

Dell PowerMax: Datenreduzierung

Inline-Komprimierung und -Deduplizierung

Juli 2022

H19254

Whitepaper

Zusammenfassung

Die PowerMax-Storage-Plattformen verfügen über mehrere Datenreduzierungstechniken wie Inline-Komprimierung und -Deduplizierung. Mustererkennung und effiziente Datenplatzierung sorgen zudem für ein ausgewogenes Verhältnis zwischen Performance und Effizienz.

Dell Technologies

Copyright

Die Informationen in dieser Veröffentlichung werden ohne Gewähr zur Verfügung gestellt. Dell Inc. macht keine Zusicherungen und übernimmt keine Gewährleistung jedweder Art im Hinblick auf die in diesem Dokument enthaltenen Informationen und schließt insbesondere jedwede implizite Gewährleistung für die Handelsüblichkeit und die Eignung für einen bestimmten Zweck aus.

Für das Nutzen, Kopieren und Verbreiten der in dieser Veröffentlichung beschriebenen Software ist eine entsprechende Softwarelizenz erforderlich.

Copyright © 2022 Dell Inc. oder deren Tochtergesellschaften. Alle Rechte vorbehalten. Dell Technologies, Dell, EMC, Dell EMC und andere Marken sind Marken von Dell Inc. oder deren Tochtergesellschaften. Intel, das Intel Logo, das Intel Inside Logo und Xeon sind Marken der Intel Corporation in den USA und/oder anderen Ländern. Alle anderen Marken können Marken ihrer jeweiligen Inhaber sein.

Veröffentlicht in den USA, Juli 2022 H19254.

Dell Inc. ist der Ansicht, dass die Informationen in diesem Dokument zum Zeitpunkt der Veröffentlichung korrekt sind. Diese Informationen können jederzeit ohne vorherige Ankündigung geändert werden.

Inhalt

Zusammenfassung	4
Datenreduzierung.....	5
Nutzung der Systemressourcen	9
Management und Monitoring	9
Unterstützte Datendienste.....	17
Fazit	18
Referenzen.....	19

Zusammenfassung

Übersicht

Die Datenreduzierung mit dem Dell EMC PowerMax-System steigert durch eine Kombination aus Inline-Komprimierung, Inline-Deduplizierung und Mustererkennung die Systemeffizienz. Mithilfe dieser Datenreduzierungstechniken können Nutzern enorme Kapazitätseinsparungen erzielen. Die Datenreduzierung komprimiert Daten und eliminiert redundante Datenkopien. In diesem Whitepaper wird erläutert, wie die Datenreduzierung bei den PowerMax-Systemen funktioniert, und das Reporting mit Dell Managementanwendungen wie Unisphere for PowerMax, Solutions Enabler und Mainframe Enabler-Software beschrieben.

Versionen

Datum	Beschreibung
Juli 2022	Erstveröffentlichung

Wir wissen Ihr Feedback zu schätzen

Dell Technologies und die Autoren dieses Dokuments freuen sich über Ihr Feedback zu diesem Dokument. Wenden Sie sich per [E-Mail](#) an das Dell Technologies Team.

Autor: Robert Tasker

Hinweis: Links zu anderen Dokumenten zu diesem Thema finden Sie im [Info Hub für PowerMax und VMAX](#).

Datenreduzierung

Übersicht

Die Datenreduzierung kombiniert Inline-Komprimierung, Inline-Deduplizierung, Mustererkennung, effiziente Datenplatzierung und maschinelles Lernen (ML). Dadurch entsteht ein System, auf das Nutzer mehr Hostdaten als die verfügbare physische Gesamtkapazität schreiben können, während sie gleichzeitig die von einem Enterprise Storage-System erwartete Performance erreichen. Diese Funktion ist standardmäßig aktiviert und kann für jede Storage-Gruppe einzeln aktiviert oder deaktiviert werden. Außerdem werden alle bei den PowerMax 2500- und 8500-Systemen verfügbaren Datendienste unterstützt. Diese Unterstützung gilt auch für die CKD-Emulation, umfasst aber keine Deduplizierung für CKD.

Durch die Komprimierung wird die Größe der Daten reduziert, während bei der Deduplizierung die Daten als eine einzelne Instanz gespeichert werden. Die Mustererkennung umfasst eine Zuweisungsfunktion für Werte ungleich Null, die ausschließt, dass Zeichenfolgen aufeinanderfolgender Nullen als Teil der komprimierten Daten gespeichert werden. Komprimierung, Deduplizierung und Mustererkennung werden mithilfe der im System integrierten Hardwareunterstützung durchgeführt, um den Overhead bei der Ausführung dieser Funktionen zu reduzieren. Maschinelles Lernen identifiziert die meist genutzten gespeicherten Daten auf der Festplatte und stellt im Hinblick auf eine optimale Performance sicher, dass diese nicht komprimiert werden. Die effiziente Datenplatzierung verwendet eine **Komprimierungsfunktion**, die Daten strategisch speichert, um nicht unnötig Speicherplatz zu belegen und den Bedarf an Funktionen zur automatischen Speicherbereinigung oder Defragmentierung zu reduzieren.

Aktivitätsbasierte Reduzierung

Die aktivitätsbasierte Reduzierung (Activity Based Reduction, ABR) reduziert die Performancekosten, die durch die Dekomprimierung häufig verwendeter Daten entstehen. Mit dieser Funktion können bis zu 20 % der meist genutzten Daten auf dem System nicht komprimiert gespeichert werden. Diese Möglichkeit kommt dem System zugute, da es die Performancelatenz minimiert, die sich aus der ständigen Dekomprimierung häufig abgerufener Daten ergibt. Zur Ermittlung der Daten mit dem häufigsten Zugriff verwendet das System ML-Algorithmen, die IO-Statistiken verarbeiten. Auf diese Weise wird eine optimale Umgebung mit einem ausgewogenen Verhältnis von Einsparungen durch Datenreduzierung und Performance aufrechterhalten.

Komprimierung

Komprimierung reduziert eingehende Schreib-Workloads auf die kleinstmögliche Größe, um die geringste Kapazität zu verbrauchen. Die Daten werden komprimiert, wenn sie die Datenreduzierungshardware, die den GZIP-Komprimierungsalgorithmus verwendet, durchlaufen. Die Datenreduzierungshardware unterteilt die Daten in vier Abschnitte, die zur Maximierung der Effizienz parallel komprimiert werden. Die Summe der vier Abschnitte entspricht der endgültigen reduzierten Größe der Daten, die auf der Festplatte gespeichert sind. Diese Funktion ermöglicht einen differenzierten Zugriff auf reduzierte Daten. Es werden nur die Abschnitte verarbeitet, die die angeforderten Daten für teilweise Lese- oder Schreibanforderungen enthalten, da jeder Abschnitt getrennt behandelt werden kann.

Deduplizierung

Bei der Deduplizierung handelt es sich um eine Methode zur Kapazitätseinsparung, die identische Kopien von Daten identifiziert und als eine einzelne Instanz speichert. Es gibt einige Elemente der Deduplizierung, die für die Bereitstellung effizienter Kapazitätseinsparungen erforderlich sind.

