução de dados é calculada usando Habilitados e Redutíveis de Efetiva usada e Habilitados e Reduzidos de Física usada.
 
$$\text{Habilitados redutíveis} \div \text{habilitados reduzidos}$$
- **Taxa geral de redução de dados:** a taxa geral de redução de dados do sistema é calculada usando os valores totais de Efetiva usada e Física usada
 
$$\text{Efetiva usada total} \div \text{Física usada total}$$

## Serviços de dados compatíveis

### Visão geral

A redução de dados é compatível com o armazenamento FBA. Sistemas mistos FBA/CKD são compatíveis com o mesmo pool de recursos de armazenamento. No entanto, a redução de dados para emulação do CKD aproveitará apenas a compactação e a redução baseada em atividade. Todos os outros serviços de dados oferecidos nos sistemas PowerMax e VMAX All Flash são compatíveis. Esses serviços incluem replicação local (SnapVX), replicação remota (SRDF), D@RE e VMware vSphere Virtual Volumes (vVols).

### Replicação local (SnapVX)

Os snapshots do SnapVX protegem os aplicativos sem o uso de volumes de destino para capturar dados de alteração conhecidos como deltas. Os deltas de snapshot são mantidos automaticamente no back-end de armazenamento usando ponteiros para as imagens point-in-time relevantes. O compartilhamento de recursos

## Conclusão

e a deduplicação de dados aproveitam automaticamente esse design para oferecer benefícios de cache, capacidade e desempenho.

Os dados de origem compactados permanecem assim ao se tornarem delta de snapshot. Os dados de origem descompactados podem se tornar compactados à medida que se tornam delta de snapshot ou à medida que se tornam menos ativos. A atividade de leitura por meio de destinos vinculados pode impedir que um delta descompactado seja compactado ou pode fazer com que os deltas de snapshots compactados sejam descompactados. Os deltas de snapshot estão disponíveis para deduplicação.

Habilitar a redução de dados em um destino vinculado só afetará os dados pertencentes ao destino vinculado. Os dados em destinos e clones vinculados estão disponíveis para deduplicação.

### Replicação remota (SRDF)

A compactação para a SRDF é compatível e é conhecida como compactação da SRDF. A compactação da SRDF é um recurso projetado para reduzir o consumo de largura de banda e, ao mesmo tempo, enviar dados de e para sistemas conectados com replicação remota. A compactação e a redução de dados da SRDF usam o mesmo hardware; no entanto, eles atendem a diferentes finalidades. Os dados que foram compactados com a redução de dados não foram compactados antes de serem enviados pelo link da SRDF. Se a compactação da SRDF e a compactação em linha forem aplicadas, os dados serão descompactados com a função de compactação da SRDF, sendo enviados para o local remoto.

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

A D@RE fornece criptografia de back-end baseada em hardware e em array, e a redução de dados é realizada como um processo em linha. Os dados são transmitidos pelo hardware de redução de dados antes de serem enviados pelo hardware de criptografia. Portanto, os dados são compactados, deduplicados ou ambos antes de serem criptografados. Em um D@RE, os dados do sistema ativado criptografados em disco já foram compactados, deduplicados ou ambos.

### Virtual Volumes

A redução de dados é compatível com a alocação de dados para vVols e segue o mesmo caminho do E/S como todos os outros dados. O caminho do E/S pode ser visto na [Figura 1](#). A redução de dados é habilitada no nível do recurso de armazenamento em um contêiner de armazenamento vVol, pois não há grupos de armazenamento para vVols.

## Conclusão

### Resumo

O uso da capacidade de armazenamento físico é uma preocupação comum dos administradores de armazenamento em todo o setor. As quantidades constantes e cada vez maiores de dados geraram a necessidade de mais eficiência no uso da capacidade física. Os sistemas de armazenamento de dados Dell PowerMax 2500 e 8500 levam essa eficiência para o próximo nível. A redução de dados proporciona uma economia de capacidade excepcional e, ao mesmo tempo, oferece o desempenho ideal. Isso resulta em menos espaço ocupado pelo data center e uma redução geral no TCO. Além da economia, o uso de redução de dados é tão simples quanto um único clique para ativá-la ou desativá-la. O sistema gerencia todo o trabalho.

## Referências

### Documentação da Dell Technologies

A documentação a seguir da Dell Technologies apresenta outras informações relacionadas a este documento. O acesso a esses documentos depende de suas credenciais de log-in. Caso você não tenha acesso a determinado documento, entre em contato com o representante da Dell Technologies.

- [Hub de informações do PowerMax e do VMAX](#)