POWERVAULT MD3000 UND MD3000i

BEST PRACTICES FÜR DIE ARRAY-OPTIMIERUNG

dell.com/PowerVault

Inhaltsverzeichnis

1 Zielgruppe und Umfang

Dieses Dokument soll Benutzer des Dell™ PowerVault™ MD3000 und MD3000i durch die komplexen Vorgänge bei der Optimierung ihres Storage-Arrays führen, um es perfekt auf ihre jeweiligen Anforderungen abzustimmen. Es enthält die Best Practices, die bei der Leistungsoptimierung eines Storage-Arrays mit Firmware der ersten Generation (06.XX.XX.XX) und der zweiten Generation (07.XX.XX.XX) zu beachten sind. Weitere Informationen zur Ermittlung der Firmware-Generation eines MD3000- oder MD3000i-Storage-Arrays finden Sie im Dell™-*Benutzerhandbuch* unter http://support.dell.com/manuals.

2 Überblick über die Leistungsoptimierung

Die Herausforderung bei der Leistungsoptimierung von Massenspeicher besteht darin, die interagierenden Komponenten (unten aufgelistet) zu verstehen und zu steuern und gleichzeitig die Anwendungsleistung zu ermitteln. Da die Leistung des Storage-Arrays nur einen Bruchteil der gesamten Anwendungsleistung ausmacht, muss die Optimierung unter Berücksichtigung der E/A-Eigenschaften der Anwendung und aller Komponenten erfolgen, die im Datenpfad involviert sind, wie zum Beispiel SAS HBA, iSCSI-Initiator, Netzwerk-Switch und der Einstellungen des Host-Betriebssystems.

Da zahlreiche Aspekte einzubeziehen sind, ist schon die Leistungsoptimierung zur Maximierung der Leistung einer einzigen Anwendung eine anspruchsvolle Aufgabe. Die Systemoptimierung zur Maximierung der Leistung mehrerer Anwendungen, die gemeinsam auf ein einziges Storage-Array zugreifen, gestaltet sich noch komplexer. Um diese Komplexität zu reduzieren, bieten Dell™-Speichersysteme Funktionen zur Leistungsüberwachung und flexiblen Optimierung, auf die über den Modular Disk Storage Manager (MDSM) zugegriffen werden kann.

2.1 Komponenten, die die Massenspeicher-Leistung beeinflussen

Dieses White Paper beschreibt einen Gesamtansatz für die Optimierung der E/A-Leistung und enthält zudem spezifische Leitlinien zur Verwendung der Optimierungsfunktionen für Storage-Arrays. Zunächst erfolgt eine grundlegende Analyse der Elemente, die die E/A-Leistung beeinflussen:

- Storage-Array
- Anwendungssoftware
- Serverplattform (Hardware, Betriebssystem, Volume Manager, Gerätetreiber)
- Netzwerk (nur bei MD3000i)

2.2 Grundlegender Ansatz zur Leistungsoptimierung

Für die Optimierung der E/A-Leistung muss zunächst folgende Frage geklärt werden:

Wie leistungsstark soll mein System sein?

Folgende Antworten sind möglich:

- "Das hängt davon ab ..." Es gibt keine absolute Antwort. Jede Umgebung ist einzigartig, und die richtigen Einstellungen hängen von den jeweiligen Zielen, der Konfiguration und den Anforderungen an die Umgebung ab.

- "Die Auslastung schwankt." Die Ergebnisse sind sehr unterschiedlich, weil die Bedingungen stark variieren.

Die Antworten auf diese Frage ergeben den folgenden grundlegenden Ansatz zur Leistungsoptimierung:

1 – Konfigurieren und Testen
2 – Messen
3 – Nach Bedarf anpassen

Aufgrund der Merkmale zur Leistungsüberwachung in allen MD3000- und MD3000i-Speichersystemen und der Optimierungsfunktionen eignen sich die Systeme hervorragend für diesen iterativen Prozess. Der erste Optimierungsschritt besteht darin, die derzeitige Leistung als Ausgangswert zu ermitteln. Zur Ermittlung eines solchen Leistungsausgangswerts empfiehlt sich das Zugrundelegen einer Auslastung, die möglichst genau der beabsichtigten Endnutzung der Speicherlösung entspricht. Diese kann ganz einfach in der tatsächlichen Anwendung oder einem SQL Replay mit einem Systemleistungsmonitor (perfmon oder sysstat/iostat ebenso wie CLI und Leistungsüberwachung zur Zustandserfassung) oder einem synthetischen Benchmark-Paket bestehen, welches das erwartete E/A-Profil fast genau imitiert (Iometer, IOZone oder Bonnie). Durch den Vergleich der Ausgangsdaten mit den geschätzten Anforderungen und der Kapazität der Konfiguration kann der Benutzer ein MD3000- oder MD3000i-Storage-Array effektiv optimieren. Dieses White Paper enthält Empfehlungen für diesen wichtigen ersten Schritt sowie Optimierungstipps, um die Kapazitäten der MD3000- und MD3000i-Speichersysteme voll auszuschöpfen.

3 Berücksichtigung der Anwendungssoftware

Bei der Ermittlung der E/A-Eigenschaften der intendierten Anwendungen ist es erforderlich, den Massenspeicher möglichst entsprechend der erwarteten

Laufzeit zu nutzen, um die optimale Speicher- und Systemkonfiguration zu bestimmen. Nur so ist eine effektive Optimierung der Gesamtlösung möglich. Hierzu gehört u. a. Folgendes:

- Anzahl der separaten E/A-Quellen, die mit der Lösung interagieren
- Zufälligkeit des Datenzugriffs durch die primäre(n) E/A-Quelle(n)
- Durchschnittsgröße typischer E/A-Bausteine (zumeist in folgende drei Kategorien unterteilt):
  - Großer Block (≥ 256 KiB) zur Übertragung
  - Mittelgroßer Block (≥ 32 KiB und < 256 KiB) zur Übertragung
  - Kleiner Block (< 32 KiB) zur Übertragung
- "Burstiness" von E/A-Strukturen, d. h. der durchschnittliche Tastgrad der E/A an das Storage-Array
- Profil der durchschnittlichen E/A-Richtung; normalerweise das Verhältnis von Lese- zu Schreibvorgängen

4 Konfiguration des MD3000/MD3000i

Es gibt zwei Verfahrensweisen zur Konfiguration der MD3000- und MD3000i-Speichersysteme. Die gebräuchlichste und einfachste Methode besteht in der Nutzung des MDSM. Der MDSM gestattet automatische Konfigurationseinstellungen, die zweckdienliche Konfigurationen mit einem geringen Wissen über Leistungsoptimierungen ermöglichen.

Es steht auch eine manuelle Konfigurationsoption mithilfe der Befehlszeilenschnittstelle (CLI) zur Verfügung, die mehr Flexibilität bietet, aber auch mehr Wissen über die Leistungsanforderungen erfordert.

Links zu den MDSM- und CLI-Handbüchern finden Sie unter Appendix C: References.

4.1 Ermitteln des optimalen RAID-Levels

Der erste Schritt bei der Optimierung eines externen MD3000- oder MD3000i-Storage-Arrays besteht darin, je nach Anwendung das geeignetste RAID-Level für die Lösungen zu ermitteln. Bitte beachten Sie, dass im Folgenden aufgrund des mangelnden Datenschutzes RAID 0 bei den meisten Informationen nicht berücksichtigt wird. Dies bedeutet nicht, dass die Nutzung von RAID 0 unerwünscht wäre, sondern lediglich, dass es nur für nicht kritische Daten verwendet werden sollte. Im Allgemeinen bietet RAID 0 eine bessere Leistung als RAID 1/10, 5 oder 6. Außerdem wird RAID 6 nicht immer speziell aufgeführt. Die meisten Anmerkungen zur Optimierung von RAID 5 sind jedoch direkt auf RAID 6 übertragbar, sofern nicht anders angegeben. In Situationen, in denen die höhere Fehlertoleranz von RAID 6 gewünscht wird, beachten Sie bitte, dass

beim direkten Vergleich mit RAID 5 eine Leistungseinschränkung aufgrund der zusätzlichen Paritätsberechnung und des zusätzlichen physischen Datenträgers für die Implementierung zu beobachten ist.

Eine wichtige Überlegung bei der Ermittlung des geeigneten RAID-Levels sind die Kosten physischer Datenträger, die für ein RAID-Level erforderlich sind. Die Kosten physischer Datenträger ergeben sich aus der Anzahl physischer Laufwerke und deren Kapazität, die für die gewünschte Datenintegrität aufgewendet werden. Die Kosten physischer Datenträger sind für jedes RAID-Level unterschiedlich und können die Entscheidung beeinflussen, welches RAID-Level für die jeweilige Umgebung am geeignetsten ist. RAID 0 bietet keine Redundanzstufe, und die Kosten physischer Datenträger betragen Null. RAID 1/10 hat die höchsten Datenträgerkosten in Datenträgergruppen mit mehr als zwei Laufwerken. Bei RAID 1/10 wird immer die Hälfte der physischen Datenträger für die Spiegelung aufgewendet. RAID 5 hat fixe Kosten von einem physischen Datenträger pro Datenträgergruppe, d. h. bei einer RAID 5-Gruppe steht nur n-1 der Kapazität zur Verfügung. Ebenso hat RAID 6 Fixkosten von zwei physischen Datenträgern pro Datenträgergruppe, also n-2. Bei RAID 5 und 6 liefern diese zusätzlichen Datenträger den Platz, der zum Erhalt der Paritätsinformationen für jeden Block erforderlich ist.

Die Kosten physischer Datenträger sind nicht der einzige Faktor, der die Entscheidung beeinflusst, welches RAID-Level für eine bestimmte Anwendung am geeignetsten ist. Die Leistung eines RAID-Levels ist stark abhängig von den Eigenschaften der E/A-Struktur, die von dem Host/den Hosts an das Storage-Array übertragen wird. Bei E/A-Strukturen mit Schreibvorgängen sollte ein E/A-Burst, der 1/3 des verfügbaren Cache-Speichers übersteigt, als lange E/A betrachtet werden. Lange Schreibvorgänge zeigen die Leistung eines RAID-Levels besser auf als kurze. Kurze Schreibvorgänge können komplett im Cache verarbeitet werden, und die Auswirkung des RAID-Levels auf die Leistung wird minimiert. Solange der Burstiness-Wert der Schreibvorgänge stets niedriger ist als die Offload-Rate vom Cache zur Festplatte, ist die Wahl des RAID-Levels unter Umständen von geringer Bedeutung.