- **Hashfunktions-ID:** Die Hashfunktions-ID ist eine eindeutige Kennung für eingehende Daten, die verwendet wird, um zu bestimmen, ob eine Deduplizierungsbeziehung erforderlich ist. Das System verwendet einen SHA-256-Algorithmus, um die Hashfunktions-ID zu generieren.
- **Hash-ID-Tabelle:** Bei einer Hashtabelle handelt es sich um eine Zuweisung des Systemspeichers, die zwischen den System Directors verteilt ist. Diese Tabellen katalogisieren die vom Deduplizierungsprozess verwendeten Hash-IDs. Anhand der Einträge in der Tabelle wird festgestellt, ob eine Deduplizierungsbeziehung vorhanden ist oder ob ein neuer Eintrag erforderlich ist und die Daten auf der Festplatte gespeichert werden können.
- **Managementobjekt für die Deduplizierung (DMO):** Das DMO ist ein 64-Byte-Objekt im Systemspeicher, das nur vorhanden ist, wenn eine Deduplizierungsbeziehung besteht. Diese Objekte speichern und managen die Zeiger zwischen

Frontend-Geräten und deduplizierten Daten, die die Backend-Kapazität im Array verbrauchen. DMOs managen die Zeiger für deduplizierte Daten zwischen Frontendgeräten und den auf der Festplatte gespeicherten Daten. Dadurch wird auch gemanagt, in welcher Hashfunktionstabelle die Hashfunktions-IDs gespeichert werden, wenn Deduplizierungsbeziehungen vorhanden sind.

Die Deduplizierung wird mit derselben Datenreduzierungshardware wie die Komprimierung durchgeführt und es wird eine eindeutige Hash-ID erzeugt, wenn die Daten von der Hardware verarbeitet werden. Anschließend wird geprüft, ob die Hash-ID-Tabelle diese Hash-ID bereits enthält. Wird eine Übereinstimmung gefunden, werden die Daten nicht auf der Festplatte gespeichert und es erfolgt eine Deduplizierungsfreigabe. Zwischen dem Frontend-Volume und der eindeutigen ID in der Hash-ID-Tabelle werden Zeiger festgelegt. Die Zeiger verknüpfen die einzelne Instanz der auf der Festplatte gespeicherten Daten mit dem Volume und ermöglichen so den zukünftigen Zugriff auf die Daten. Das DMO verwaltet die Zeiger zwischen den Daten, den Frontend-Volumes, die auf die Daten zugreifen, und der Hash-ID-Tabelle. Wenn keine Übereinstimmung in der Hash-ID-Tabelle vorhanden ist, wird ein neuer Eintrag für zukünftige Hash-ID-Vergleiche hinzugefügt.

Deduplizierungsalgorithmus

PowerMax-Systeme verwenden den SHA-256-Hashfunktionsalgorithmus, der in der Hardware zur Datenreduzierung implementiert ist, um doppelte Daten zu finden. Die Daten werden dann als einzelne Instanz gespeichert, damit sie von mehreren Quellen gemeinsam genutzt werden können. Dieser Prozess verbessert die Dateneffizienz bei gleichzeitiger Aufrechterhaltung der Datenintegrität.

Der SHA-256-Algorithmus generiert einen 32-Byte-Code für jeden 32-KB-Datenblock. Denken Sie dabei an ein System mit 1 PB geschriebener Daten, von denen pro Tag 5 % aktualisiert werden. In einer Million Betriebsjahren liegt die Wahrscheinlichkeit einer Hashfunktionskollision bei 20 %. Da jeder 128-KB-Track als vier Blöcke mit je 32 KB behandelt wird, müsste für eine tatsächliche Hashfunktionskollision ein Hashfunktionskonflikt auf allen vier Blöcken auf demselben 128 KB-Track auftreten. Die Wahrscheinlichkeit, dass alle vier Blöcke kollidieren, ist rein theoretisch (und liegt bei einer Wahrscheinlichkeit von 1 % im Laufe von einer Billion Betriebsjahren). Wenn in der Vergleichsphase der Deduplizierung eine Übereinstimmung gefunden wird, wird außerdem ein Byte-für-Byte-Vergleich durchgeführt. Mit diesem Vergleich wird noch einmal bestätigt, dass eine Übereinstimmung vorliegt, bevor die Tabellen aktualisiert und die Zeiger festgelegt werden, um den Zugriff auf die Daten zu ermöglichen.

Komprimierung

Die Datenplatzierung erfolgt mithilfe eines Komprimierungsprozesses. Bei dieser Komprimierung werden reduzierte oder nicht reduzierte Daten intuitiv am bestmöglichen verfügbaren Ort auf der Festplatte platziert. Beim Speichern von Daten auf der Festplatte werden Schreibobjekte verwendet. Jedes Objekt hat 6 MB zusammenhängende Backend-Datengerätekapazität für die im System konfigurierten Laufwerke. Schreibobjekte sind auf 1-K-Grenzen ausgerichtet und werden nacheinander in einer einzigen Verwendung verbraucht. Schreibobjekte werden für alle unterstützten RAID-Typen über vollständige Stripes verteilt, um die Schreibvorgänge zu optimieren. Jedes Objekt unterstützt reduzierte oder nicht reduzierte Daten für FBA- und CKD-Emulation.

- **FBA-Schreibobjekt:** Ein nicht reduziertes Schreibobjekt besteht aus 48 FBA-Tracks. Ein reduziertes Schreibobjekt besteht aus 1.000 reduzierten Tracks. Die Größe reduzierter Einträge für Schreibobjekte reicht von 1 KB bis 96 KB.
- **CKD-Schreibobjekt:** Ein nicht reduziertes Schreibobjekt besteht aus 108 CKD-Tracks. Ein reduziertes Schreibobjekt besteht aus 1.000 reduzierten Tracks. Die Größe reduzierter Einträge für Schreibobjekte reicht von 1 KB bis 52 KB.

Extended Data Compression

PowerMax 2500- und 8500-Systeme verfügen über eine zusätzliche Funktion, die so genannte Extended Data Compression (EDC), die bereits komprimierte Daten komprimiert, um weitere Kapazitätseinsparungen zu erzielen. Dies wird durch die Identifizierung von Daten erreicht, auf die längere Zeit nicht zugegriffen wurde. Die nachstehenden Faktoren qualifizieren die Daten als Kandidaten für EDC:

- Die Daten gehören zu einer Storage-Gruppe, für die die Datenreduzierung aktiviert wurde.
- Auf die Daten wurde seit 30 Tagen nicht zugegriffen.
- Die Daten wurden noch nicht von EDC komprimiert.

Daten, die die Kriterien für EDC erfüllen, werden mithilfe des Algorithmus Def9_128_SW komprimiert, um die zum Speichern der Daten beanspruchte Kapazität weiter zu reduzieren. Hierbei handelt es sich um einen automatisierten Hintergrundprozess innerhalb des Systems. Die zusätzlichen Einsparungen spiegeln sich im erreichten Komprimierungsverhältnis auf Ebene der Storage-Gruppe wider. EDC ist nur mit PowerMax-Storage-Arrays verfügbar.

CKD Komprimierung

Die aktivitätsbasierte Reduzierung (Activity Based Reduction, ABR) reduziert die Performancekosten, die durch die Dekomprimierung häufig verwendeter Daten entstehen. Mit dieser Funktion können bis zu 20 % der meist genutzten Daten auf dem System nicht komprimiert gespeichert werden. Dies kommt dem System zugute, da dadurch die negative Auswirkung auf die Performance minimiert wird, die sich aus der ständigen Dekomprimierung häufig abgerufener Daten ergibt. Zur Ermittlung der Zugriffshäufigkeit auf die Daten verwendet das System ML-Algorithmen, die von eingehenden I/O-Vorgängen an die Frontend-Geräte erfasste Statistiken verarbeiten. Dies gewährleistet eine ausgewogene Nutzung der Systemressourcen und ermöglicht so eine Umgebung mit einem optimalen Verhältnis von Einsparungen durch Datenreduzierung und Performance.

