# Dell PowerMax: Data Reduction

## Inline compression and deduplication

July 2022

H19254

White Paper

Abstract

PowerMax storage platforms feature multiple data-reduction techniques such as inline compression and deduplication. They also include pattern detection and efficient data placement to deliver a great balance of performance and efficiency.

Dell Technologies

Copyright

# Contents

# Executive summary

## Overview

Data reduction with Dell PowerMax boosts system efficiency by combining inline compression, inline deduplication, and pattern detection. Using these data-reduction techniques permits users to achieve great capacity savings. Data reduction compresses data and eliminates redundant copies of data. This white paper explains how data reduction functions in the PowerMax systems and describes reporting using Dell management applications such as Unisphere for PowerMax, Solutions Enabler, and Mainframe Enabler Software.

## Revisions

| Date | Description |
|---|---|
| July 2022 | Initial release |

## We value your feedback

Dell Technologies and the authors of this document welcome your feedback on this document. Contact the Dell Technologies team by email.

**Author:** Robert Tasker

**Note**: For links to other documentation for this topic, see the PowerMax and VMAX Info Hub.

# Data reduction

## Overview

Data reduction combines inline compression, inline deduplication, pattern detection, efficient data placement, and machine learning (ML). This combination creates a system that users can write more host data to than the total amount of physical capacity available, while achieving the performance expected from an enterprise storage system. This feature is on by default, and you can enable or disable it at the storage group level. Also, all data services available in PowerMax 2500 and 8500 systems are supported. This support applies to CKD emulation also, but it does not include deduplication for CKD.

Compression reduces the size of data while deduplication (dedupe) stores data as a single instance. Pattern detection includes a non-zero allocate function that excludes strings of consecutive zeros from being stored as part of the compressed data. Compression, dedupe, and pattern detection are performed using hardware assistance integrated within the system to reduce the overhead of performing these functions. Machine learning identifies the busiest data stored on disk and ensures it remains unreduced for optimal performance. Efficient data placement uses a function called **compaction** which strategically stores data to minimize wasted space and reduces the need for garbage collection or defragmentation (defrag) functions.

## Activity Based Reduction

Activity Based Reduction (ABR) reduces the performance cost incurred by decompressing data that is accessed frequently. This function allows up to 20% of the busiest data to be stored on the system uncompressed. This ability benefits the system as it minimizes performance latency that results from constantly decompressing frequently accessed data. To determine which data is the busiest, the system uses ML algorithms that process IO statistics. Doing this task maintains a balanced, optimal environment for both data reduction savings and performance.

## Compression

Compression reduces incoming write workloads to the smallest possible size to consume the least amount of capacity. Data is compressed when passed through data reduction hardware that uses the GZIP compression algorithm. When passed through data reduction hardware, data is divided into four sections that are compressed in parallel to maximize efficiency. The sum of the four sections is the final reduced size of the data that is stored on disk. This ability provides granular access of reduced data. Only the sections that contain the requested data for partial read or write requests are processed as each section can be handled independently.

## Deduplication

Deduplication is a capacity-savings method that identifies identical copies of data and stores a single instance of each copy. There are a few deduplication components that are required for it to provide efficient capacity savings.

- **Hash ID**: The Hash ID is a unique identifier for incoming data that is used to determine if a dedupe relationship is needed. The system uses a SHA-256 algorithm to generate the Hash ID.
- **Hash ID Table**: Hash Tables are an allocation of system memory distributed between the system directors. These tables catalog the Hash IDs used by the dedupe process. Entries in the table are used to determine if a dedupe relationship exists, or if a new entry is required and the data can be stored on disk.
- **Dedupe Management Object (DMO):** The DMO is a 64-byte object within system memory that only exists when a dedupe relationship exists. These objects store

and manage the pointers between front-end devices and the deduplicated data that consumes backend capacity in the array. DMOs manage the pointers for deduped data between front-end devices and the data stored on disk. This also manages what Hash Table the Hash IDs are stored in when dedupe relationships exist.

Deduplication is performed using the same data reduction hardware as compression, and a unique Hash ID is generated when data is processed by the hardware. Then, the Hash ID is compared to the Hash ID table looking for the same ID. When there is a match found, the data is not stored on disk, and a dedupe share is created. Pointers are set between the front-end volume and the unique ID in the Hash ID table. The pointers link the single instance of data stored on disk to the volume, providing future access to the data. The DMO manages the pointers between the data, the front-end volumes accessing the data, and the Hash ID table. When there is no match in the Hash ID table, a new entry is added to for future Hash ID comparisons.

## Deduplication algorithm

PowerMax systems use the SHA-256 hashing algorithm implemented in the data reduction hardware to find duplicate data. Then, the data is stored as a single instance for multiple sources to share. This process provides enhanced data efficiency while maintaining data integrity.

The SHA-256 algorithm generates a 32-byte code for each 32 KB block of data. Consider a system with 1 PB of written data with 5% updated per day. In one million years of operation, there is a 20% likelihood of a hash collision. As each 128 KB track is handled as four blocks of 32 KB, there would need to be a hash collision on all four blocks in the same 128 KB track to have an actual hash collision. The odds of having all four blocks collide makes this scenario only theoretical (less than a 1% chance in a trillion years of operation). Also, when there is a match found during the compare phase of deduplication, a byte-for-byte comparison is performed. This comparison is done to confirm there is a match before updating the tables and setting the pointers to allow access to the data.