Grundsätzlich sind die folgenden RAID-Level unter den jeweiligen Bedingungen am besten geeignet:

- RAID 5 und RAID 6 für sequenzielle, große E/As (> 256 KiB)
- RAID 5 oder RAID 1/10 für kleine E/As (< 32 KiB)
- Bei dazwischenliegenden E/A-Größen richtet sich das RAID-Level nach anderen Anwendungseigenschaften:
    - RAID 5 und RAID 1/10 sind für die meisten Lese- und sequenziellen Schreibumgebungen ähnlich gut geeignet.

- RAID 5 und RAID 6 weisen hauptsächlich durch zufällige Schreibvorgänge die schlechteste Leistung auf.
- In zufälligen E/A-Anwendungen mit mehr als 10 % Schreibvorgängen bietet RAID 1/10 die beste Leistung.

Table 1 bietet eine Übersicht dieser Punkte für eine ideale Umgebung. Eine ideale Umgebung besteht aus abgestimmten Blöcken oder Segmenten, Lese- und Schreibvorgängen, bei denen Cache-Speicher und RAID-Controller-Module nicht übermäßig durch E/A-Vorgänge ausgelastet werden.

Tabelle 1: E/A-Größe und optimales RAID-Level

	Signifikant zufällig		**Signifikant sequenziell**	
Blockgröße	Lesen	Schreiben	Lesen	Schreiben
Klein (< 32 KiB)	1/10, 5, 6	1/10	1/10, 5, 6	1/10, 5
Mittelgroß (Zwischen 32 und 256 KiB)	1/10, 5, 6	1/10	1/10, 5, 6	5
Groß (> 256 KiB)	1/10, 5, 6	1/10	1/10, 5, 6	5

4.1.1 Auswahl eines RAID-Levels – Szenario mit hohem Schreibanteil

In zufälligen E/A-Anwendungen mit einem Anteil von > 10 % Schreibvorgängen und einem geringen Grad an Burstiness bietet RAID 1/10 die beste Gesamtleistung für redundante Datenträgergruppen.

Die Leistung von RAID 1/10 kann in solchen Umgebungen > 20 % höher sein als die von RAID 5, bringt jedoch die höchsten Datenträgerkosten mit sich. So müssen beispielsweise mehr physische Datenträger angeschafft werden. RAID 5 bietet Schutz und minimiert die Datenträgerkosten für die Nettokapazität, wird aber durch den Schreibleistungs-Overhead von Paritätsaktualisierungen stark beansprucht.

RAID 6 bietet zwar mehr Schutz als RAID 5 bei minimalen Datenträgerkosten, ist jedoch aufgrund der erforderlichen Double Parity-Berechnungen stärker vom Schreibleistungs-Overhead betroffen.

In sequenziellen E/A-Anwendungen mit relativ geringen Schreibübertragungsgrößen wirkt sich das RAID-Level nicht so stark aus. Bei mittleren Übertragungsgrößen kann RAID 1/10 gegenüber RAID 5/6 Vorteile bieten, jedoch wiederum verbunden mit höheren Datenträgerkosten. Bei sehr großen sequenziellen Schreibvorgängen kann RAID 5 eine genauso gute oder bessere Leistung aufweisen als RAID 1/10, insbesondere wenn die Datenträgerkosten für gleichwertige Kapazität berücksichtigt werden. Darüber hinaus wird immer dann eine bessere Leistung erzielt, wenn die Anwendung

oder das Betriebssystem Schreibvorgänge puffern oder zusammenfließen lassen kann, sodass diese ein ganzes Segment bzw. einen ganzen Block ausfüllen.

4.1.2 Auswahl eines RAID-Levels – Szenario mit geringem Schreibanteil

In zufälligen E/A-Anwendungen mit einem geringen Anteil (< 10 %) von Schreibvorgängen bietet RAID 5 in etwa die gleiche Leistungsstärke wie RAID 1/10 bei kleinen Übertragungsgrößen, aber zu geringeren Datenträgerkosten. RAID 0 liefert eine etwas bessere Leistung als 5 oder 1/10, aber keinen Datenschutz. In Umgebungen mit größeren Übertragungsgrößen kann die RAID 1/10-Leistung etwas besser sein, sie bringt aber wesentlich höhere Datenträgerkosten mit sich.

4.2 Auswählen der Anzahl der Laufwerke in einer Datenträgergruppe

Bei der Optimierung der Leistung müssen viele Faktoren berücksichtigt werden, z. B. Laufwerkstyp und -kapazität sowie die Anzahl der Laufwerke.

Die folgenden Empfehlungen gelten für die Gruppierung von Laufwerken in eine Datenträgergruppe:

- Teilen Sie zufällige und sequenzielle Arbeitslasten auf unterschiedliche Datenträgergruppen auf: Trennung des E/A-Verkehrs, um die gemeinsame Nutzung von Datenträgergruppen unter virtuellen Laufwerken zu minimieren.
- Wählen Sie schnellere Laufwerke aus: Im Allgemeinen bringt ein Laufwerk mit 15.000 1/min bei gemischten sequenziellen und zufälligen Vorgängen etwa 20 % mehr Leistung als eines mit 10.000 1/min. Bitte beachten Sie die Datenblätter des Herstellers, um das geeignetste Laufwerk zu ermitteln.
- Durch Hinzufügen von mehr Laufwerken zu einer Datenträgergruppe bei gleich bleibender Blockgröße können Sie die E/A-Rate für sequenzielle E/A-Bausteine bis zur kompletten Controller-Auslastung steigern: Mehr Laufwerke bedeutet eine bessere Bedienung der E/A-Anforderungen.
- Zur Optimierung der Datenübertragungsrate multiplizieren Sie die Anzahl der physischen Datenträger mit der Segmentgröße, um die E/A-Größe zu erzielen. Es gibt jedoch immer auch Ausnahmen. Bei kleinen/mittelgroßen E/A-Bausteinen ist es nicht sinnvoll, die E/As so aufzuteilen, dass eine noch kleinere E/A an die Laufwerke gesendet wird. Bitte beachten Sie, dass zu Datenlaufwerken keine Paritäts- oder Spiegelungslaufwerke zählen, die in einer RAID-Gruppe genutzt werden.

Bei IOPs oder transaktionsorientierten Anwendungen spielt die Anzahl der Laufwerke eine größere Rolle, weil die zufälligen E/A-Raten der Laufwerke relativ gering sind. Wählen Sie eine Anzahl von Laufwerken, die bei der jeweiligen virtuellen Laufwerksgruppe der E/A-Rate entspricht, die für die Unterstützung der Anwendung benötigt wird. Achten Sie darauf, die für den Datenschutz des

gewählten RAID-Levels zu implementierenden E/As zu berücksichtigen. Gestalten Sie die Segmentgröße mindestens so groß wie die typische Größe der Anwendungs-E/A.

Eine Segmentgröße von 128.000 ist ein sinnvoller Ausgangspunkt für die meisten Anwendungen. Für die meisten Anwendungen gilt: je höher die Anzahl an Laufwerken in einer Datenträgergruppe, desto besser die durchschnittliche Leistung. Die Laufwerkszahl einer vorhandenen Speicherlaufwerksgruppe kann mithilfe der CLI erhöht werden.

4.3 Position und Kapazität virtueller Laufwerke

Die Position virtueller Laufwerke in einer Datenträgergruppe, die Anzahl und Position zugewiesener virtueller Laufwerke innerhalb einer Datenträgergruppe sowie die Kapazität eines virtuellen Laufwerks sind wichtige Faktoren, die bei der Leistungsoptimierung eines Arrays berücksichtigt werden sollten.

Wenn rotierende Speichermedien verwendet werden, wirken sich die Kapazität einer virtuellen Festplatte und ihre Position innerhalb einer Datenträgergruppe enorm auf die erzielte Leistung aus. Dies liegt in erster Linie an Unterschieden in der Winkelgeschwindigkeit in den äußeren Bereichen. Der Effekt der Zuweisung der äußersten Randbereiche eines rotierenden Speichermediums zur Steigerung der Leistung wird *Short-Stroking* eines Laufwerks genannt. Die technischen Einzelheiten des Short-Strokings können nicht im Rahmen dieses White Papers erläutert werden. Man kann jedoch sagen, dass normalerweise das äußere Drittel eines rotierenden Speichermediums am schnellsten ist, während die inneren Zonen am langsamsten sind. Short-Stroking ist leicht umzusetzen, indem man eine Datenträgergruppe erstellt, die aus einem einzelnen virtuellen Laufwerk besteht, dem weniger als ein Drittel der Gesamtkapazität zugewiesen wird. Der klare Nachteil des Short-Stroking eines Speicherlaufwerks besteht in dem Verlust zusätzlicher nutzbarer Kapazität. Die Leistungssteigerung muss daher unmittelbar gegen diesen Kapazitätsverlust abgewogen werden.

Neben den Leistungszunahmen durch Short-Stroking sollte der Effekt der Laufwerkkopfsuche bei der Unterteilung einer Datenträgergruppe in virtuelle Laufwerke berücksichtigt werden. Virtuelle Laufwerke werden in einer Datenträgergruppe in Reihen angeordnet, wobei das erste virtuelle Laufwerk in den schnellsten äußeren Bereichen liegt und die Reihe sich dann nach Innen fortsetzt. Wenn man dies berücksichtigt, sollte eine Datenträgergruppe mit möglichst wenig virtuellen Laufwerken konzipiert werden.

Für optimale Leistung empfiehlt Dell™, nicht mehr als vier virtuelle Laufwerke oder Repositorys pro Datenträgergruppe zu verwenden. Wo Leistung ein entscheidender Faktor ist, sollten virtuelle Laufwerke außerdem möglichst in separate Datenträgergruppen ausgelagert werden. Wenn mehrere virtuelle Laufwerke mit starkem Datenverkehr in einer Datenträgergruppe vorhanden sind, wird das E/A-Verhalten dieser Gruppe selbst bei rein sequenziellen Verwendungsmodellen zunehmend zufällig gestaltet, was die Gesamtleistung

beeinträchtigt. Ist die gemeinsame Nutzung einer Datenträgergruppe unumgänglich, sollte das Laufwerk mit dem stärksten Verkehr stets am Anfang der Datenträgergruppe stehen.