Komprimierung reduziert eingehende Schreib-Workloads auf die kleinstmögliche Größe, um die geringstmögliche Kapazität zu verbrauchen. Die Daten werden komprimiert, wenn sie die im System integrierte Datenreduzierungshardware, die den GZIP-Komprimierungsalgorithmus verwendet, durchlaufen. Die Datenreduzierungshardware unterteilt die Daten in vier Abschnitte, die zur Maximierung der Effizienz der Hardware parallel komprimiert werden. Die Summe der vier Abschnitte entspricht der endgültigen reduzierten Größe der Daten, die auf der Festplatte gespeichert sind. Dadurch ist ein differenzierter Zugriff auf die reduzierten Daten möglich, da diese mit einer teilweisen Lese- oder Schreibanforderung abgerufen werden. Es werden nur die Abschnitte verarbeitet, die die angeforderten Daten enthalten, da jeder Abschnitt getrennt behandelt werden kann.

Die Datenplatzierung erfolgt mithilfe eines Komprimierungsprozesses während der Datenplatzierung. Bei dieser Komprimierung werden reduzierte oder nicht reduzierte Daten intuitiv am bestmöglichen verfügbaren Ort auf der Festplatte platziert. Beim Speichern von Daten auf der Festplatte werden Schreibobjekte verwendet. Jedes Objekt hat 6 MB zusammenhängende Backend-Datengerätekapazität für die im System konfigurierten Laufwerke. Schreibobjekte sind auf 1-K-Grenzen ausgerichtet und werden nacheinander in einer einzigen Verwendung verbraucht. Schreibobjekte werden für alle unterstützten RAID-Typen über vollständige Stripes verteilt, um die Schreibvorgänge zu optimieren. Jedes Objekt unterstützt reduzierte oder nicht reduzierte Daten. Ein nicht reduziertes Schreibobjekt besteht aus 108 CKD-Tracks. Ein reduziertes Schreibobjekt besteht aus 1.000 reduzierten Tracks. Die Größe reduzierter Einträge für Schreibobjekte reicht von 1 KB bis 52 KB.

I/O-Vorgang bei der Datenreduzierung

Alle I/O-Vorgänge werden über den Cache geleitet und anschließend vom System verarbeitet. Die Aktionen zur Datenreduzierung werden durchgeführt, nachdem die Daten vom System empfangen wurden, jedoch bevor sie auf der Festplatte gespeichert werden. Die Verwendung eines Inline-Prozesses erfordert zusätzliche Prüfungen innerhalb des I/O-Vorgangs, bei dem eine Datenreduzierung angewendet wird. Das System verwendet diese Prüfungen, um zu ermitteln, ob eingehende Daten die Hardware zur Datenreduzierung passieren müssen oder nicht. Eingehende Daten für eine Storage-Gruppe mit aktivierter Datenreduzierung folgen dem Datenreduzierungsablauf. Die Funktion zur aktivitätsbasierten Reduzierung (ABR) sorgt jedoch dafür, dass aktive

Daten den Datenreduzierungsschritt bei einer Storage-Gruppe mit aktivierter Datenreduzierung zur Performance-Optimierung überspringen. Daten, die aufgrund von ABR nicht komprimiert wurden, können zu einem späteren Zeitpunkt komprimiert und in einen Komprimierungspool verschoben werden. Daten für eine Storage-Gruppe mit deaktivierter Datenreduzierung ignorieren den Datenreduzierungsablauf und werden unreduziert in das System geschrieben.

Es gibt drei verschiedene I/O-Typen, die zu berücksichtigen sind: Lesen, Schreiben und Schreiben/Aktualisieren.

- **Lesen:** Eine Anforderung für den Zugriff auf Daten, die das Array bereits belegen.
- **Schreiben:** Eingehender I/O-Vorgang, der Speicherplatz verbraucht.
- **Schreiben/Aktualisieren:** Eingehender I/O-Vorgang, der die Daten ändern kann, die dem Speicherplatz auf dem Array zugewiesen sind.

Die folgende Abbildung beschreibt den Pfad, dem der I/O-Vorgang folgt. Er wird bestimmt durch die Eigenschaften des Datenvolumens oder der zugehörigen Storage-Gruppe.

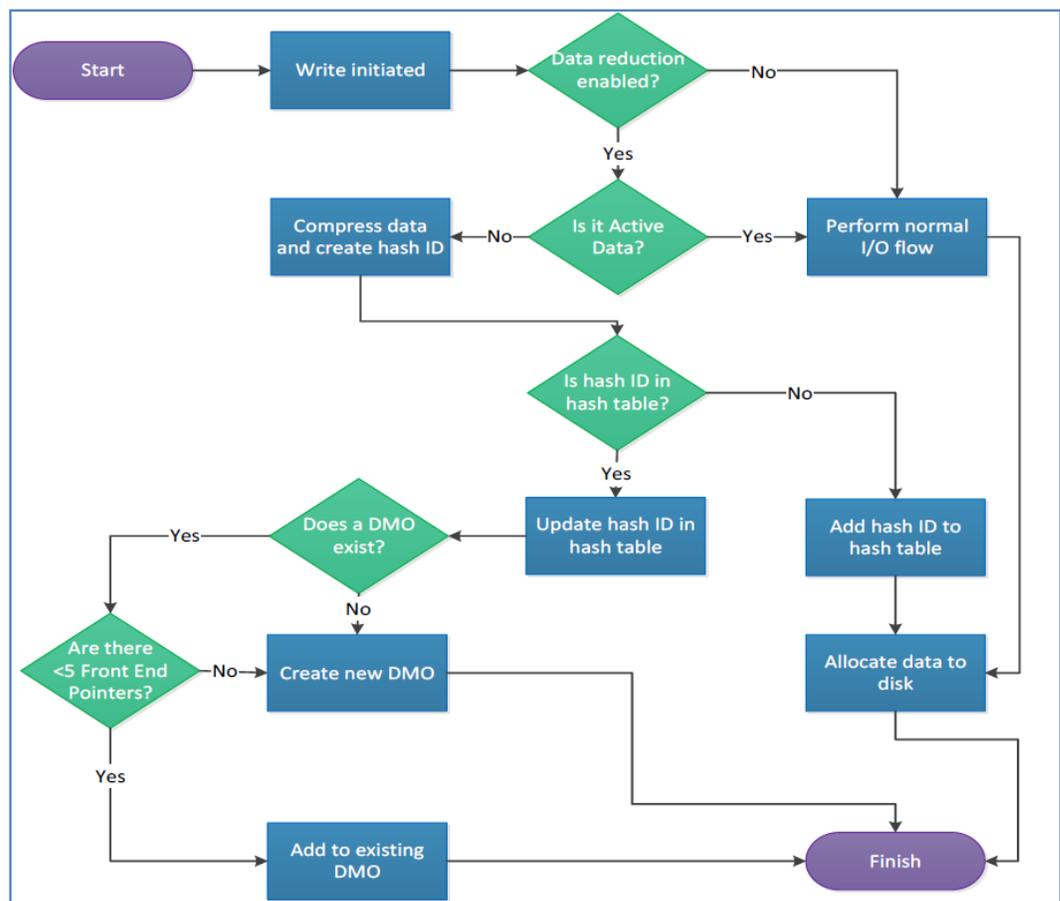


Abbildung 1. I/O-Vorgang für die Datenreduzierung bei PowerMax Enterprise Storage-Systemen

Nutzung der Systemressourcen

Übersicht

Kapazität und Arbeitsspeicher sind die beiden primären Ressourcen, die in jedem PowerMax-System konfiguriert sind. Die Kapazität wird in die Kategorien „physisch“, „effektiv“ und „bereitgestellt“ unterteilt. Der Arbeitsspeicher fällt unter die Systemressourcen.