## Compaction

Data placement is performed using a process called compaction. Compaction intuitively places reduced or unreduced data on disk in the best possible location available. The operation of storing data on disk uses write objects. Each object is 6 MB of contiguous back-end data-device capacity across the drives configured in the system. Write objects are aligned on 1 K boundaries and are consumed sequentially in a single use. Write objects are spread across full stripes for all supported RAID types to optimize writes. Each object supports reduced or un-reduced data for both FBA and CKD emulation.

- **FBA Write Object**: An unreduced write object consists of 48 FBA tracks. A reduced write object consists of 1000 reduced tracks. Reduced entries for write objects range from 1 KB to 96 KB.
- **CKD Write Object**: An unreduced write object consists of 108 CKD tracks. A reduced write object consists of 1000 reduced tracks. Reduced entries for write objects range from 1 KB to 52 KB.

## Extended data compression

PowerMax 2500 and 8500 systems include an extra function called extended data compression (EDC) that compresses already compressed data to gain further capacity savings. This task is accomplished by identifying data that has not been accessed for an extended time. The factors that make data a candidate for EDC is listed below:

- The data belongs to a data reduction enabled storage group.
- The data has not been accessed for 30 days.
- The data is not already compressed by EDC.

Data that qualifies for EDC is compressed using the Def9_128_SW algorithm to further reduce the amount of capacity used to store the data. This process is an automated background process within the system. Extra savings are included in the storage group. level achieved compression ratio. EDC is only available with PowerMax storage arrays.

## CKD compression

Activity Based Reduction (ABR) reduces the performance cost incurred by decompressing data that is accessed frequently. This function allows up to 20% of the busiest data to be stored on the system uncompressed. This result benefits the system as it eliminates the negative performance impact that results from constantly decompressing frequently accessed data. To determine the busy level of data, the system uses ML algorithms that process statistics collected from incoming I/O to the front-end devices. This action maintains balance between the system resources providing an optimal environment for both data reduction savings and performance.

Compression reduces incoming write workloads to the smallest possible size to consume the least amount of capacity possible. Data is compressed when passed through data reduction hardware built into the system that uses the GZIP compression algorithm. When passed through data reduction hardware, data is divided into four sections that are compressed in parallel to maximize the efficiency of the hardware. The sum of the four sections is the final reduced size of the data that is stored on disk. This result provides granular access for reduced data when there is a partial read or write request. Only the sections that contain the requested data are processed as each section can be handled independently.

Data placement is performed using a data placement process called compaction. Compaction intuitively places reduced or unreduced data on disk in the best possible location available. The operation of storing data on disk uses write objects. Each object is 6 MB of contiguous back-end data device capacity across the drives configured in the system. Write objects are aligned on 1 K boundaries and are consumed sequentially in single use. Write objects spread across full stripes for all RAID types supported to optimize writes. Each object supports reduced or un-reduced data. An unreduced write object consists of 108 CKD tracks.  A reduced write object consists of 1000 reduced tracks. Reduced entries for write objects range from 1 KB to 52 KB.

## Data reduction I/O flow

All I/O is passed through cache and then processed by the system. Data reduction actions are performed after the data is received by the system before it is placed on disk. Using an inline process requires additional checks within the I/O flow where data reduction applies. The system uses these checks to determine whether incoming data needs to pass through the data reduction hardware or not. Incoming data for a storage group with data reduction enabled will follow the data reduction flow. However, due to activity-based reduction (ABR), active data for a storage group with data reduction enabled will skip the data reduction flow for performance optimization. Data not compressed due to ABR may be compressed later and moved to a compression pool. Data for a storage group with data reduction disabled will ignore the data reduction flow and will be written to the system unreduced.

There are a few different I/O types to consider: Read, Write, and Write-update.

- **Read**: A request to access data that is already populating the array.
- **Write**: Incoming I/O that will consume disk space.
- **Write-update**: Incoming I/O that can change data that is allocated to disk space on the array.

The following figure describes the path the I/O will follow which is determined by characteristics of the dataset or the related storage group.