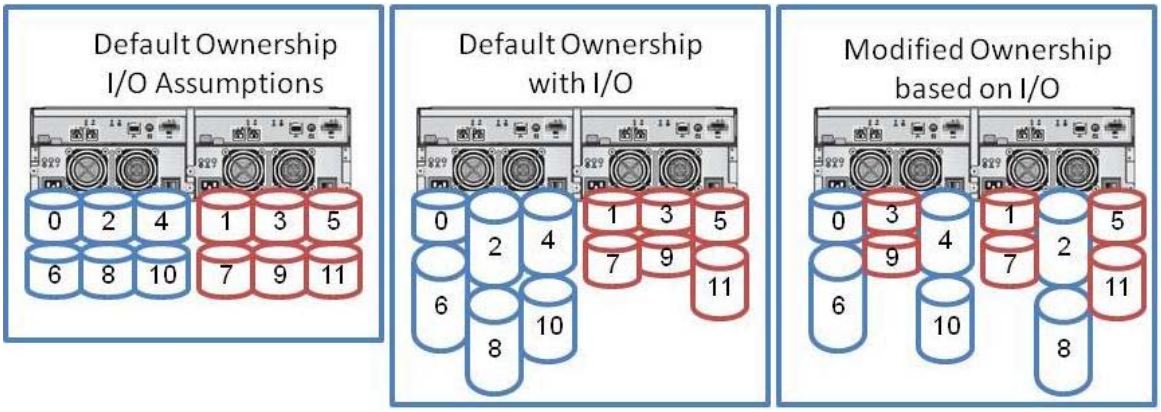
4.4 *Zuordnung von virtuellen Laufwerken*

Der Dell™ MDSM kann dazu verwendet werden, virtuelle Laufwerke automatisch zu erstellen und anzuzeigen. Er verwendet die optimalen Einstellungen für das Striping der Datenträgergruppe. Virtuelle Laufwerke werden bei der Erstellung wechselweise anderen RAID-Controllern zugeordnet. Diese Standardzuordnung stellt eine einfache Maßnahme für eine ausgeglichene Auslastung der RAID-Controller dar. Die Zuordnung kann später geändert werden, um einen Ausgleich auf Basis der tatsächlichen Auslastung zu schaffen. Wenn die Zuordnung virtueller Laufwerke nicht manuell abgestimmt wird, ist es möglich, dass ein Controller die Hauptlast trägt, während der andere gar nicht genutzt wird.

Begrenzen Sie die Anzahl der virtuellen Laufwerke in einer Datenträgergruppe. Wenn sich mehrere virtuelle Laufwerke in einer Datenträgergruppe befinden, berücksichtigen Sie Folgendes:

- Bedenken Sie die Rückwirkungen der einzelnen virtuellen Laufwerke auf die anderen virtuellen Laufwerke in der Datenträgergruppe.
- Vergegenwärtigen Sie sich die Nutzungsmuster für die einzelnen virtuellen Laufwerke.
- Die Auslastung kann je nach Tageszeit unterschiedlich sein.

Abbildung 1: Ausgewogene Zuordnung virtueller Laufwerke



4.5 *Berechnung der optimalen Segment- und Blockgröße*

Die Wahl der Segmentgröße kann sich entscheidend auf die Leistung in Bezug auf IOPS sowie auf die Datenübertragungsrate auswirken.

Der Begriff *Segmentgröße* bezieht sich auf die Menge der Daten, die auf ein Laufwerk in einer virtuellen Datenträgergruppe geschrieben werden, bevor Daten auf das nächste Laufwerk in der Gruppe geschrieben werden. Eine zusammenhängende Reihe von Segmenten über mehrere Laufwerke hinweg bildet einen *Block*. In einer virtuellen Datenträgergruppe der RAID-Level 5, 4 + 1 mit einer Segmentgröße von 128 KiB werden zum Beispiel die ersten 128 KiB eines E/A-Bausteins auf das erste Laufwerk geschrieben, die nächsten 128 KiB auf das nächste Laufwerk usw., bis zu einer Gesamtblockgröße von 512 KiB. Bei einer virtuellen Laufwerksgruppe der RAID-Level 1, 2 + 2 würden 128 KiB auf die beiden Laufwerke geschrieben (genauso bei gespiegelten Laufwerken). Wenn die E/A-Größe die Anzahl der Laufwerke mal 128 KiB überschreitet, wird das Muster wiederholt, bis der ganze E/A-Baustein abgearbeitet ist.

Bei sehr großen E/A-Aufrufen ist die Segmentgröße für eine RAID-Datenträgergruppe im Idealfall so zu wählen, dass ein einzelner Host-E/A-Baustein über alle Datenlaufwerke in einem Block verteilt wird. Die Formel für die maximale Blockgröße lautet also:

> LUN-Segmentgröße = Maximale E/A-Größe ÷ Anzahl an Datenlaufwerken

Die LUN-Segmentgröße sollte jedoch auf die nächste unterstützte Zweierpotenz aufgerundet werden.

Bei RAID 5 und 6 ist die Anzahl der Datenlaufwerke gleich der Anzahl der Laufwerke in der Datenträgergruppe minus 1 bzw. 2. Beispiel:

> RAID 5, 4 + 1 mit einer Segmentgröße von 64 KiB => (5-1) x 64 KiB = 256 KiB Blockgröße

Im Idealfall reicht diese RAID-Gruppe aus, um E/A-Aufrufe unter oder bis zu 256 KiB zu verarbeiten.

Bei RAID 1 ist die Anzahl der Datenlaufwerke gleich der Anzahl der Laufwerke geteilt durch 2. Beispiel:

> RAID 1/10, 2 + 2 mit 64 KiB Segmentgröße => (4-2) x 64 KiB = 128 KiB Blockgröße

Es darf nicht vergessen werden, dass Segment- und Blockgröße je nach den E/A-Parametern der Anwendung variieren.

Für Anwendungsprofile mit kleinen E/A-Aufrufen sollten Sie die Segmentgröße groß genug wählen, um die Anzahl der Segmente (Laufwerken im LUN), auf die zur Abdeckung des E/A-Aufrufs zugegriffen wird, zu minimieren, d. h. um die Überschreitungen von Segmentgrenzen zu minimieren. Falls von der

Anwendung nicht anders vorgegeben, wird als Ausgangspunkt eine Standardsegmentgröße von 128 KiB empfohlen.

Es ist entscheidend, dass die Blockgröße korrekt gewählt wird, damit das Betriebssystem des Hosts, sofern möglich, stets Aufrufe für ganze Blöcke oder ganze Segmente vornimmt.

4.6 Cache-Einstellungen

Read-Ahead-Cache kann im MDSM und über die CLI konfiguriert werden. Im MDSM stehen nur die Standardeinstellungen zur Verfügung, während über die CLI der Read-Ahead-Cache komplett konfiguriert werden kann. Außerdem kann die globale Cache-Blockgröße für Read-and-Write-Cache über die CLI eingestellt werden.

Eine vollständige Liste der unterstützten Befehle einschließlich der folgenden Cache-spezifischen Befehle finden Sie im *Dell™ PowerVault™-Handbuch für den Modular Disk Storage Manager und CLI* auf der Dell™-Website für technischen Support (http://support.dell.com/manuals).

4.6.1 Einstellen des für virtuelle Laufwerke spezifischen Schreib-Cache und der Schreib-Cache-Spiegelung

Über die CLI konfiguriert: Diese Befehle stehen auf der Ebene des virtuellen Laufwerks zur Verfügung.

Schreib-Cache: Durch Deaktivierung des Schreib-Cache stellen Sie die Controller auf einen Write-Through-Modus ein, was zusätzliche Latenz schafft, während Daten an das Laufwerk weitergeleitet werden. Außer bei speziellen schreibgeschützten Umgebungen wird empfohlen, diese Einstellung beizubehalten. Schreib-Cache wird im Falle eines Cache-Akkuausfalls oder eines Lernzyklus des Cache-Akkus automatisch deaktiviert.

Schreib-Cache-Spiegelung: Die Schreib-Cache-Spiegelung bietet eine zusätzliche Ebene der Redundanz und Fehlertoleranz beim MD3000 und MD3000i. Als Nebeneffekt reduziert sie bei diesem Vorgang den verfügbaren physischen Speicher sowie die Intra-Controller-Bandbreite. In ausgewählten, nicht datenkritischen Fällen kann es vorteilhaft sein, diesen Parameter abzuändern. Für den normalen Gebrauch empfiehlt Dell™ stets die Aktivierung der Cache-Spiegelung. Im Fall eines Controllerausfalls oder wenn Schreib-Cache deaktiviert ist, wird die Cache-Spiegelung automatisch deaktiviert.

***ACHTUNG**: Es kann zu Datenverlusten kommen, wenn ein RAID-Controllermodul ausfällt, während in den Schreib-Cache geschrieben wird, ohne dass Cache-Spiegelung auf einem virtuellen Laufwerk aktiviert ist.*

4.6.2 Einstellung der für virtuelle Laufwerke spezifischen Prefetch-Funktion des Lese-Cache

Über die CLI konfiguriert: Dieser Befehl steht auf der Ebene des virtuellen Laufwerks zur Verfügung.

Prefetch-Funktion des Lese-Cache: Die Einstellung des Lese-Cache kann auf der Ebene eines virtuellen Laufwerks geändert werden. Die Deaktivierung der Lese-Prefetch-Funktion ist vor allem in zufälligen Lese-Umgebungen mit geringen Übertragungsgrößen sinnvoll, wo ein Prefetching zufälliger Daten keinen ausreichenden Wert liefern würde. Die normal zu beobachtenden Kosten der Lese-Prefetch-Funktion sind jedoch nicht nennenswert. Für die meisten Umgebungen empfiehlt Dell™ stets die Aktivierung von Lese-Cache-Prefetch.

4.6.3 Einstellung der Cache-Blockgröße des Storage-Arrays

Über die CLI konfiguriert: Dieser Befehl steht auf der Ebene des Storage-Arrays zur Verfügung und wirkt sich auf alle virtuellen Laufwerke und Datenträgergruppen aus.

Cache-Blockgröße: Die Cache-Blockgröße bezieht sich darauf, wie der Cache-Speicher bei der Belegung segmentiert wird, und betrifft alle virtuellen Laufwerke in einem Array. Auf dem MD3000 und MD3000i stehen die Einstellungen 4 KiB und 16 KiB zur Verfügung, wobei 4 KiB die Standardeinstellung ist. Drastische Auswirkungen auf die Leistung sind möglich, indem Sie die korrekte Einstellung der Cache-Blockgröße auswählen, die dem E/A-Profil des Systems entspricht. Wenn die übliche E/A-Größe ≥ 16 KiB beträgt, was für sequenzielle E/As typisch ist, stellen Sie die Cache-Blockgröße des Storage-Arrays auf 16. Bei kleineren E/A-Größen (≤ 8 KiB), insbesondere bei stark zufälliger oder transaktionaler Nutzung, wird die Standardeinstellung von 4 KiB bevorzugt. Da diese Einstellung alle virtuellen Laufwerke eines Storage-Arrays betrifft, sollte eine Änderung nur unter Berücksichtigung der E/A-Anforderungen der Anwendung erfolgen.