Storage-Ressourcen unterstützen die Metadatenstrukturen für die bereitgestellte Kapazität sowie die physische Kapazität. Die verfügbare effektive Kapazität setzt sich zusammen aus der physischen Kapazität, den verfügbaren Systemressourcen und der Reduzierbarkeit der auf das System geschriebenen Daten. Geschriebene Daten, die hochgradig reduzierbar sind, verbrauchen weniger physische Kapazität, was zu einer höheren effektiven Kapazität führt. Umgekehrt gilt: Geschriebene Daten, die nicht reduzierbar sind, können die verfügbare effektive Kapazität verringern. Die im Abschnitt [Datenreduzierung](#) beschriebenen Angaben (Kapazität, Systemressourcen) sind in den Managementanwendungen verfügbar, die für PowerMax 2500- und 8500-Systeme, Unisphere for PowerMax, Solutions Enabler und Mainframe Enabler-Software verwendet werden. Unisphere for PowerMax ist eine Benutzeroberfläche (UI), die Daten in Grafiken, Diagrammen und Listenform bereitstellt. Solutions Enabler ist eine Standard-Befehlszeilenschnittstelle, die die gleichen Daten, jedoch nicht in Form von Grafiken und Diagrammen, bereitstellt. Mainframe Enablers sind eine Suite von Komponenten, die die Dell Storage-Systeme in einer Mainframeumgebung überwachen und verwalten. Die Abbildungen in den nächsten Abschnitten dieses Whitepapers zeigen Unisphere für PowerMax, das ein PowerMax 2500- oder 8500-System verwaltet.

Physische Kapazität

Die physische Kapazität ist die Größe des Festplattenspeicherplatzes, der basierend auf den installierten Festplatten und dem angewendeten RAID-Schutz im System konfiguriert wird. In einer Konfiguration ohne verwendete Datenreduzierung entspricht die physische Kapazität der Gesamtkapazität, die für Hostdaten verfügbar ist. Beispiel: Ein System mit 100 TB physischer Kapazität bedeutet, dass es 100 TB Hostdaten ohne Datenreduzierung aufnehmen kann.

Effektive Kapazität

Die effektive Kapazität ist die Menge des verfügbaren Speicherplatzes, wenn die Datenreduzierung verwendet wird. Die Gesamtkapazität bei der Erstinstallation hängt von der im System konfigurierten Speichermenge ab und basiert auf einem standardmäßigen Datenreduzierungsverhältnis von 4:1 (3:1 für CKD-Emulation). Beispielsweise weist dasselbe System mit 100 TB physischer Kapazität eine effektive Kapazität von 400 TB auf. Dieser Wert von 400 TB bildet einen Ausgangspunkt für die effektive Kapazität und ändert sich, wenn Daten auf das System geschrieben und die Datenreduzierung angewendet wird.

Bereitgestellte Kapazität

Die bereitgestellte Kapazität ist die Darstellung der verfügbaren Kapazität in Form von Geräten, die erstellt und Hosts und Anwendungen präsentiert werden, die physische oder effektive Kapazität im System verbrauchen möchten.

Management und Monitoring

Übersicht

Unisphere for PowerMax ist eine Benutzeroberfläche, die zur Verwaltung und Überwachung der Kapazität und Ressourcennutzung des Systems verwendet wird. Auf Systemebene werden Informationen zu Kapazitätsauslastung, Datenreduzierung und Systemressourcen im Kapazitäts-Dashboard angezeigt. Über das Kapazitäts-Dashboard können Nutzer verschiedene Seiten aufrufen, auf denen Angaben zur effektiven und bereitgestellten Kapazität, zur Snapshot-Kapazität und Datenreduzierung sowie zu den Systemressourcen angezeigt werden.

Kapazitäts-Dashboard

In Unisphere for PowerMax werden Informationen zur Kapazitätsauslastung in verschiedenen grafischen Darstellungen bereitgestellt.

Das Hauptdashboard umfasst ein interaktives Diagramm, das die effektive Kapazitätsauslastung und Datenreduzierung auf einer Zeitachse anzeigt. Das Diagramm stellt die effektive Kapazitätsauslastung im Zeitverlauf und die Beziehung zwischen Datenreduzierungsverhältnis und effektiver Kapazität dar. Anhand dieser Informationen lassen sich Trends der effektiven Kapazitätsauslastung in Korrelation zum angezeigten Datenreduzierungsverhältnis überwachen und nachverfolgen. PowerMax 2500- oder 8500-Systeme können mit FBA und CKD-Emulation innerhalb desselben Storage-Ressourcenpools konfiguriert werden, aber das Verlaufsdiagramm ist spezifisch für die ausgewählte Emulationsansicht.



Abbildung 2. Verlaufsdiagramm im Kapazitäts-Dashboard mit effektiver Kapazität und Datenreduzierung für FBA-Emulation

Das Hauptdashboard liefert außerdem Daten in Form von Balkendiagrammen für bereitgestellte Kapazität, effektive Kapazität, Snapshot-Nutzung und Datenreduzierung. Jeder Abschnitt kann auf eine Detailanzeige mit genaueren Daten für jedes Element erweitert werden.