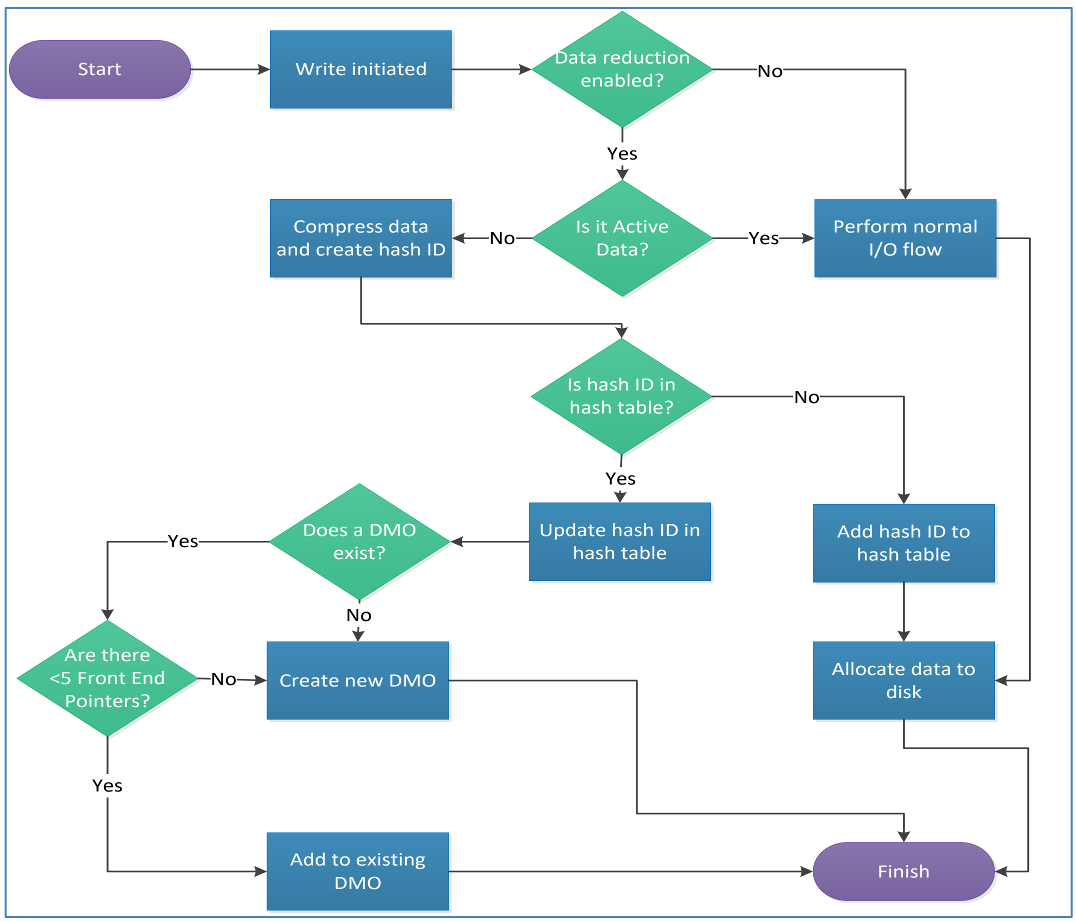


**Figure 1. Data reduction I/O flow for PowerMax enterprise storage systems**

# System resource usage

## Overview

Capacity and memory are the two primary resources configured in each PowerMax system. Capacity is categorized into Physical, Effective and Provisioned. Memory is categorized as system resources.

Memory resources support the metadata structures for provisioned capacity as well as the physical capacity. The amount of effective capacity available is related to the amount of physical capacity, the amount of system resources available, and the reducibility of data written to the system. Data written that is highly reducible consume less physical capacity resulting in more effective capacity. The adverse is also true: Data written that is not reducible can result in less available effective capacity. The information described in the Data reduction section (capacity, system resources) is available within the management applications used for PowerMax 2500 and 8500 systems, Unisphere for PowerMax, Solutions Enabler, and Mainframe Enabler Software. Unisphere for PowerMax is a user interface (UI) that provides data in graphs, charts, and list form. Solutions Enabler is a standard command-line interface providing the same data, however not in the form of charts and graphs. Mainframe Enablers are a suite of components that monitor and manage the Dell Storage systems in a mainframe environment. The images shown in the next sections of this paper depict Unisphere for PowerMax managing a PowerMax 2500 or 8500 system.

## Physical capacity

The physical capacity is the amount of disk space configured in the system based on the disks installed and RAID protection applied. In a configuration where data reduction is not in use, the physical capacity is the total amount of capacity available for host data. For example, a system showing 100 TB of physical capacity indicates it can accommodate 100 TB of host data that is not using data reduction.

## Effective capacity

The effective capacity is the amount of space available when data reduction is in use. The total amount at initial installation depends on the amount of memory configured in the system and is based on a default data reduction savings of 4:1 (3:1 for CKD emulation). For example, that same system with 100 TB of physical capacity will show 400 TB of effective capacity. This value of 400 TB is a starting point of effective capacity and will change as data is written to the system and data reduction applied.

## Provisioned capacity

The provisioned capacity is the representation of available capacity in the form of devices that are created and presented to hosts and applications that intend on consuming physical or effective capacity in the system.

# Managing and monitoring

## Overview

Unisphere for PowerMax is a user interface used to manage and monitor capacity and resource usage of the system. At a system level, information about capacity usage, data reduction, and system resources is displayed in the capacity dashboard. From the capacity dashboard, users can go to screens that display information for Effective and Provisioned capacity, Snapshot capacity, and Data Reduction, as well as system resources.

## Capacity dashboard

Within Unisphere for PowerMax, there are multiple displays that provide information related to capacity usage.

The main dashboard displays an interactive graph that shows the effective capacity usage and data reduction over time. This display shows the history of effective capacity usage and how the data reduction ratio relates to effective capacity.  This information can be used to monitor and track trends of effective capacity usage relative to the data reduction ratio being shown. PowerMax 2500 or 8500 systems can be configured with FBA and CKD emulation within the same storage resource pool, but the historical graph is specific to the emulation view selected.