Tabelle 2: Standardkonfigurationseinstellungen für Storage-Arrays

Option	MDSM-Konfigurationsvorlagen für Benutzeroberflächen			CLI-Optionen
	Dateisystem	**Datenbank**	**Multimedia**	
Laufwerkstyp	Wählbar	Wählbar	Wählbar	Wählbar
RAID-Level	Wählbar	Wählbar	Wählbar	0, 1/10, 5, 6
Segmentgröße [1]	128 KiB	128 KiB	256 KiB	8 KiB, 16 KiB, 32 KiB, 64 KiB, 128 KiB, 256 KiB, 512 KiB
Schreib-Cache mit Spiegelung	Immer ein	Immer ein	Immer ein	Ein oder aus
Read-Ahead-Cache	Ein	Aus	Ein	Ein oder Aus
Cache-Blockgröße [1]	Array-Standardeinstellung: 4 KiB			4 KiB, 16 KiB

4.7 Optimierung mithilfe von Array-Leistungsdaten

4.7.1 Sammeln statistischer Leistungsdaten

Die Dateien **stateCaptureData.txt** und **performanceStatistics.csv**, die über die Registerkarte **Support** des MDSM als Teil eines Technischen Support-Pakets verfügbar sind, liefern wertvolle statistische Daten in einem leicht lesbaren Format. Im folgenden Abschnitt finden Sie einige Beispieldaten aus der Datei **stateCaptureData.txt** und Konfigurationsempfehlungen auf der Basis der Leistungsgesichtspunkte, die im vorherigen Abschnitt aufgeführt wurden.

Weitere hilfreiche Informationen sind über das Array-Profil verfügbar. Öffnen Sie MDSM, und wählen Sie die Registerkarte "Support – View Storage Array Profile" (Storage Array-Profil anzeigen) aus.

Bevor Sie statistische Leistungsdaten sammeln, sollte die zu testende E/A-Auslastung ausgeführt werden. Dadurch stellen Sie die Gültigkeit der Leistungsdaten als Teil des Messvorgangs der eigentlichen Leistungsoptimierung sicher.

[1] Im MDSM CLI-Handbuch und in der SMcli-Anwendung wird unter Umständen noch die veraltete Bezeichnung KB für Kilobyte bzw. 2^{10} Byte verwendet. In diesem White Paper wird stattdessen der SI-Begriff Kibibyte benutzt. Bei der Formulierung eines SMcli-Befehls ist jedoch weiterhin das KB-Postfix erforderlich. In den Normen IEEE 1541-2002 und IEC 60027-2 finden Sie die Binärpräfixe der Maßeinheiten.

***Hinweis**: Die folgenden Abbildungen stammen aus dem Leistungs-Tool IOmeter.*

4.7.2 RAID-Level

Die Datei **stateCaptureData.txt** liefert statistische Daten in den Prozentspalten "reads" und "writes", die bei der Auswahl des geeignetsten RAID-Levels hilfreich sind. In Figure 2 liefern die E/A-Prozentangaben unter "small reads" und "small writes" Informationen zur Verteilung der E/A-Typen in der getesteten Auslastung. Das ist insbesondere dann hilfreich, wenn man Table 1 heranzieht, auf die auf Seite 8 verwiesen wird, und das aktuelle Lese-/Schreibverhältnis der Anwendungen bestimmt. Das gewählte RAID-Level kann sich auf die E/A-Leistung auswirken. Im Allgemeinen bietet RAID 1/10 die beste Gesamtleistung, allerdings bei den höchsten Kosten physischer Datenträger. Nutzen Sie bei dieser Ermittlung die E/A-Prozentverteilung sowie die durchschnittliche Blockgröße aus den gesammelten Daten. Diese Felder finden Sie in den hervorgehobenen Bereichen von Figure 2 und Figure 3 für Firmware der ersten bzw. zweiten Generation. Es ist zu beachten, dass die Werte in diesen Abbildungen in Blocknotation aufgeführt sind; die Blockgröße für die spezifische Konfiguration des virtuellen Laufwerks ist in der Datei **stateCaptureData.txt** zu finden und beträgt fast immer 512 Byte. Die durchschnittlich eingegangene E/A ist nicht die E/A-Größe, die von der Anwendung genutzt wird, sondern die, die der Host sendet. Eine Anwendung könnte also versuchen, größere E/As zu senden, der E/A-Stack des Hosts kann diese dann jedoch nach Bedarf zusammenfließen lassen oder aufteilen. Bitte schlagen Sie diese Werte in den jeweiligen Unterlagen zu Ihrem Betriebssystem oder HBA nach.

Abbildung 2: Firmware der ersten Generation – RAID-Level. Datei: stateCaptureData.txt

```
Virtual Disk Unit 0 Configuration
   Volume Type:          13+1 RAID 5
   User Label:           MyRAID5_1
   Block Size:           512 bytes
   Large IO:             4096 blocks
   Segment Size:         256 blocks
   Stripe Size:          3328 blocks
...
   E/A-Statistik:
                 small      small      large      large                cache
                 reads     writes      reads     writes      total       hits
  requests  2028332119  147699066          0          0 2176031185 1289775370
    blocks  3091968111 2518067526          0          0 1315068341 4019884678
avg blocks           4         17          0          0          0          3
   IO pct.       93.21%      6.78%     0.00%      0.00%      0.00%     59.27%

                  IOs    stripes        /IO   clusters        /IO
     reads  2028332119 2034477363        1.00 2107869128        1.03
    writes   147699066  148449472        1.00  157404718        1.06

     write        Full    Partial         RMW  No Parity        RMW2        FSWT
algorithms     1105611   12598366    32120072          0           0           0
```

Abbildung 3: Firmware der zweiten Generation – RAID-Level. Datei: stateCaptureData.txt

```
Volume 0 Attributes:
     Volume Type:       RAIDVolume
     User Label:        MyRAID10_One
...
     BlockSize:         512 bytes
     LargeIoSize:       4096 blocks
...
     Perf. Stats:       Requests     Blocks  Avg. Blks  IO Percent
        Reads           67456452 5943724625         88      71.20%
        Writes          27283249 1144902648         41      28.80%
        Large Reads            0          0          0       0.00%
        Large Writes           0          0          0       0.00%
        Total           94739701 7088627273         74     100.00%
```

4.7.3 E/A-Verteilung

E/A kann nach Verteilung und Struktur charakterisiert werden. Die zwei Hauptfaktoren bei der Bestimmung der E/A-Verteilung einer Anwendung sind die Zufälligkeit der E/A und die Richtung der E/A. Die Zufälligkeit der E/A gibt an, wie sequenziell oder zufällig der Datenzugriff ist, und beschreibt die Struktur dieses Datenzugriffs. Die Richtung der E/A kann ganz einfach auf die Lese- und

Schreibprozentsätze der E/A bezogen werden, d. h. welche Richtung E/As vom Speichergerät aus nehmen. Der Begriff E/A-Struktur beschreibt, wie eng die Varianz sequenzieller oder zufälliger Datenzugriffe im Speicherlaufwerk gehalten wird. Dabei kann es sich um rein zufälligen Zugriff innerhalb eines gesamten virtuellen Laufwerks handeln oder zufällig innerhalb gewisser Grenzen, z. B. eine große Datei, die auf einem virtuellen Laufwerk gespeichert wird, im Vergleich zu großen, nicht zusammenhängenden Häufungen sequenzieller Datenzugriffe, die innerhalb gewisser Grenzen zufällig verteilt werden. Hierbei handelt es sich um unterschiedliche E/A-Strukturen, für die bei der Optimierung des Massenspeichers verschiedene Fälle angewendet werden müssen.

Die Daten aus der Datei **stateCaptureData.txt** können Ihnen helfen, diese Eigenschaften zu ermitteln. Der Prozentsatz sequenzieller Lesevorgänge kann anhand des Prozentsatzes aller Cache-Treffer insgesamt bestimmt werden. Sind Cache-Treffer und Leseprozentsatz hoch, können Sie zunächst davon ausgehen, dass die E/A-Struktur eher sequenziell ist. Da Cache-Treffer jedoch nicht statistisch in Lese- und Schreibvorgänge unterteilt werden, muss unter Umständen ein wenig mit dem repräsentativen Datensatz experimentiert werden, falls die Struktur nicht bekannt ist. Im Falle von E/A-Hoststrømen mit nur einem Thread kann das Verhalten geprüft werden, indem man die Menge an Lesevorgängen mit der Prefetch-Statistik vergleicht.

Wenn zahlreiche sequenzielle Lesevorgänge erwartet werden, wird empfohlen, die Lese-Prefetch-Funktion im Cache zu aktivieren. Wenn der Prozentsatz an Cache-Treffern gering ist, ist die Anwendung eher zufällig, und die Read-Ahead-Funktion sollte deaktiviert werden. Mittlere Prozentwerte weisen eventuell auf Häufungen sequenzieller E/As hin, lassen aber nicht unbedingt auf ihren Zusammenhang mit Lese- oder Schreib-E/A schließen. Auch hier wäre dann ein Testen mit aktiviertem/deaktiviertem Read-Ahead erforderlich.

Bei der Firmware der zweiten Generation wurde die Segment-, Block- und Prefetch-Statistik neu organisiert, dargestellt in Figure 4 aus der unteren Hälfte von Figure 2.

Abbildung 4: Firmware der zweiten Generation – Aufgeteilte Leistungsdaten. Datei: stateCaptureData.txt

```
*** Performance stats ***

   Cluster Reads   Cluster Writes      Stripe Reads
         6252626          3015009           5334257
   Stripe Writes       Cache Hits    Cache Hit Blks
         2040493          4685032         737770040
    RPA Requests        RPA Width        RPA Depth
          982036          3932113         418860162
     Full Writes   Partial Writes        RMW Writes
          653386                29           328612
No Parity Writes      Fast Writes    Full Stripe WT
                0                0                0
```

4.7.4 Blockgröße

Um die beste Leistung zu erzielen, sollte die Blockgröße stets größer sein als die maximale E/A-Größe, die der Host ausführt. Wie bereits erwähnt, sollten Blöcke die Größe einer geraden Zweierpotenz haben. Die durchschnittliche Blockgröße kann anhand der gesammelten Daten ermittelt werden. Darüber hinaus werden E/As über 2 MiB als groß betrachtet und in der Statistik getrennt von kleineren E/As aufgeführt. Obwohl alle RAID-Level von der sorgfältigen Optimierung der Block- und Segmentgröße profitieren, sind RAID 5 und 6 aufgrund ihrer Paritätsberechnungen am meisten davon abhängig.