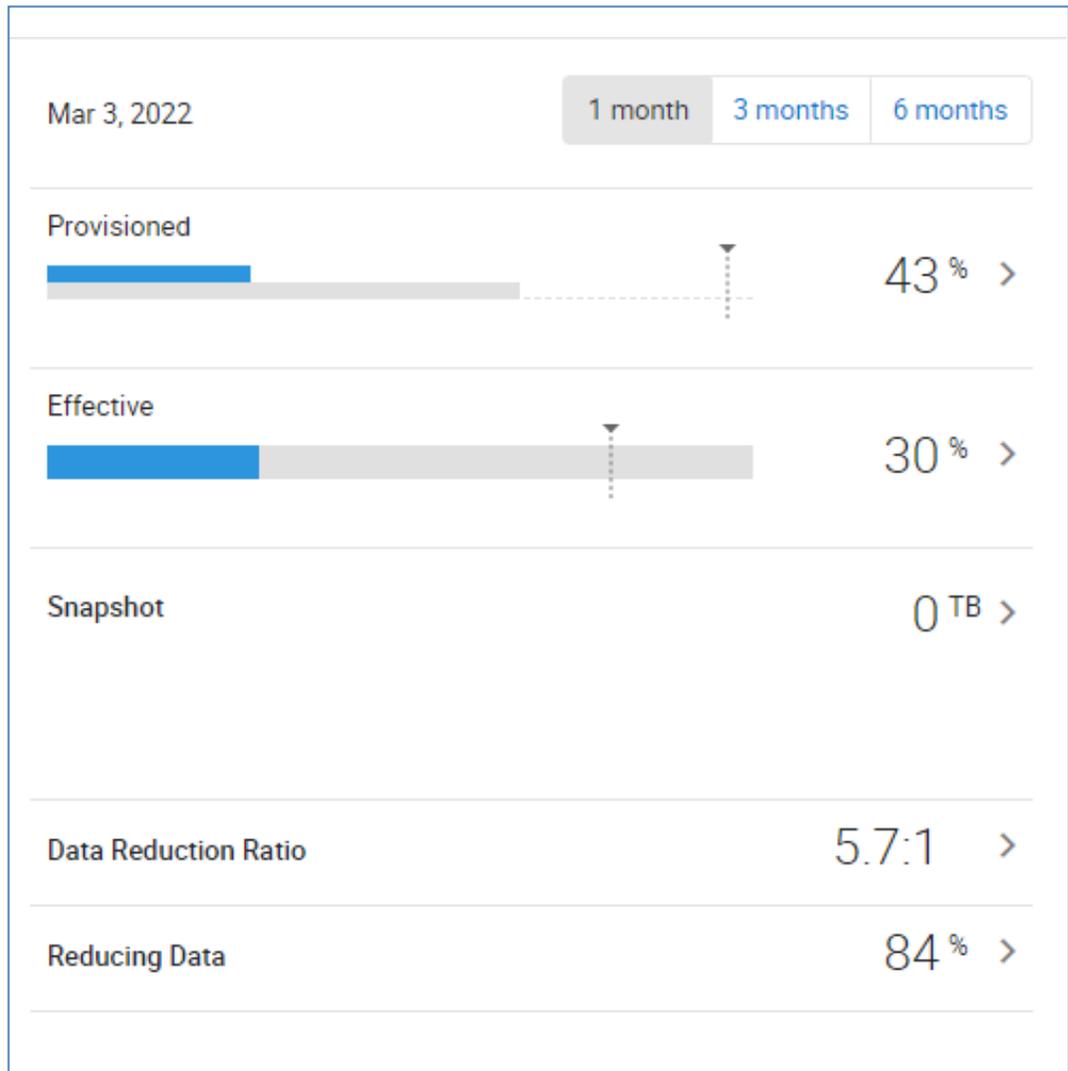


Abbildung 3. Balkendiagramme im Kapazitäts-Dashboards für bereitgestellte Kapazität, effektive Kapazität, Snapshot-Nutzung und Datenreduzierung

Provisioned

Die bereitgestellte Kapazität entspricht der in Form von Geräten bereitgestellte Kapazität, die Hosts und Anwendungen als verfügbare Kapazität zur Verfügung steht. Die bereitgestellte Kapazität wird mithilfe von zwei Metriken nachverfolgt: SRP-Kapazität und Systemressourcen.

- **SRP Capacity** zeigt die bereitgestellte Kapazität als absoluten Wert in TB und als Prozentsatz der effektiven Kapazität an. Die effektive SRP-Kapazität basiert anfänglich auf einem Standard-Datenreduzierungsverhältnis von 4:1 (3:1 für CKD-Emulation). Dieser Wert wird auf Basis der im System konfigurierten physischen Kapazität berechnet. Wenn Nutzer Geräte erstellen, erhöht sich die bereitgestellte Kapazität. Der angezeigte Prozentwert ist der Prozentsatz des Abonnements und wird anhand der bereitgestellten Kapazität und der effektiven Kapazität berechnet.

- **System Resources** entspricht dem verfügbaren Arbeitsspeicher zur Unterstützung der bereitgestellten Kapazität in Form von Metadaten. Der angezeigte Gesamtwert ändert sich nur, wenn dem System Cache hinzugefügt wird. Der Wert für die belegten Systemressourcen stellt genau das dar. Die freien Systemressourcen geben an, wie viel zusätzlich bereitgestellte Kapazität das System unterstützen kann. Wenn Nutzer Geräte erstellen, nimmt der Wert der belegten Systemressourcen zu. Bei Abweichungen zwischen den beiden Werten werden Storage-Ressourcen verwendet, um die Datenreduzierung oder andere Funktionen zu unterstützen, die Speicher verwenden, wie z. B. das Erstellen von Snapshots vorhandener Geräte.

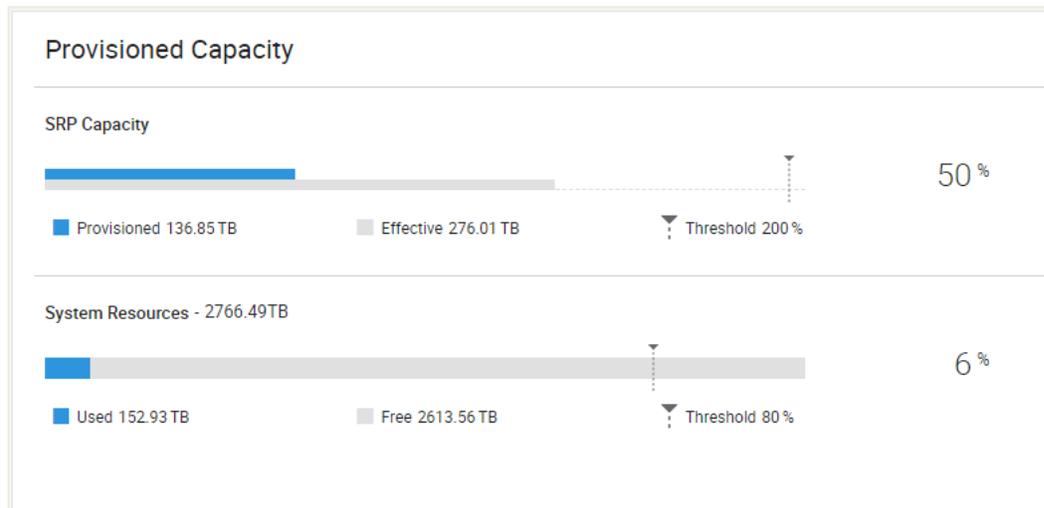


Abbildung 4. Systemressourcennutzung in der Anzeige „Provisioned Capacity“

Effective

Die effektive Kapazität entspricht der Kapazität, die dem Nutzer auf Basis der erwarteten Einsparungen durch die Verwendung der Datenreduzierung zur Verfügung steht. Die Anzeige der effektiven Kapazität bietet eine detaillierte Ansicht der verfügbaren physischen und effektiven Ressourcen. Dies wird in drei Abschnitten dargestellt: „Physical Capacity“, „Effective Capacity Resources“ und „Effective Capacity Usage“.

- **Physical Capacity** zeigt die physische Kapazität an, die von den im System konfigurierten Festplatten verfügbar ist. Die angezeigten Werte entsprechen der Kapazität nach Anwendung der Formatierung und des RAID-Schutzes. Der angezeigte Wert ist die Kapazität, die das System für Hostdaten unterstützen kann, wenn keine Datenreduzierung verwendet wird.
- **Effective Capacity Resources** gibt die erreichbaren Werte auf Basis der aktuellen Nutzung der Systemressourcen an. Der angezeigte Wert für die effektiven Kapazitätsressourcen passt sich je nach aktuellen Einsparungen durch Datenreduzierung und der physischen und effektiven Kapazitätsauslastung automatisch an.
- **Effective Capacity Usage** zeigt die aktuelle effektive Kapazität an, die auf Basis der Nutzung der Systemressourcen und der aktuellen Einsparungen durch Datenreduzierung verfügbar ist. Der im Kreisdiagramm angezeigte Wert ist die aktuell verfügbare effektive Kapazität. Die auf der rechten Seite dargestellten Werte unterteilen die Nutzung in drei Kategorien: von Snapshots belegt, von Nutzern belegt, und frei.