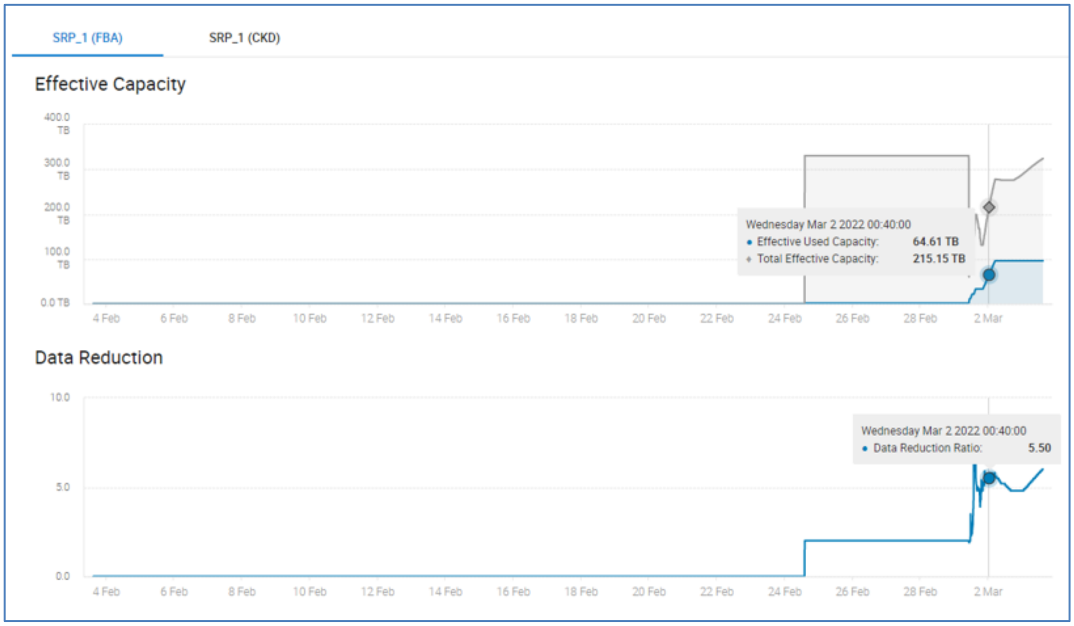


**Figure 2. Capacity Dashboard historical graph showing Effective Capacity and Data Reduction for FBA emulation**

The main dashboard also offers data in the form of bar graphs for provisioned capacity, effective capacity, snapshot usage, and data reduction. Each section can be expanded to a more detailed display showing more granular data for each item.

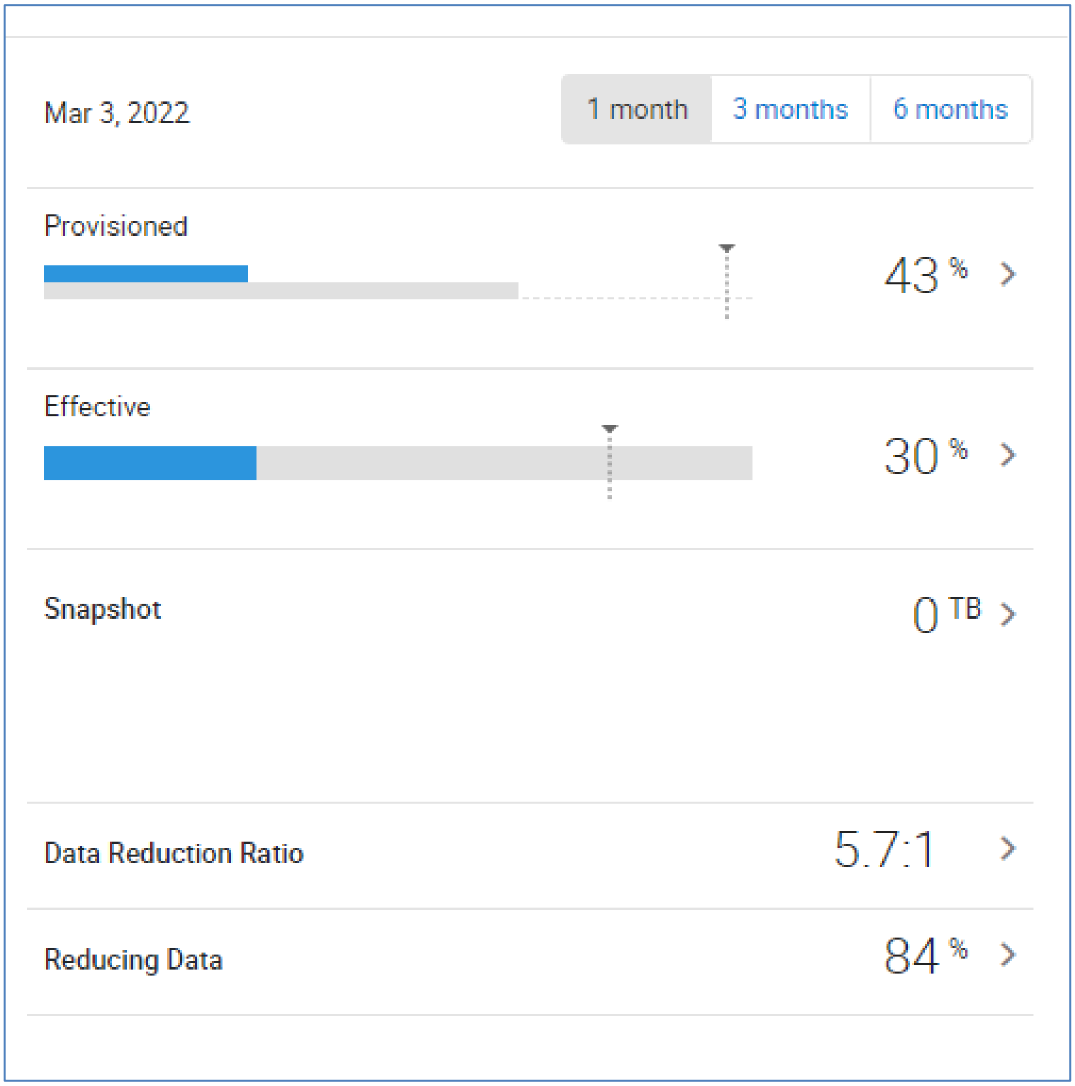


**Figure 3. Capacity dashboard bar graphs for provisioned capacity, effective capacity, snapshot usage, and data reduction**

## Provisioned

Provisioned capacity is the amount of capacity that is provisioned in the form of devices that is presented to hosts and application as available capacity. Provisioned capacity is tracked using two metrics, SRP Capacity and System Resources.