Bei Firmware der ersten Generation (siehe Figure 5) kann diese anhand der Zeile "Avg. Blocks" ermittelt werden, welche die durchschnittliche E/A-Blockgröße angibt. Bei der ersten Generation steht das Feld "Large IO" für einen 4096 Block oder eine 2 MiB Größe mit null registrierten großen Lese- oder Schreibvorgängen während der Testdauer. Jede vom Host empfangene E/A, die über die Large IO-Größe hinausgeht, wird in Teile zerlegt, die kleiner sind als der angegebene Wert unter Large IO oder diesem entsprechen. Es kommt äußerst selten vor, dass ein Host derart große E/As sendet.

Bei Firmware der zweiten Generation (siehe Figure 6), gibt "Avg. Blks" die durchschnittlich zu beobachtende E/A-Blockgröße an. In Figure 6 steht das Feld "LargeIoSize" für eine 2 MiB-Größe mit null registrierten großen Lese- oder Schreibvorgängen während der Testdauer.

Abbildung 5: Firmware der ersten Generation – Blockgröße Datei: stateCaptureData.txt

```
Virtual Disk Unit 0 Configuration
   Volume Type:          13+1 RAID 5
   User Label:           MyRAID5_1
   Block Size:           512 bytes
   Large IO:             4096 blocks
   Segment Size:         256 blocks
   Stripe Size:          3328 blocks
...
   E/A-Statistik:
                small      small      large      large                cache
                reads     writes      reads     writes      total       hits
  requests  2028332119  147699066          0          0 2176031185 1289775370
    blocks  3091968111 2518067526          0          0 1315068341 4019884678
avg blocks           4         17          0          0          0          3
   IO pct.       93.21%      6.78%     0.00%      0.00%      0.00%     59.27%
```

Abbildung 6: Firmware der zweiten Generation – Leistungsdaten der Speicherlaufwerksattribute des RAID 1-Datenträgers Datei: stateCaptureData.txt

```
Volume 0 Attributes:
     Volume Type:        RAIDVolume
     User Label:         MyRAID10_One
...
     BlockSize:          512 bytes
     LargeIoSize:        4096 blocks
...
     Perf. Stats:        Requests     Blocks  Avg. Blks  IO Percent
        Reads            67456452 5943724625         88       71.20%
        Writes           27283249 1144902648         41       28.80%
        Large Reads             0          0          0        0.00%
        Large Writes            0          0          0        0.00%
        Total            94739701 7088627273         74      100.00%
```

Zusätzlich bietet die Datei **stateCaptureData.txt** eine granularere Methode zur Ermittlung der E/A-Verteilung innerhalb von Blöcken und Segmenten. In Figure 7 und Figure 8 enthält Kasten 1 die Anzahl vollständig gelesener oder geschriebener Blöcke, während Kasten 2 die Anzahl vollständig gelesener oder geschriebener Cluster zeigt. Die Zahl der Blöcke pro E/A-Aufruf innerhalb von Lese- und Schreibvorgängen ist ebenfalls nützlich, um zu bestimmen, ob die Block- bzw. Segmenteinstellungen für die getestete Datenzugriffsstruktur optimal sind. Die Firmware der zweiten Generation spaltet im Gegensatz zur ersten Generation nicht mehr das pro-E/A-Verhältnis der Datenausgabe speziell ab. Es kann jedoch immer noch manuell berechnet werden, indem man einfach den Wert in Kasten 1 oder 2 durch den entsprechenden E/A-Aufrufwert aus Kasten 3 (siehe Figure 8) teilt.

In den meisten Fällen wird die beste Leistung erzielt, wenn das Verhältnis der Segmente und Blöcke pro E/A möglichst nah bei 1,00 liegt. Bei der Optimierung auf maximale E/As pro Sekunde gilt grundsätzlich: Ist das Verhältnis Segmente pro E/A hoch, so ist die derzeitige Segmentgröße eventuell zu gering für die Anwendung. Wird auf höchstmögliche Datentransferrate optimiert, sollte das Verhältnis Block pro E/A am besten bei 1,00 oder sogar höher liegen. Ist dieser Wert hoch, kann die Leistung gesteigert werden, indem die Anzahl an physischen Laufwerken und/oder die Segmentgröße erhöht wird.

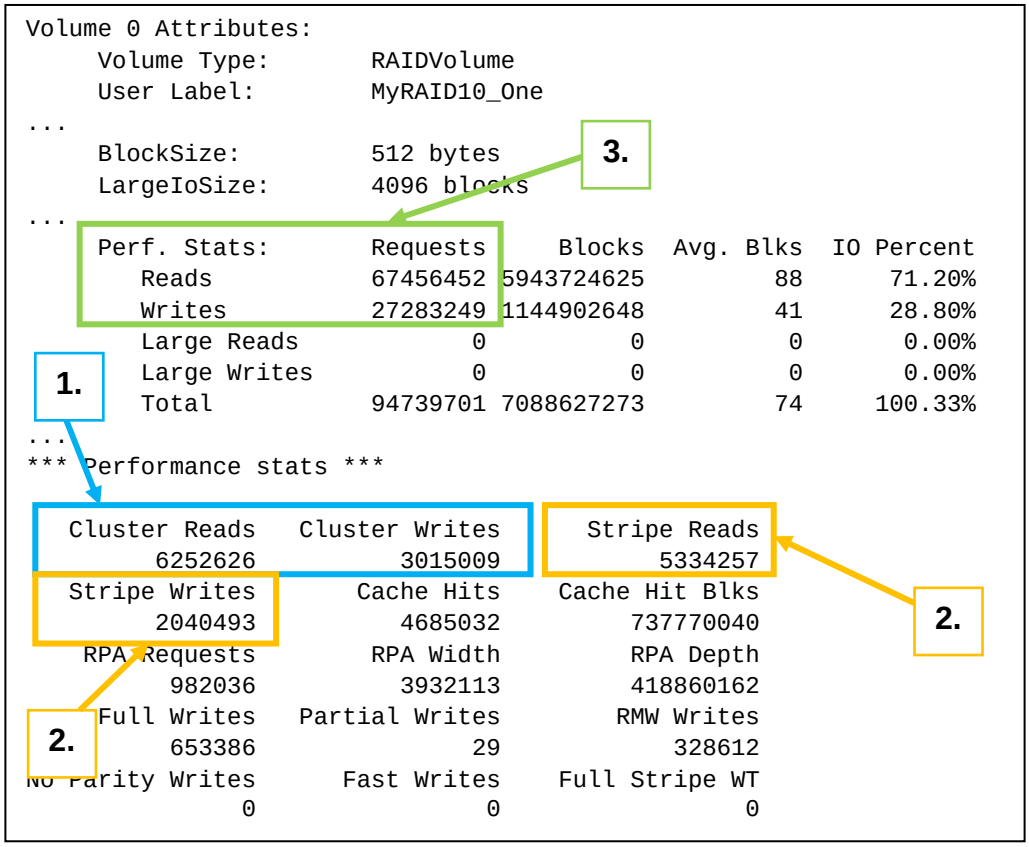
Abbildung 7: Firmware der ersten Generation – Blockverteilung Datei: stateCaptureData.txt

```
Virtual Disk Unit 0 Configuration
   Volume Type:          13+1 RAID 5
   User Label:           MyRAID5_1
   Block Size:           512 bytes
   Large IO:             4096 blocks
   Segment Size:         256 blocks
   Stripe Size:          3328 blocks
...
   E/A-Statistik:
                small      small      large      large                 cache
                reads     writes      reads     writes      total       hits
  requests  2028332119  147699066          0          0 2176031185 1289775370
    blocks  3091968111 2518067526          0          0 1315068341 4019884678
avg blocks           4         17          0          0          0          3
   IO pct.       93.21%      6.78%      0.00%      0.00%      0.00%     59.27%

                   IOs    stripes        /IO   clusters        /IO
     reads  2028332119 2034477363        1.00 2107869128        1.03
    writes   147699066  148449472        1.00  157404718        1.06

     write        Full    Partial         RMW  No Parity        RMW2        FSWT
algorithms     1105611   12598366    32120072          0           0           0
```

Abbildung 8: Firmware der zweiten Generation – Blockverteilung Datei: stateCaptureData.txt



4.7.5 Daten für Schreibalgorithmen

Es ist wichtig sich bewusst zu machen, dass die Ermittlung des geeignetsten RAID-Levels eine beängstigende Aufgabe sein kann. Man muss wissen, welche Auswirkungen der Einsatz verschiedener Algorithmen hat, um sich für ein RAID-Level entscheiden zu können. Bei der Firmware der ersten Generation hat man, wie in Figure 9 dargestellt, die Wahl zwischen Full, Partial, RMW, RMW2 und Full Stripe Write-Through. Bei der Firmware der zweiten Generation wurde RMW2 in die RMW-Statistik mit einbezogen (siehe Figure 10).

Bei "Full" (Vollständig) nimmt der Algorithmus einen gesamten Datenblock und lädt ihn auf dem Laufwerk ab, und je nach dem gewählten RAID-Level werden P oder P und Q an dieser Stelle berechnet. Dies ist der effizienteste Schreibvorgang, der ausgeführt werden kann, und das Design einer Datenträgergruppe sollte stets die maximale Anzahl vollständiger Schreibvorgänge zum Ziel haben.

Partielle Schreibvorgänge liegen vor, wenn weniger als ein vollständiger Block von Daten, die nicht mit Segmentgrenzen abgestimmt sind, verändert und geschrieben werden. Bei RAID-Level 5 und 6 ist das Verfahren komplexer, da

Paritätsdaten für den ganzen Block neu berechnet werden müssen. Partielle Schreibvorgänge sind ein "Worst-Case"-Algorithmus und sollten möglichst gering gehalten werden. Kommt es zu mehr partiellen als vollständigen Schreibvorgängen, kann dies ein Hinweis auf eine unangemessene Segmentgröße sein.