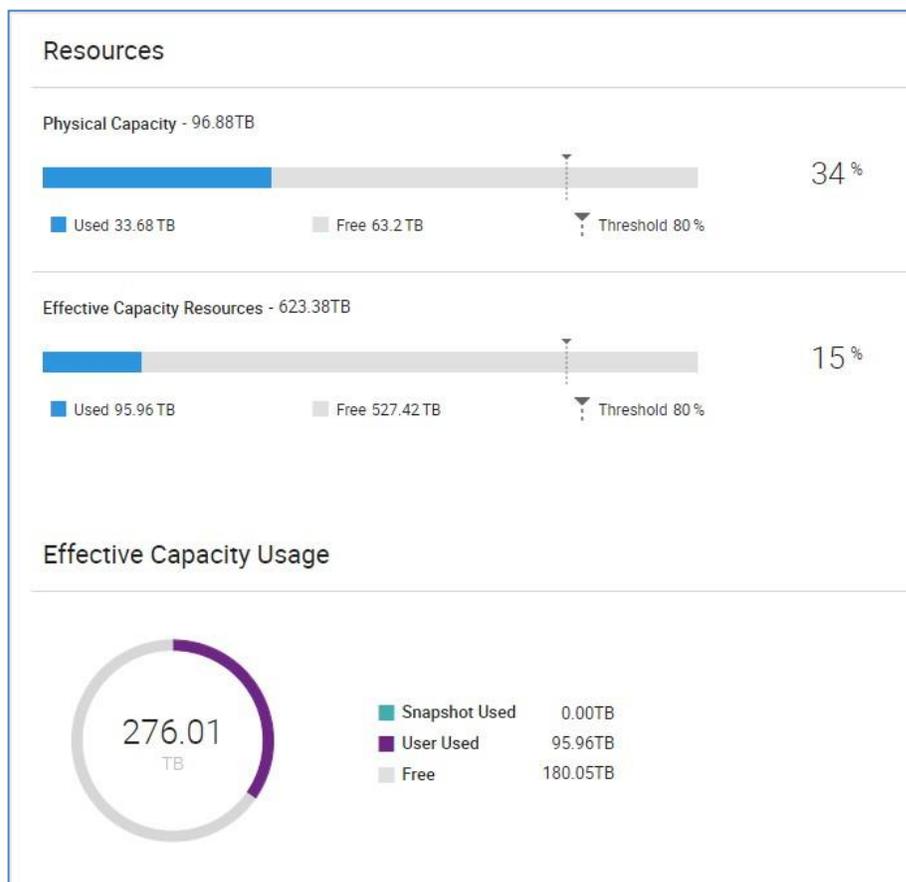


Abbildung 5. Ressourcennutzung in der Anzeige „Effective Capacity“

Snapshot

Die Backend-Snapshot-Kapazität kann aufgrund von Effizienzfunktionen wie gemeinsamen Zuweisungen und Datenreduzierung deutlich geringer als die logische Snapshot-Kapazität sein.

Bewegen Sie den Mauszeiger über das Snapshot-Balkendiagramm im Kapazitäts-Dashboard, um allgemeine Details anzuzeigen. Die Snapshot-Werte sind wie folgt definiert:

- **Used:** Genutzte effektive Kapazität für Snapshot-Änderungsdaten
- **Free:** Verbleibende Kapazität für Snapshot-Änderungsdaten basierend auf der genutzten Kapazität und den verbleibenden Metadaten
- **Total:** Summe aus genutzter und freier Kapazität („Used“ + „Free“)
- **Threshold:** Schwellenwert für Warnmeldungen zu Snapshot-Änderungsdaten

Klicken Sie auf die Balkendiagramme, um zum Dashboard „Effective Capacity“ und zum Dashboard „Snapshot Capacity“ zu gelangen.

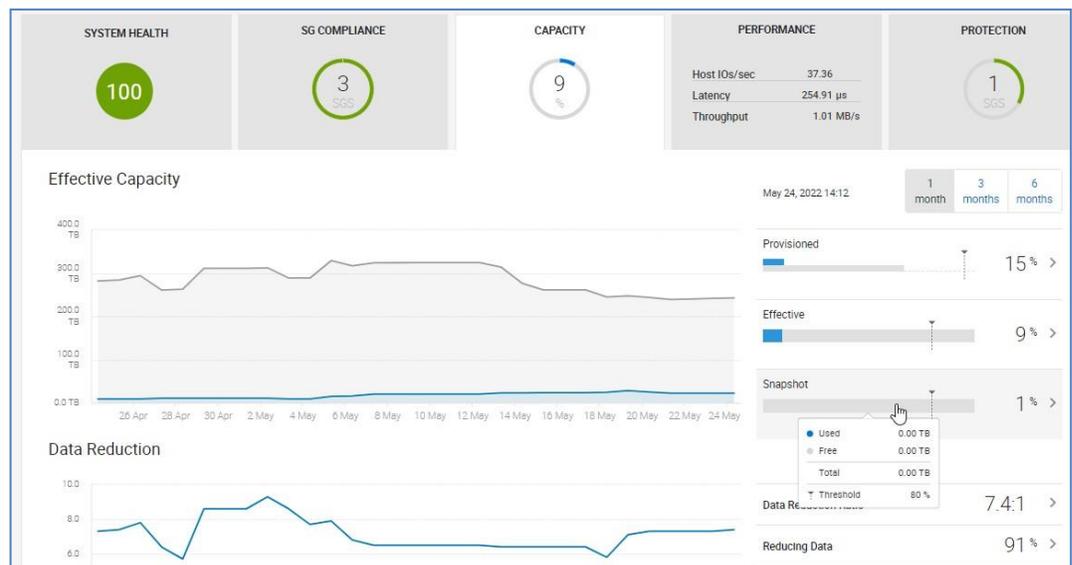


Abbildung 6: Kapazitäts-Dashboard

Dashboard „Effective Capacity“:

- **Snapshot Used:** Effektive Kapazität, die von Snapshot-Änderungsdaten genutzt wird

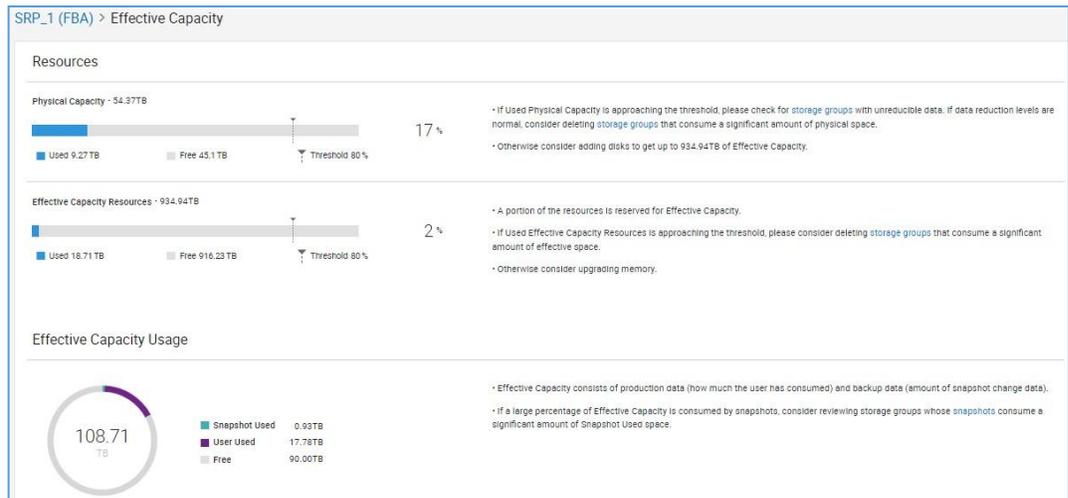


Abbildung 7. Dashboard „Effective Capacity“

Die Werte auf der Seite „Snapshot Capacity“ sind wie folgt definiert:

- **Snapshot Effective Used:** Prozentsatz der effektiven Kapazität, die von Snapshot-Daten für den SRP verwendet wird
- **Snapshot Physical Used:** Prozentsatz der nutzbaren Kapazität, die von Snapshot-Daten für den SRP verbraucht wird
- **Snapshot Resources:** Prozentsatz der Snapshot-Metadaten, die für das gesamte System verbraucht werden
- **Snapshot Resources Used:** Kapazität, die von Snapshots als Teil der Snapshot-Metadatenkapazität verwendet wird