- **SRP Capacity** shows the amount of capacity provisioned as a value of TB provisioned and the available amount of effective capacity. The amount of effective SRP capacity is initially based on a default data reduction savings of 4:1 (3:1 for CKD emulation). This amount is calculated using the amount of physical capacity configured in the system. As users create devices provisioned capacity increases. The percentage value shown is the subscription percent and is calculated using the amount provisioned and the amount of Effective capacity.

- **System Resources** represent the amount of memory available to support provisioned capacity in the form of metadata. The total value presented will only change if cache is added to the system. The amount used represents just that. The amount free is an indication of how much extra provisioned capacity the system can support. As users create devices, the amount used will increase. When there is a variation between the two values memory resources are being used to support data reduction or other features that use memory such as taking snapshots of existing devices.

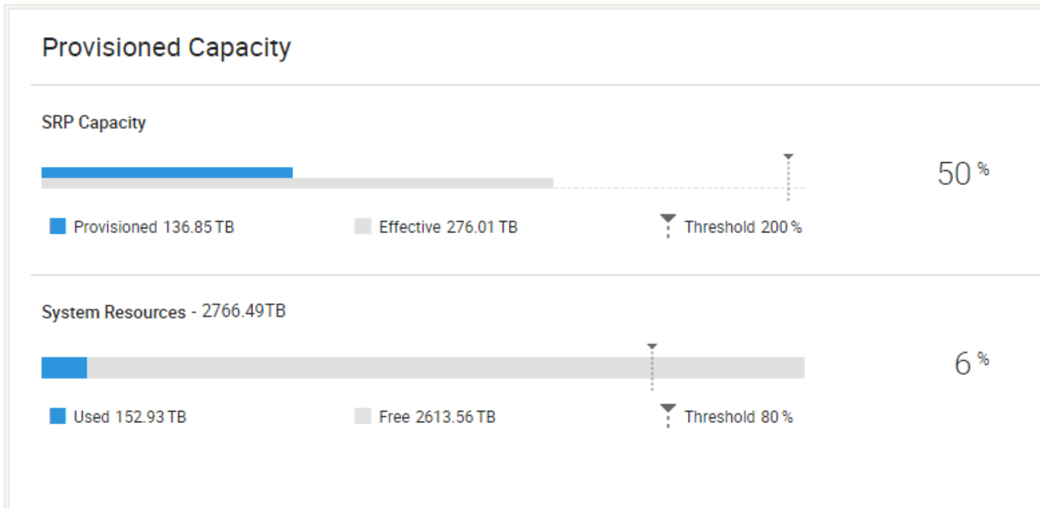


**Figure 4. System resource usage within the Provisioned capacity display**

## Effective

Effective capacity represents the amount of capacity available to the user based on an expectation of savings from the use of Data Reduction. The effective capacity display provides a detailed view of the physical and effective resources available. This is shown in three sections: Physical capacity, Effective Capacity Resources, and Effective Capacity Usage.

- **Physical Capacity** shows the amount of physical capacity available from the hard drives that are configured in the system. The amounts shown are the values after the formatting and RAID protection are applied. The value shown is the amount of capacity that the system can support for host data when data reduction is not being used.
- **Effective Capacity Resources** indicate the achievable values based on the current system resource usage. The effective capacity resource value shown will adjust relative to the current data reduction savings, and the physical and effective capacity usage.
- **Effective Capacity Usage** displays the current amount of effective capacity that is available based on the system resource usage and the current data reduction savings. The value displayed within the circular chart is the current available effective capacity. The values presented to the right breakdown the usage into three categories, snapshot usage, User Used and Free.