RMW (Read-Modify-Write) ist der zweitbeste Schreibalgorithmus, der für RAID 5 und 6 verfügbar ist. Zu einem RMW kommt es, wenn eine Menge an Bits abgeändert wird, die kleiner ist als ein einzelnes Segment oder diesem entspricht. Dies stellt einen Zwei-Schritt-Lesevorgang in RAID 5 und einen Drei-Schritt-Lesevorgang in RAID 6 dar, wobei eines der Segmente verändert und das Paritätslaufwerk bzw. die Paritätslaufwerke gelesen werden. Danach wird die Parität für den betroffenen Bereich neu berechnet, und die Daten und Parität in dem Block werden neu auf das Laufwerk geschrieben. Bei der Verarbeitung kleiner Transaktionen ist mit einer sehr hohen Anzahl an RMWs zu rechnen. Diese RMW-Schreibvorgänge können zu einem erheblichen Leistungsverlust führen. Es ist jedoch möglich, die Auswirkungen durch die richtige Optimierung der Blockgrößen auf dem virtuellen Laufwerk zu verringern. .

RMW2 wird dazu genutzt, zwischen so genannten Write-to-Cache-RMWs und Write-Through-RMWs zu differenzieren, wobei sich RMW2 auf Letzteres bezieht. Diese statistischen Daten wurden bei der Firmware der zweiten Generation konsolidiert. RMW2-Vorgänge ereignen sich auch insbesondere, wenn der Cache gewaltsam deaktiviert wird oder bei Versagen des Spiegelungscontrollers (falls aktiv) bzw. Versagen des Cache-Akkus. Darüber hinaus verfolgt die Firmware der zweiten Generation Write-Through-Bedingungen ganzer Blöcke, und beide Generationen verfolgen Daten zur Anzahl an Paritätsblöcken, die neu berechnet werden.

Abbildung 9: Firmware der ersten Generation – Schreibalgorithmen

```
Virtual Disk Unit 0 Configuration
   Volume Type:          13+1 RAID 5
   User Label:           MyRAID5_1
   Block Size:           512 bytes
   Large IO:             4096 blocks
   Segment Size:         256 blocks
   Stripe Size:          3328 blocks
...
   E/A-Statistik:
                small      small      large      large                cache
                reads     writes      reads     writes      total       hits
  requests  2028332119  147699066          0          0 2176031185 1289775370
    blocks  3091968111 2518067526          0          0 1315068341 4019884678
avg blocks           4         17          0          0          0          3
   IO pct.       93.21%      6.78%      0.00%      0.00%      0.00%     59.27%

                  IOs    stripes        /IO   clusters        /IO
     reads  2028332119 2034477363       1.00 2107869128       1.03
    writes   147699066  148449472       1.00  157404718       1.06

     write        Full    Partial        RMW  No Parity       RMW2       FSWT
algorithms     1105611   12598366   32120072          0          0          0
```

Abbildung 10: Firmware der zweiten Generation – Schreibalgorithmen

```
Volume 1 Attributes:
     Volume Type:       RAIDVolume
     User Label:        MyRAID5vd
...
     BlockSize:         512 bytes
     LargeIoSize:       4096 blocks
...
     Perf. Stats:       Requests     Blocks  Avg. Blks  IO Percent
        Reads          577761063 1155647889          2      87.50%
        Writes          82542765  175687877          2      12.50%
        Large Reads            0          0          0       0.00%
        Large Writes           0          0          0       0.00%
        Total          660303828 1331335766          2     100.00%
...
*** Performance stats ***

   Cluster Reads   Cluster Writes      Stripe Reads
         6252626          3015009           5334257
   Stripe Writes       Cache Hits    Cache Hit Blks
         2040493          4685032         737770040
    RPA Requests        RPA Width         RPA Depth
          982036          3932113         418860162
     Full Writes   Partial Writes        RMW Writes
          653386              29            328612
No Parity Writes      Fast Writes    Full Stripe WT
               0                0                 0
```

4.8 Verwenden des CLI Performance Monitor

Der CLI Performance Monitor ist ein befehlszeilenbasiertes Scripting-Dienstprogramm, das Zugriff auf alle Funktionen des PowerVault-Storage-Arrays und einige zusätzliche Leistungsstatistiken bietet. In Microsoft Windows heißt das Scripting-Dienstprogramm `smcli.exe` und befindet sich standardmäßig im Verzeichnis `c:\Programme\Dell\MD Storage Manager\client`. Das komplette *CLI-Handbuch* für MD3000/3000i befindet sich auf der Support-Website von Dell unter http://support.dell.com/manuals.

Es folgen einige Beispiele für CLI-Befehle.

Der Befehl zur Ausführung des **CLI Performance Monitor** lautet:

SMcli -n ArrayName -c "set session performanceMonitorInterval=5 performanceMonitorIterations=30;save storageArray performanceStats file=\“performance.csv\";“

Eine vollständige Liste der Befehle und Anleitungen zur Verwendung des CLI Performance Monitor finden Sie im *Dell™ PowerVault™-Handbuch für den Modular Disk Storage Manager und CLI* auf http://support.dell.com/manuals.

4.9 Weitere Array-Gesichtspunkte

4.9.1 Globale Medienscanrate

Die Einstellungen für den Medienscan werden in MDSM auf der Registerkarte **Tools** vorgenommen bzw. geändert. Der globale Medienscan nutzt CPU-Zyklen und beeinträchtigt die Leistung, wenn er zu unangemessener Zeit ausgeführt wird, zum Beispiel in Phasen mit hohem Benutzerzugriff oder während Datensicherungen.

Hinweis: *Dell™ empfiehlt nicht die Deaktivierung des Medienscans oder die Herabsetzung des Medienscanintervalls auf unter 15 Tage. Durch Deaktivierung des Medienscans kann das Risiko unvorhergesehener Ausfälle steigen.*

4.9.2 Einstellen des für virtuelle Laufwerke spezifischen Medienscans

Die Einstellungen für den Medienscan werden in MDSM auf der Registerkarte "Tools" vorgenommen bzw. geändert. Um den Medienscan auf bestimmten virtuellen Laufwerken auszuführen, markieren Sie das virtuelle Laufwerk, das Sie scannen möchten, und aktivieren Sie das Kontrollkästchen **Scan selected virtual disks** (Ausgewählte virtuelle Laufwerke scannen).

4.10 Leistung für Zusatzfunktionen

4.10.1 Erzielen der optimalen Leistung für Snapshots

Wenn Sie Snapshot-Repositorys verteilen, ordnen Sie die virtuellen Repository-Laufwerke auf Datenträgern an, die von den virtuellen Produktions-Laufwerken getrennt sind, um die Repository-Schreibvorgänge zu isolieren und die Copy-on-Write-Einbußen zu minimieren. Verlegen Sie, wenn möglich, Lese-E/As auf das virtuelle Snapshot-Laufwerk auf Nebenzeiten, wenn weniger E/A-Aktivitäten auf dem virtuellen Quell-Laufwerk ausgeführt werden, z. B. in den Abendstunden.

4.10.2 Erzielen der optimalen Leistung für virtuelle Datenträgerkopien

Bei der Zusatzfunktion Virtual Disk Copy werden optimierte große Blöcke eingesetzt, um die Kopie so schnell wie möglich durchzuführen. Demzufolge erfordert diese Funktion wenig Optimierung, außer dass die Kopierpriorität auf die höchste Stufe gesetzt wird, die noch eine akzeptable Host-E/A-Leistung ermöglicht. Auf die Leistung für virtuelle Datenträgerkopien wirken sich andere Controller-Aktivitäten, das RAID-Level und Laufwerksparameter des virtuellen Quell-Datenträgers und des virtuellen Ziel-Datenträgers aus. Ein optimales Verfahren zur Nutzung von Virtual Disk Copy besteht darin, alle virtuellen Snapshot-Datenträger zu deaktivieren, die einem virtuellen Quell-Datenträger zugeordnet sind, bevor man den virtuellen Quell-Datenträger als Zielvolumen für die virtuelle Datenträgerkopie auswählt. Der virtuelle Ziel- und der virtuelle Quell-Datenträger sollten sich im Idealfall in separaten Datenträgergruppen befinden, sofern möglich. Belässt man sie in der gleichen Datenträgergruppe, steigt das Risiko leistungsschwächerer zufälliger E/As beim Kopiervorgang.

5 Überlegungen zu Host-Servern

5.1 Hardwareplattform des Hosts

5.1.1 Überlegungen zur Serverhardware-Architektur

Die verfügbare Bandbreite hängt von der Serverhardware ab. Mit der Anzahl von Bussen kann die gesamte Bandbreite erhöht werden, aber die Anzahl der HBAs, die einen einzelnen Bus gemeinsam nutzen, kann die Bandbreite drosseln. Darüber hinaus hat manche Serverhardware langsamere PCI-E-Anschlüsse (4x) sowie Hochgeschwindigkeitsanschlüsse (8x). Die Dell SAS5e HBAs sind 8x-PCI-E-Geräte und sollten in 8x-Steckplätzen installiert werden, um die maximale Leistung zu erzielen. Wo zusätzliche PCI-E-Steckplätze verfügbar sind, sollten zwei SAS HBAs verwendet werden, um den E/A-Host redundant mit den einzelnen Controllermodulen des Storage-Arrays zu verbinden und damit sowohl Leistung als auch Redundanz zu maximieren.