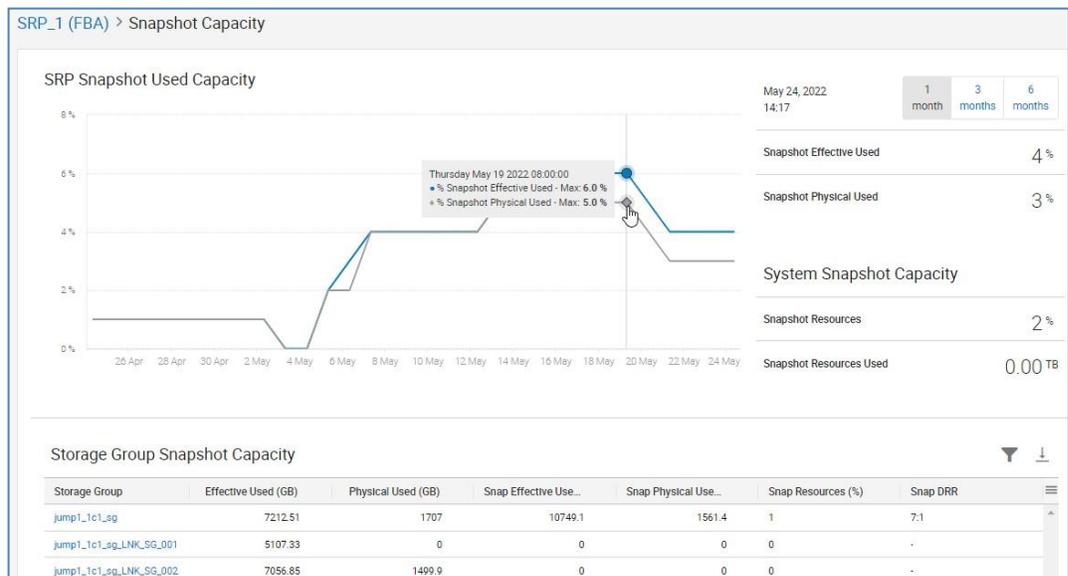


Abbildung 8. Kapazitätsauslastung durch Snapshots

Datenreduzierungsverhältnis

Die Datenreduzierungsanzeige bietet Nutzern eine zentrale Übersicht über die Effizienz der Datenreduzierung. Es gibt drei Abschnitte: Datenreduzierungsverhältnis, ein interaktives Verlaufsdiagramm und eine Tabelle aller Storage-Gruppen. Das angezeigte Datenreduzierungsverhältnis berücksichtigt nur aktivierte und reduzierte Daten, die auf das System geschrieben werden.

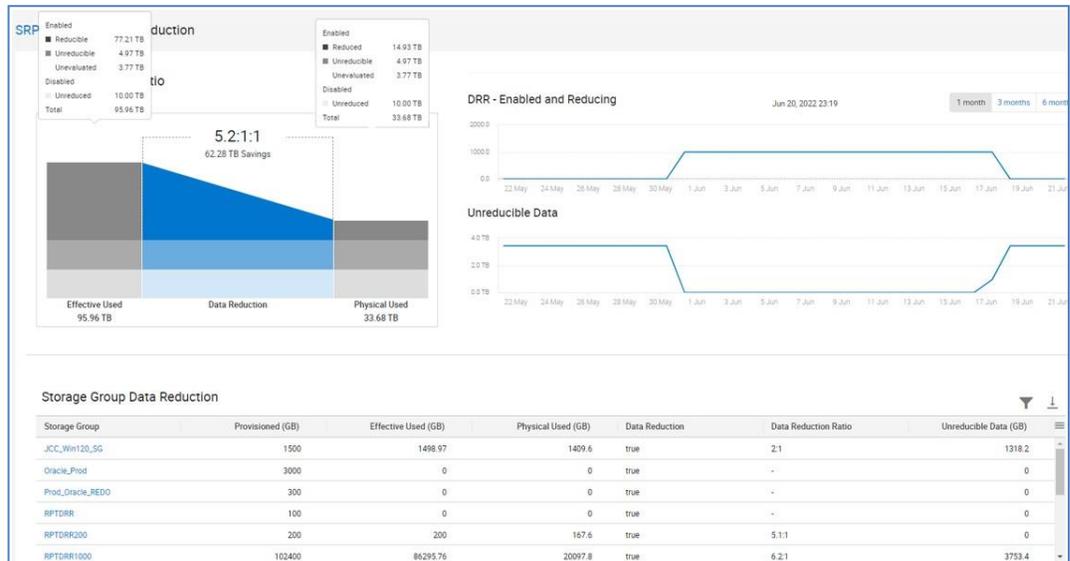


Abbildung 9. Anzeige „Data Reduction“

Das **Datenreduzierungsverhältnis** wird als Diagramm angezeigt, das die effektive genutzte Kapazität, das Datenreduzierungsverhältnis und die genutzte physische Kapazität darstellt. „Physical Used“ bezieht sich auf die tatsächlich belegte physische Kapazität. Die Datenreduzierung stellt die Einsparungen als Verhältnis dar.

Effective Used steht für Daten, die auf das System geschrieben werden, bevor bei verwendeter Datenreduzierung Einsparungen erzielt werden. Alle angezeigten Werte stellen die volle Größe dar, die vom Host oder der Anwendung geschrieben wurde. Es gibt zwei Kategorien, in die geschriebene Daten platziert werden: **Enabled** (Aktiviert) und **Disabled** (Deaktiviert).

- **Enabled** gibt an, dass für die berücksichtigten Daten die Datenreduzierung aktiviert ist und sie den Datenreduzierungsprozess und die aktivitätsbasierte Reduzierungsfunktion durchlaufen. Es gibt drei zusätzliche Kategorien, unter die Daten fallen können, wenn die Datenreduzierung aktiviert ist: „Reducible“, „Unreducible“ und „Unevaluated“.
 - **Reducible** entspricht der Menge der geschriebenen Daten, die der Datenreduzierungsprozess als reduzierbar erkannt hat, um weniger physische Kapazität zu belegen, als auf das System geschrieben wurde.
 - **Unreducible** sind Daten, die nicht reduziert werden können.
 - **Unevaluated** sind Daten, die vom Datenreduzierungsprozess noch nicht ausgewertet wurden. Es wurde noch nicht ermittelt, ob die Daten reduzierbar oder nicht reduzierbar sind.
- **Disabled** bedeutet, dass die auf das System geschriebenen Daten keinen Einsparungen durch Datenreduzierung unterliegen. Alle „Disabled“ gekennzeichneten Daten werden ohne Reduzierung angezeigt.

Physical Used steht für Daten, die auf das System geschrieben wurden, nachdem sie auf der Festplatte gespeichert wurden. Dies gilt für alle aktivierten und deaktivierten Daten sowie für alle reduzierten und nicht reduzierten Daten.

Die auf der Festplatte gespeicherten Daten fallen in zwei Kategorien: „Enabled“ und „Disabled“.

- **Enabled** zeigt Daten an, die den Datenreduzierungsprozess durchlaufen haben. Es gibt drei Unterkategorien dieser Daten: „Reduced“, „Unreducible“ und „Unevaluated“.
 - **Reduzierte** sind Daten, die den Datenreduzierungsprozess durchlaufen haben. Dieser Prozess umfasst das Durchlaufen der Datenreduzierungshardware und die Speicherung der Daten auf der Festplatte. Reduzierte Daten, die auf der Festplatte gespeichert sind, verbrauchen weniger Speicherplatz als das, was vom Host oder der Anwendung geschrieben wurde.
 - **Unreducible** bedeutet, dass die Daten dem Datenreduzierungsprozess einschließlich der Datenreduzierungshardware unterzogen wurden, aber nicht reduziert werden konnten. Einige nicht reduzierbare Daten, die im Abschnitt „Physical Used“ berücksichtigt werden, können zu den Einsparungen durch Datenreduzierung beitragen, da Daten aufgrund der Deduplizierung gemeinsam genutzt werden.
 - **Unevaluated** sind Daten, die vom Datenreduzierungsprozess noch nicht ausgewertet wurden. Es wurde daher noch nicht ermittelt, ob die Daten reduzierbar oder nicht reduzierbar sind.
- **Disabled** bedeutet, dass die auf das System geschriebenen Daten keinen Einsparungen durch Datenreduzierung unterliegen. Alle „Disabled“ gekennzeichneten Daten werden ohne Reduzierung angezeigt.