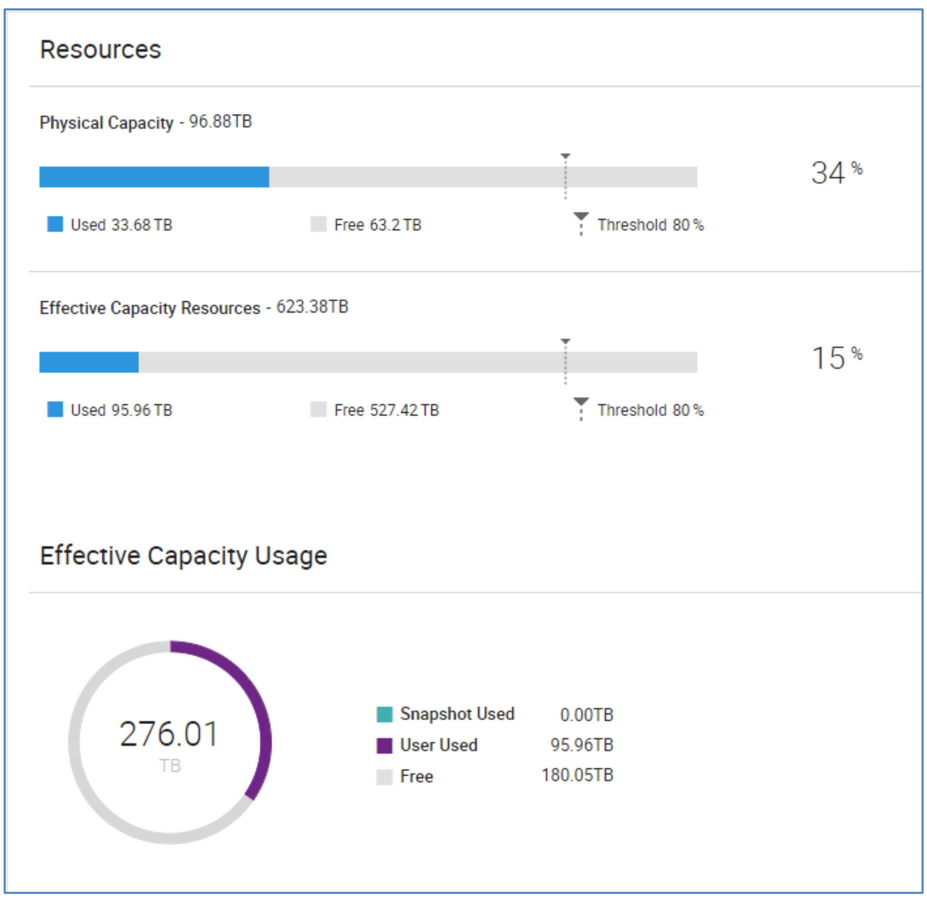


**Figure 5. Resource usage showing the Effective Capacity display**

## Snapshot

Back-end snapshot capacity may be significantly less than deltas per snapshot due to efficiency of features such as shared allocations and data reduction.

Hover over the Snapshot bar graph on the Capacity Dashboard for high-level details. The snapshot values are defined as:

- **Used**: Used Effective Capacity for snapshot change data
- **Free**: Remaining amount of snapshot change data based on amount used and remaining metadata
- **Total**: Used + Free
- **Threshold**: Threshold for alerting on snapshot change data

Click the bar graphs to go to the Effective Capacity dashboard and Snapshot Capacity dashboard.

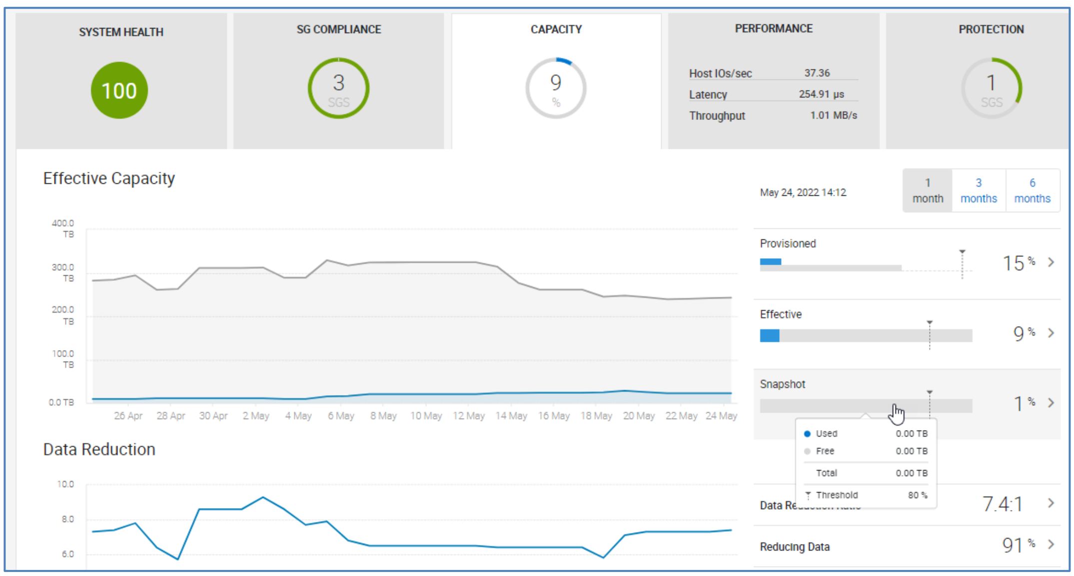


**Figure 6. Capacity dashboard**

On the Effective Capacity dashboard:

- **Snapshot Used**: Amount of effective capacity used by Snapshot Change data

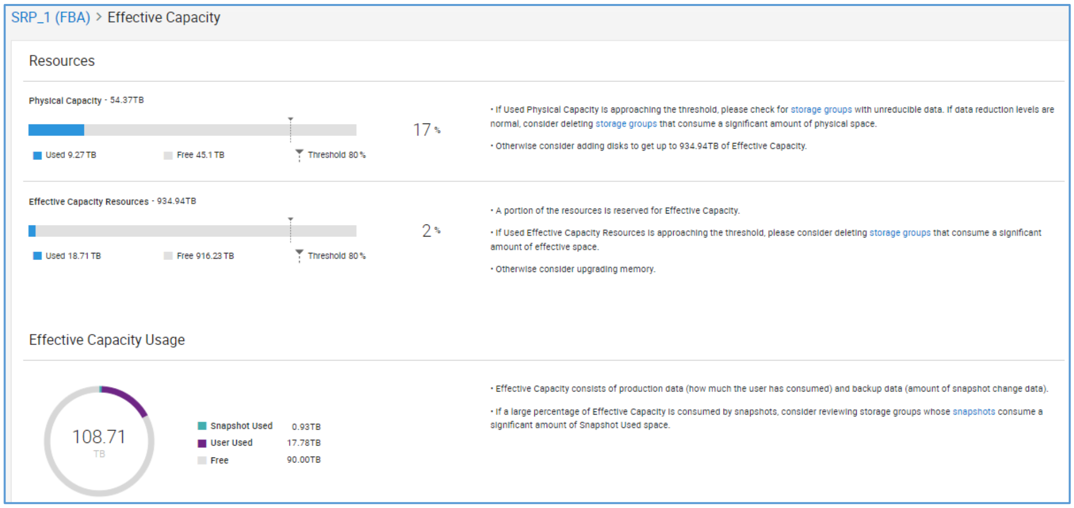


**Figure 7. Effective Capacity dashboard**

The values in the Snapshot Capacity page are defined as:

- **Snapshot Effective Used**: Percentage of Effective Capacity used by snapshot data for the SRP
- **Snapshot Physical Used**: Percentage of Usable Capacity consumed by snapshot data for the SRP
- **Snapshot Resources**: Percentage of snapshot metadata consumed for the entire system
- **Snapshot Resources Used**: Capacity used by snapshots as part of Snapshot Metadata Capacity

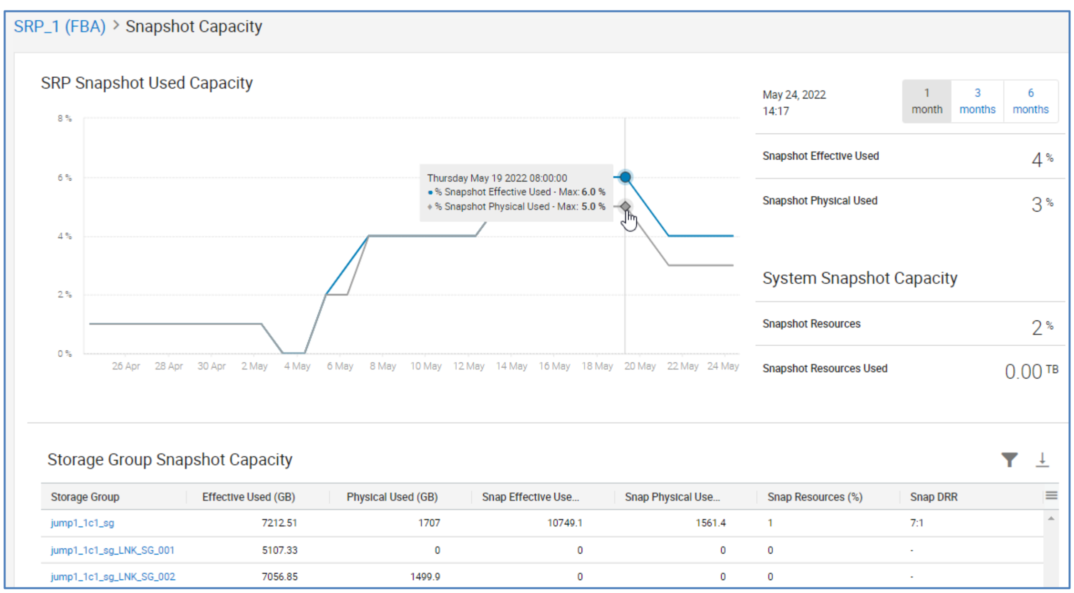


**Figure 8. Snapshot Capacity usage**

## Data Reduction Ratio

The data reduction display provides a single location for users to view the data reduction efficiency. There are three sections: Data Reduction Ratio, a historical interactive graph, and a table of all storage groups. The data reduction ratio shown accounts for only enabled and reducing data written to the system.

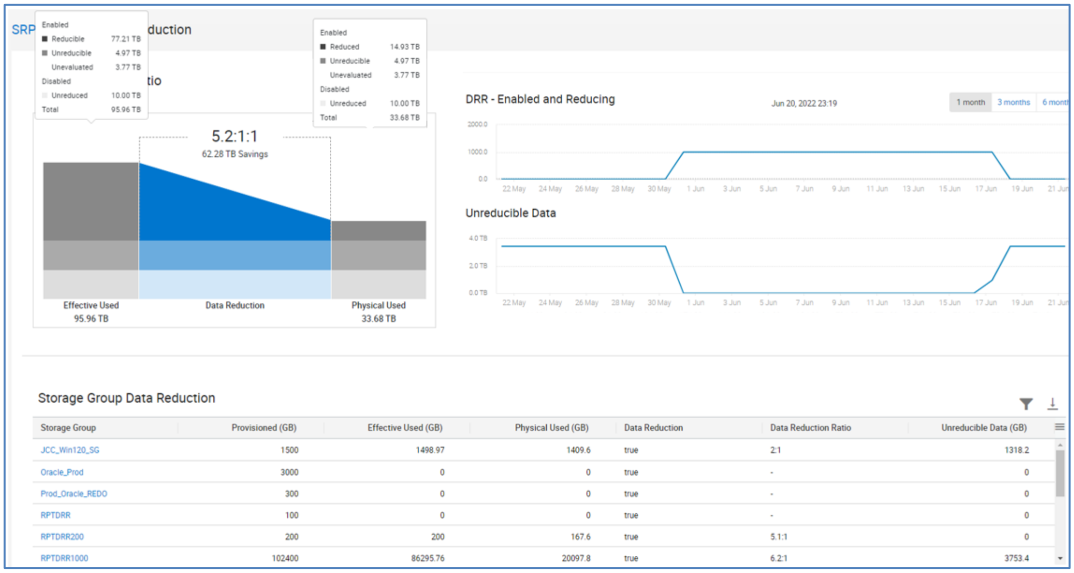


**Figure 9. Data Reduction display**

**Data Reduction Ratio** is displayed as a graph that presents effective used capacity the data reduction ratio and physical used capacity. Physical Used refers to the actual amount of physical capacity that is being used. Data reduction presents the savings as a ratio.

**Effective Used** represents data written to the system before any savings are achieved when data reduction is applied. All values shown represent the full size as it was written by the host or application. There are two categories that written data is placed into: Enabled and Disabled.