***Hinweis**: Dell™ stellt auf der Abdeckung aller Server ein Bus-Layout dar. Ziehen Sie dieses Diagramm heran, und verwenden Sie für jeden im Host installierten HBA einen anderen Bus.*

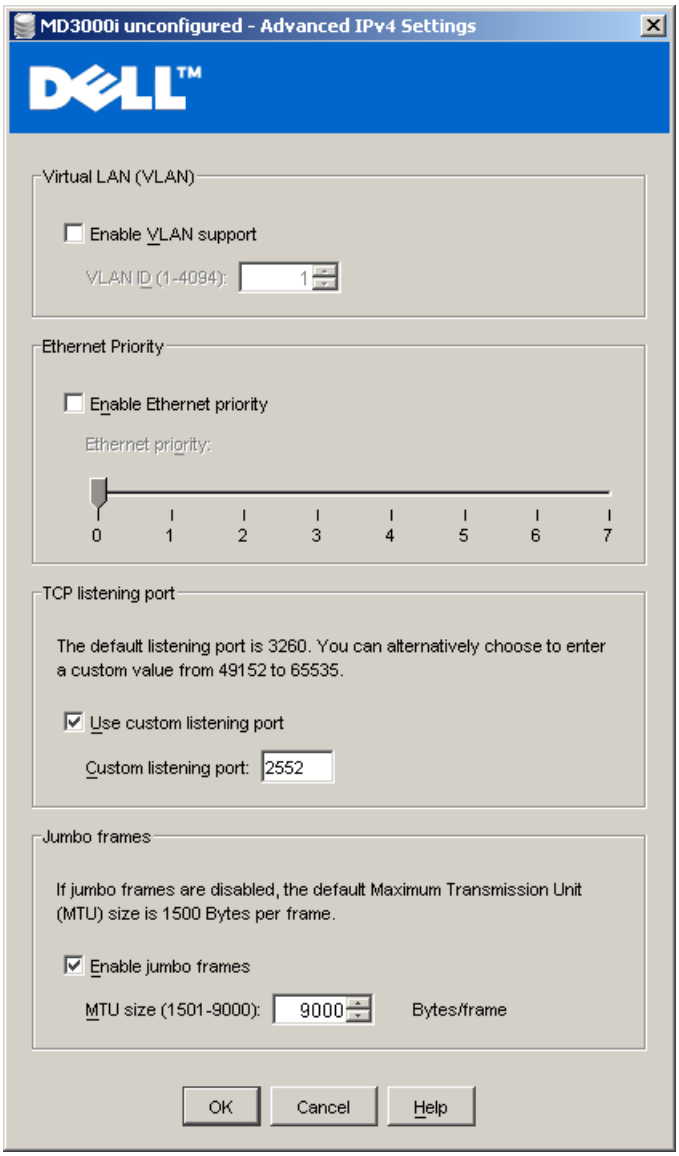
5.1.2 Gemeinsame Nutzung von Bandbreite auf dem Dell™ MD3000i mit mehreren Netzwerkkarten

Beachten Sie folgende Punkte, wenn Sie Bandbreite auf dem MD3000i mit mehreren Netzwerkkarten nutzen.

- Weisen Sie jeder Netzwerkkarte eine eigene IP-Adresse zu.
- Verbinden Sie jede Netzwerkkarte mit einem separaten Switch, oder verbinden Sie Host-Netzwerkkarten direkt mit den Ziel-iSCSI-Anschlüssen.
- Verwenden Sie separate Netzwerkkarten für den Zugriff auf das öffentliche Netzwerk und iSCSI-Verkehr des Storage-Arrays. Installieren Sie zusätzliche Netzwerkkarten je nach Bedarf.
- Richten Sie separate redundante Netzwerke dediziert für den iSCSI-Datenverkehr ein. Wenn dies nicht möglich ist, richten Sie ein separates VLAN für den iSCSI-Traffic ein.
- Nutzen Sie Jumbo-Frames. (Mit Jumbo-Frames erhöht sich die TCP-Framegröße von 1500 Byte auf 9000 Byte.)
- Der Microsoft iSCSI-Initiator funktioniert *nicht* mit zusammengeschalteten Netzwerkkarten.
- Auf einem einzelnen Host können nicht HBAs und Netzwerkkarten kombiniert werden, um eine Verbindung zum selben oder zu unterschiedlichen Arrays aufzubauen.

Um die Einstellungen der Jumbo-Frames im MDSM zu ändern, wählen Sie die Registerkarte **iSCSI, Configure iSCSI Host Ports (iSCSI-Host-Ports konfigurieren), Advanced (Erweitert)** (siehe Figure 11). Jumbo-Frames können außerdem auch über die CLI eingestellt werden. Wenn Sie ein MD3000i mit Jumbo-Frames verwenden, müssen auch auf den Ethernet-Switches und Host-Netzwerkkarten Jumbo-Frames aktiviert und auf einen entsprechenden Wert eingestellt sein.

Abbildung 11: Erweiterte IPv4-Einstellungen für das MD3000i: Support von VLAN, QOS, Jumbo-Frames



5.1.3 Gemeinsame Nutzung von Bandbreite mit mehreren SAS HBAs

Jeder SAS-Anschluss integriert vier Vollduplex-Verbindungen in einem einzigen Anschluss. Die einzelnen SAS 1.1-Verbindungen haben eine Höchstgeschwindigkeit von 3 Gbit/s. Eine einzige Übertragungsstrecke wird für die Verbindung mit den Laufwerken verwendet. Die zweite, dritte und vierte Strecke dient als Überlauf, wenn gleichzeitig ausgeführte E/As den primären Kanal überlasten. Wenn die erste Verbindung beispielsweise Daten bei einer Geschwindigkeit von 3 Gbit/s überträgt, verwendet SAS 10-Bit-Verschlüsselung gegenüber 8-Bit für Byte-Übertragung, sodass jede einzelne 3 Gbit/s-Verbindung auf 300 MiB/s beschränkt ist. Wenn nun

beispielsweise ein anderer Datenblock auf Festplatte übertragen werden muss und Verbindung 1 noch ausgelastet ist, übernimmt Verbindung 2 den Datenüberlauf, der nicht über Verbindung 1 übertragen werden kann. Wenn Verbindung 1 die Datenübertragung beendet hat, wird der nächste Datenblock wieder über Verbindung 1 übertragen, andernfalls wird eine andere Verbindung genutzt. Auf diese Weise ist es bei starker E/A-Auslastung möglich, dass in bestimmten Zeiten alle Verbindungen bei Bereitstellung einer simultanen Datengeschwindigkeit von bis zu 12 Gbit/s genutzt werden. Bitte beachten Sie, dass diese Rohgeschwindigkeit nicht die Übertragungskosten bzw. die operativen Grenzen der Geräte auf beiden Seiten der SAS-Verbindung berücksichtigt und lediglich einen gecachten E/A-Vorgang darstellt.

Darüber hinaus müssen die Busse innerhalb des Hosts sorgfältig ausgewählt werden. Durch die Installation von HBAs, die denselben Bus verwenden, wird die Datenübertragungsrate beeinträchtigt. Stellen Sie sicher, dass alle HBAs, die im Hostsystem installiert sind, einen anderen Bus nutzen (siehe Hinweis in 5.1)

5.2 Überlegungen zur Systemsoftware

5.2.1 Pufferung der E/A

Der E/A-Typ, gepuffert oder nicht gepuffert, der der Anwendung vom Betriebssystem zur Verfügung gestellt wird, spielt bei der Analyse der Speicherleistung eine wichtige Rolle.
Ungepufferte E/A-Bausteine (auch als *Raw oder* Direct E/A bezeichnet) verschieben Daten direkt zwischen der Anwendung und den Laufwerken. Gepufferte E/A-Bausteine sind ein Dienst, der vom Betriebssystem oder vom Dateisystem bereitgestellt wird. Durch die Pufferung wird die Anwendungsleistung verbessert, indem Schreibdaten in einem Puffer des Dateisystems zwischengespeichert werden, der vom Betriebs- oder Dateisystem regelmäßig in den nicht volatilen Speicher entleert wird.

Gepufferte E/A wird im Allgemeinen für kürzere und häufigere Übertragungen bevorzugt. Durch die Dateisystempufferung ändern sich möglicherweise die E/A-Strukturen, die von der Anwendung erzeugt werden. Das heißt, Schreibvorgänge können zusammenfließen, sodass die Struktur, die vom Speichersystem erkannt wird, sequenzieller und schreibintensiver ist als der Anwendungs-E/A-Baustein selbst. Direct E/A wird für umfangreichere, weniger häufige Übertragungen bevorzugt und für Anwendungen, die eine eigene umfassende Pufferung vorsehen, zum Beispiel Oracle. Das E/A-Leistungs-Tool Iometer ist eine weitere Anwendung, die ungepuffert genutzt werden kann, um eine direktere Leistung zu testen. Unabhängig vom E/A-Typ lässt sich die E/A-Leistung grundsätzlich verbessern, wenn das Speichersystem mit einem konstanten Strom von E/A-Aufrufen aus der Host-Anwendung ausgelastet wird. Machen Sie sich mit den Parametern vertraut, die das Betriebssystem zur Steuerung der E/A-Bausteine vorsieht, zum Beispiel die maximale Übertragungsgröße.

5.2.2 Abstimmen der Host-E/A mit RAID-Striping

Die meisten Host-Betriebssysteme verlangen oder profitieren von unterschiedlichen Graden der Abstimmung von E/A-Partitionen und der Vermeidung leistungssenkender Segmentübergänge. Das heißt, E/As sollten Segmentgrenzen nicht überschreiten. Durch das Anpassen der E/A-Größe (i. d. R. über eine Zweierpotenz) an das Layout der Datenträgergruppe können überall im Laufwerk E/A-Bausteine abgestimmt werden. Dies trifft allerdings nur dann zu, wenn der Startsektor sachgemäß mit einer Segmentgrenze abgestimmt ist. Segmentübergänge sind oft im Microsoft Windows-Betriebssystem zu beobachten, wo Partitionen, die durch Microsoft Windows 2000 oder Microsoft Windows 2003 erstellt wurden, bei Sektor 64 beginnen. Der Start bei Sektor 64 führt zu einer mangelnden Abstimmung mit dem zugrunde liegenden RAID-Striping und ermöglicht es, dass sich ein einzelner E/A-Vorgang über mehrere Segmente erstreckt. Neuere Versionen des Microsoft Windows-Betriebssystems haben eine Standardabstimmung von typischerweise 2048 Blöcken, die je nach Umgebung unter Umständen noch geändert werden muss.

Microsoft stellt hierfür das Dienstprogramm diskpar.exe als Teil des Windows 2000 Resource Kits zur Verfügung, das in Microsoft Windows 2003 und späteren Versionen in diskpart.exe umbenannt wurde. Der KB-Artikel 929491 von Microsoft deckt dieses Thema ab. Dell™ empfiehlt stets, die ordnungsgemäße Partitionsabstimmung auf die Blockgröße zugewiesener virtueller Laufwerke zu überprüfen. Mithilfe dieser Dienstprogramme kann der Startsektor im Master Boot Record (MBR) auf einen Wert gesetzt werden, der die Sektorabstimmung für alle E/As sicherstellt. Verwenden Sie als Wert ein Vielfaches von 64, z. B. 64 oder 128. Anwendungen wie Microsoft Exchange verlassen sich auf die korrekte Abstimmung von Partitionen auf die Blockgrenze des Laufwerks.

Nähere Informationen von Microsoft zur Nutzung von diskpart.exe finden Sie unter http://technet.microsoft.com/en-us/library/aa995867.aspx.