Das interaktive Diagramm **DRR Enabled and Reducing** und **Unreducible Data** enthält Verlaufsdaten. Es zeigt die Auswirkung nicht reduzierbarer Daten auf das Datenreduzierungsverhältnis. Dieses Diagramm ermöglicht es Nutzern, die durch nicht reduzierbare Daten verursachten Änderungen im Datenreduzierungsverhältnis nachzuerfolgen und zu überwachen.

Die Liste der Storage-Gruppen enthält Informationen zur Kapazitätsauslastung und Datenreduzierung, die für jede Storage-Gruppe im System spezifisch sind. Wenn Sie das interaktive Diagramm verwenden, um Änderungen des Datenreduzierungsverhältnisses nachzuerfolgen, lassen sich anhand der Storage-Gruppenliste Storage-Gruppen identifizieren, die große Mengen nicht reduzierbarer Daten enthalten, die sich auf das Datenreduzierungsverhältnis auswirken.

Berechnung von Effizienzverhältnissen: Die zur Berechnung des Datenreduzierungsverhältnisses erforderlichen Daten sind in einem Pop-up-Fenster im Datenreduzierungsdiagramm verfügbar.

- **Datenreduzierungsverhältnis:** Das Datenreduzierungsverhältnis wird aus den Werten für „Enabled“ und „Reducible“ der genutzten effektiven Kapazität und den Werten für „Enabled“ und „Reduced“ der verwendeten physischen Kapazität berechnet.

Aktiviert und reduzierbar ÷ Aktiviert und reduziert

- **Allgemeines Datenreduzierungsverhältnis:** Das allgemeine Datenreduzierungsverhältnis des Systems wird anhand der Gesamtwerte von „Effective Used“ und „Physical Used“ berechnet.

Effektiv genutzt gesamt ÷ Physisch genutzt gesamt

Unterstützte Datendienste

Übersicht

Die Datenreduzierung wird für FBA-Storage unterstützt. Gemischte FBA-/CKD-Systeme werden innerhalb desselben Storage-Ressourcenpools unterstützt, die Datenreduzierung für die CKD-Emulation nutzt jedoch nur Komprimierung und aktivitätsbasierte Reduzierung. Alle anderen Datendienste, die sowohl für die PowerMax- als auch die VMAX All Flash-Systeme angeboten werden, werden unterstützt. Dazu gehören die lokale Replikation (SnapVX), die Remotereplikation (SRDF), D@RE und VMware vSphere Virtual Volumes (vVols).

Lokale Replikation (SnapVX)

SnapVX-Snapshots schützen Anwendungen ohne die Verwendung von Ziel-Volumes zur Erfassung von Änderungsdaten, so genannter „Deltas“. Snapshot-Deltas werden automatisch am Speicher-Backend mit Zeigern auf die relevanten Point-in-Time-Images verwaltet. Die gemeinsame Ressourcennutzung

Design automatisch, um Cache-, Kapazitäts- und Performancevorteile zu erzielen.

Komprimierte Quelldaten bleiben komprimiert, wenn sie zu Snapshot-Deltas werden. Nicht komprimierte Quelldaten können komprimiert werden, wenn sie zu Snapshot-Deltas werden oder weniger aktiv sind. Leseaktivitäten über verknüpfte Ziele können verhindern, dass ein nicht komprimiertes Delta komprimiert wird, oder dazu führen, dass komprimierte Snapshot-Deltas dekomprimiert werden. Snapshot-Deltas sind für die Deduplizierung verfügbar.

Die Aktivierung der Datenreduzierung auf einem verknüpften Ziel wirkt sich nur auf Daten aus, die im Besitz des verknüpften Ziels sind. Daten auf verknüpften Zielen und Clones sind für die Deduplizierung verfügbar.

Remote-replikation (SRDF)

Die Komprimierung für SRDF wird unterstützt und als SRDF-Komprimierung bezeichnet. Bei der SRDF-Komprimierung handelt es sich um eine Funktion zur Reduzierung der Bandbreitennutzung beim Senden von Daten an und von Systemen, die mithilfe der Remotereplikation verbunden sind. DIE SRDF-Komprimierung und die Datenreduzierung verwenden die gleiche Hardware, dienen jedoch unterschiedlichen Zwecken. Durch Datenreduzierung komprimierte Daten werden dekomprimiert, bevor sie über die SRDF-Verknüpfung gesendet werden. Wenn SRDF-Komprimierung und Inline-Komprimierung verwendet werden, werden die Daten dekomprimiert, anschließend über die SRDF-Komprimierungsfunktion komprimiert und dann an den Remotestandort gesendet.

Data-at-Rest-Verschlüsselung (D@RE)

D@Re bietet eine Hardware- und Array-basierte Backend-Verschlüsselung, während die Datenreduzierung als ein Inline-Prozess ausgeführt wird. Die Daten werden über die Hardware zur Datenreduzierung weitergeleitet, bevor sie die Verschlüsselungshardware durchlaufen. Aus diesem Grund werden Daten komprimiert und/oder dedupliziert, bevor sie verschlüsselt werden. Auf einem D@Re-aktivierten System wurden Daten, die auf einer Festplatte verschlüsselt wurden, bereits komprimiert und/oder dedupliziert.

Virtuelle Volumes

Die Datenreduzierung wird für die Zuweisung von Daten zu vVols unterstützt und folgt dem gleichen I/O-Pfad wie alle anderen Daten. Der IO-Pfad wird auch in [Abbildung 1](#) angezeigt. Die Datenreduzierung ist auf Storage-Ressourcenebene in einem vVol-Storagecontainer aktiviert, da es keine Storage-Gruppen für vVols gibt.

Fazit

Zusammenfassung

Die Verwendung von physischer Storage-Kapazität ist ein häufig diskutiertes Thema für Storage-Administratoren in der gesamten Storage-Branche. Die stetig wachsenden Datenmengen haben den Bedarf nach mehr Effizienz bei der Nutzung physischer Kapazitäten verursacht. Dell PowerMax 2500- und 8500-Daten-Storage-Systeme diese Effizienz einen Schritt weiter. Die Datenreduzierung bietet außerordentliche Kapazitätseinsparungen bei gleichzeitig optimaler Performance. Dies verringert die benötigte Stellfläche im Rechenzentrum und reduziert die Gesamtbetriebskosten (TCO). Neben den Einsparungen ist die Aktivierung und Deaktivierung der Datenreduzierung mit einem Klick ein Kinderspiel. Das System übernimmt alle Aufgaben.

Referenzen

Dell Technologies Dokumentation

Die folgende Dell Technologies Dokumentation enthält weitere Informationen im Zusammenhang mit diesem Dokument. Der Zugriff auf diese Dokumente hängt von Ihren Anmeldedaten ab. Falls Sie auf ein Dokument nicht zugreifen können, wenden Sie sich an Ihren Dell Technologies Vertriebsmitarbeiter.

- [Info Hub für PowerMax und VMAX](#)