- **Enabled** indicates that the data being accounted for is data reduction enabled and subject to the data reduction process and the activity-based reduction function. There are three additional categories data can fall under when data reduction is enabled, Reducible, Unreducible, and Unevaluated.
  - **Reducible** data is the amount of data written that the data reduction process has identified as data that can be reduced to use less physical capacity than was written to the system.
  - **Unreducible** data is data that cannot be reduced.
  - **Unevaluated** data is not yet evaluated by the data reduction process. It has not been determined if the data is reducible or unreducible yet.
- **Disabled** indicates that the data written to the system is not subject to any data reduction savings. All data identified as disabled will be shown as unreduced.

**Physical Used** represents data written to the system after it has been stored on disk. This accounts for all data enabled and disabled as well as all data reduced and unreduced. There are two categories that represent the data stored on disk: Enabled, and Disabled.

- **Enabled** Indicates data that has gone through the data reduction process. There are three sub-categories of this data: Reduced, Unreducible, and Unevaluated.
  - **Reduced** data has been sent through the data reduction process. This process includes both passing through the data reduction hardware and being stored on disk. Reduced data stored on disk consumes less disk space than what was written by the host or application.
  - **Unreducible** indicates that the data has been sent through the data reduction process including the data reduction hardware but could not be reduced. Some unreducible data accounted for within the physical used section may be contributing to the data reduction savings as data that is shared due to deduplication.
  - **Unevaluated** data is data not yet evaluated by the data reduction process. It therefore has not been determined if the data is reducible or unreducible yet.
- **Disabled** indicates that the data written to the system is not subject to any data reduction savings. All data identified as disabled will be shown as unreduced.

The Interactive Graph charting **DRR Enabled and Reducing**, and **Unreducible Data** provides historical data. This shows the effect of unreducible data on the data reduction ratio. This graph allows the user to track and monitor the changes in the data reduction ratio that can be caused by unreducible data.

The storage group list provides capacity usage and data reduction information specific to each storage group in the system. When using the interactive graph to track changes to the data reduction ratio the storage group list can be used to identify storage groups that have large amounts of unreducible data that are impacting the data reduction ratio.

**Calculating Efficiency Ratios**: The data required to calculate the data reduction ratio is available in a pop-up window in the Data Reduction graph.

- **Data Reduction Ratio**: The data reduction ratio is calculated using enabled and reducible from Effective used and Enabled and Reduced from Physical used.

  *Enabled reducible ÷ Enabled Reduced*

- **Overall Data Reduction Ratio**: The overall system data reduction ratio is calculated using the total values from Effective Used and Physical Used

  *Effective Used total ÷ Physical Used total*

# Supported data services

## Overview

Data reduction is supported for FBA storage. Mixed FBA/CKD systems are supported within the same storage resource pool/s however data reduction for CKD emulation will only leverage compression and activity-based reduction. All other data services offered in both the PowerMax and VMAX All Flash systems are supported. These services include local replication (SnapVX), remote replication (SRDF), D@RE, and VMware vSphere Virtual Volumes (vVols).

## Local replication (SnapVX)

SnapVX snapshots protect applications without the use of target volumes to capture change data known as deltas. Snapshot deltas are automatically maintained at the storage back end using pointers to the relevant point-in-time images. Resource sharing and data deduplication automatically take advantage of this design to provide cache, capacity, and performance benefits.

Compressed source data remains compressed when becoming snapshot delta. Uncompressed source data may become compressed as it becomes snapshot delta or as it becomes less active. Read activity through linked targets may prevent an uncompressed delta from being compressed or may cause compressed snapshot deltas to be uncompressed. Snapshot deltas are available for deduplication.

Enabling data reduction on a linked target will only affect data owned by the linked target. Data on linked targets and clones are available for deduplication.

### Remote replication (SRDF)

Compression for SRDF is supported and known as SRDF compression. SRDF compression is a feature designed to reduce bandwidth consumption while sending data to and from connected systems using remote replication. SRDF compression and Data Reduction both use the same hardware; however, they serve different purposes. Data that has been compressed using data reduction is uncompressed before being sent across the SRDF link. If SRDF compression and inline compression apply, the data is uncompressed, compressed using the SRDF compression function, and sent to the remote site.

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

D@RE provides hardware-based, on-array, back-end encryption, Data Reduction is performed as an inline process. Data is passed through Data Reduction hardware before being sent through the encryption hardware. Therefore, data is compressed, deduped, or both before being encrypted. On a D@RE enabled system data encrypted on disk has already been compressed, deduped, or both.

### Virtual Volumes

Data reduction is supported for the allocation of data to vVols and follows the same I/O path as all other data. The IO path can be seen in Figure 1. Data Reduction is enabled at the storage resource level in a vVol storage container as there are no storage groups for vVols.

## Conclusion

### Summary

The use of physical storage capacity is a common concern of storage administrators across the storage industry. The constant and ever-growing amounts of data have created the need for more efficiency in the use of physical capacity. Dell PowerMax 2500 and 8500 data storage systems take this efficiency to the next level. Data Reduction provides exceptional capacity savings while delivering optimal performance. This result leads to a smaller data center footprint and an overall reduction in TCO. In addition to the savings, using data reduction is as simple as a single click to enable or disable. The system handles all the work.

# References

## Dell Technologies documentation

The following Dell Technologies documentation provides other information related to this document. Access to these documents depends on your login credentials. If you do not have access to a document, contact your Dell Technologies representative.

- PowerMax and VMAX Info Hub