***ACHTUNG**: Durch Änderungen an der Abstimmung bestehender Partitionen werden Daten **zerstört**.*

Anhang A: Zusätzliche Leistungs-Tools

Table 3 enthält eine Anzahl weit verbreiteter Tools, Benchmarks und Dienstprogramme. Einige dieser Tools werden von nicht-kommerziellen Organisationen entwickelt und sind kostenlos.

Tabelle 3: Leistungs-Tools

Name	Beschreibung	Plattform	Erhältlich von
IOBench	Benchmark für E/A-Datenrate und feste Auslastung	Unix/Linux	http://www.acnc.com/benchmarks.html
IOmeter	Tool zur Messung und Charakterisierung von E/A-Subsystemen	Windows, Unix/Linux	http://www.iometer.org
IOZone	Benchmark-Tool für Dateisystem	Windows, Unix/Linux	http://www.iozone.org
Xdd	Tool für Messung und Charakterisierung der Subsystem-E/A	Windows, Unix/Linux	http://www.ioperformance.com
FIO	Benchmarking- und E/A-Tool für Unix/Linux	Unix/Linux	http://freshmeat.net/projects/fio/
Bonnie++	E/A-Benchmark-Suite für Unix/Linux-Dateisysteme	Unix/Linux	http://www.coker.com.au/bonnie++/

Anhang B: Fehlerbehebung der Systeme

Informationen zur Fehlerbehebung bei den MD3000- und MD3000i-Storage-Arrays finden Sie im Kapitel über die Fehlerbehebung im *Dell™ PowerVault™ - Handbuch für den Modular Disk Storage Manager*. Weitere Infos finden Sie hier:

MD3000:

http://support.dell.com/support/edocs/systems/md3000/en/index.htm

MD3000i:

http://support.dell.com/support/edocs/systems/md3000i/en/index.htm

Anhang C: Referenzunterlagen

Dell™ PowerVault™-Handbuch für den Modular Disk Storage Manager und CLI,
http://support.dell.com/support/edocs/systems/md3000/en/index.htm

Dell™ PowerVault™MD3000,
http://support.dell.com/support/edocs/systems/md3000/en/index.htm

PowerVault MD3000i-SAN-Array für Speicherkonsolidierung,
http://www.dell.com/content/products/productdetails.aspx/pvaul_md3000i?c=us&1=de&s=bsd&cs=04

Verwenden von iSCSI: Dell™ PowerVault™-Handbuch für den Modular Disk Storage Manager und CLI,
http://support.dell.com/support/edocs/systems/md3000/en/UG/HTML/iscsi.htm

Dell-iSCSI-Clusterinformationen,
http://www.dell.com/content/topics/global.aspx/sitelets/solutions/cluster_grid/clustering_ha?c=us&cs=555&l=en&s=biz&~page=3&~tab=4

Microsoft, Verwenden von Diskpart,

http://technet.microsoft.com/en-us/library/aa995867.aspx

Microsoft, KB929491, Richtige Blockabstimmung für NTFS mithilfe von Diskpart,

http://support.microsoft.com/kb/929491

Microsoft, *Optimieren von Speichersystemen für Exchange Server 2003,*
http://www.microsoft.com/technet/prodtechnol/exchange/2003/library/optimizestorage.mspx

Hochverfügbarkeits-Clustering von Dell: iSCSI

http://www.dell.com/content/topics/global.aspx/sitelets/solutions/cluster_grid/clustering_ha?&~page=3&~tab=4

IEC 60027-2 Ed. 2.0 (2000-11): SI Prefixes for Binary multiples

http://www.iec.ch/zone/si/si_bytes.htm

Anhang D: Glossar

Begriff	Definition
Burstiness	Eine Eigenschaft von Datenverkehr, die als Verhältnis der höchsten E/A-Rate zur durchschnittlichen E/A-Rate definiert wird. In diesem Fall geht es um den durchschnittlichen Tastgrad der E/A, die von einem Storage-Array übertragen oder empfangen wird. Der Begriff "Burstiness" wurde aus seinem üblichen Gebrauch zur Beschreibung von Netzwerkauslastungen entlehnt.
Controller-Auslastung	Controller-Auslastung tritt ein, wenn ein RAID-Controller-Modul seine maximale Betriebsauslastung erreicht hat und keine weiteren Vorgänge in seiner verfügbaren Bandbreite ausführen kann. Weitere Vorgänge jenseits dieses Spitzenwerts gehen nicht verloren, sondern werden vorübergehend in eine Warteschlange gestellt. Dies wird als Inflektionspunkt betrachtet, an dem die Leistung eine feste Höchstgrenze erreicht.
GiB	Gibibyte; siehe Kibibyte.
HBA	Host-Bus-Adapter
FESTPLATTENLAUFWERK	Festplattenlaufwerk
IOPS	Input/Output Operations Per Second; Maßeinheit im IT-Benchmarking, die die E/A-Rate quantifiziert.
Interposer	Integrierte Schaltung mit einer Schnittstelle für das Routing zwischen zwei entgegengesetzten signalgebenden Seiten. Wird im MD3000/MD3000i als Schnittstelle zwischen SATA-Festplatten und SAS-Rückwandplatine verwendet, um Übersetzung zu leisten.
iSCSI	Internet Small Computer Systems Interface. Das iSCSI-Protokoll wird von der IETF in RFC 3720 definiert.
KiB	Siehe Kibibyte.
Kibibyte	Kilo binary byte; IEC SI-Einheit, um 1024 bzw. 2^{10} Byte (Basis: 2) klar von Kilobyte auf Zehnerbasis (10^3 oder 1000 Byte) zu unterscheiden. Siehe IEC/ISO ISO 80000, IEC 60027-2 oder IEEE 1541-2002.
Lange E/A	Jegliche E/A-Häufung, die 1/3 der verfügbaren Cache-Speichergröße übersteigt und ein erhöhtes Risiko mit sich bringt, nicht komplett im Cache verarbeitet werden zu können.
MD3000	Dell™ PowerVault MD3000 Expandable Storage Array mit SAS-Front-End.
MD3000i	Dell™ PowerVault MD3000i Expandable Storage Array

Begriff	Definition
	mit iSCSI-Front-End.
MDSM	Dell™ Modular Disk Storage Manager. Suite von Hostverwaltungsdienstprogrammen zur Konfiguration und Wartung eines MD3000-/MD3000i-Storage-Arrays.
MiB	Mebibyte, siehe Kibibyte.
NIC	Netzwerkschnittstellen-Controller
NL-SAS	Near-Line-SAS; eine Hybridtechnologie von kapazitätsreicheren SATA-Festplatten mit einer speziellen SAS-Controllerplatine direkt auf dem Laufwerk, anstatt eine SATA Interposer Card zu verwenden.
RAID	Redundant Array of Independent Disks
RAID 0	RAID-Level 0; RAID 0 ist eine Blockgruppe ohne redundante Informationen. Es handelt sich dabei grundsätzlich um ein voll degradiertes RAID-Set ohne Festplattenredundanzkosten.
RAID 1/10	RAID-Level 1/10: Die RAID 1/10-Implementierung auf dem MD3000/MD3000i folgt Berkley RAID 1-Standard und erweitert diesen zu einem redundanten, gespiegelten N+N-Set. Funktional betrachtet entspricht diese Implementierung einem generischen Nested RAID 1+0 und kommt mit nur zwei physischen Laufwerken aus. Dadurch kann eine beliebige Anzahl an Laufwerken ausfallen, solange auch nur ein Laufwerkspaar bei Laufwerkskosten in Höhe der Hälfte der vorhandenen physischen Laufwerke zur Verfügung steht.
RAID 5	RAID-Level 5; mit einem Block-Striping-Algorithmus, bei dem n-1 Laufwerke pro Block im RAID-Set die Daten und ein Paritätslaufwerk P die Parität oder Checksumme enthält, die zur Validierung der Datenintegrität und Bereitstellung von Konsistenzinformationen verwendet wird. Die Parität wird auf alle Laufwerke in einer Datenträgergruppe verteilt, um zusätzliche Fehlertoleranz zu schaffen. RAID 5 bietet Schutz beim Ausfall eines Laufwerks.
RAID 6	RAID-Level 6; mit einem Block-Striping-Algorithmus, bei dem n-2 Laufwerke pro Block im RAID-Set die Daten und Paritätsblöcke P und Q die Parität oder Checksumme enthalten, die zur Validierung der Datenintegrität und Bereitstellung von Konsistenzinformationen verwendet wird. Die Parität wird auf alle Laufwerke in einer Datenträgergruppe verteilt, um zusätzliche Fehlertoleranz zu schaffen. RAID 6 bietet Schutz beim Ausfall von zwei

Begriff	Definition
	Laufwerken.
RPA	Read Prefetch Algorithm: Eine Kurzform für den Read-Ahead-Cache, der auf dem MD3000\MD3000i verwendet wird.
RMW	Read, Modify, Write; der zweitbeste Algorithmus, der bei RAID 5- und RAID 6-Schreibvorgängen zur Verfügung steht. Zu einem RMW kommt es, wenn eine Menge an Bits abgeändert wird, die kleiner ist als ein einzelnes Segment oder diesem entspricht.
RMW2	Eine Anpassung von RMW in Firmware der ersten Generation speziell für Write-Through-Bedingungen, bei der Schreib-Cache nicht aktiviert ist oder auf ein virtuelles Laufwerk aufgerufen wird.
SAS	Serial Attached SCSI Protokoll wird von t10.org geführt
SATA	Serial Attached Technology Attachment bzw. Serial ATA. Dabei handelt es sich um die nächste Phase des veralteten Parallel ATA. In diesem White Paper bezieht sich SATA in erster Linie auf SATA-Festplattentechnologie.
Auslastung	Siehe Controller-Auslastung
SCSI	Small Computer System Interface; Protokoll wird von t10.org geführt
Segment	Ein Segment sind die Daten, die auf ein Laufwerk im Block einer virtuellen Datenträgergruppe geschrieben werden, bevor Daten auf das nächste Laufwerk im Block der virtuellen Datenträgergruppe geschrieben werden.
Kurze E/A	Jegliche E/A, die weniger als 1/3 des verfügbaren Cache-Speichers beansprucht und im Cache verarbeitet werden kann, also ein gecachter Vorgang.
SQL	Structured Query Language; flexible Auszeichnungssprache für Computerdatenbanken auf ANSI- und ISO-Basis.
Block	Eine Gruppe zusammenhängender Segmente, die mehrere Laufwerke umfasst.