(19) **Bundesrepublik Deutschland**
Deutsches Patent- und Markenamt

(10) **DE 695 33 575 T2** 2006.01.05

(12) # Übersetzung der europäischen Patentschrift

(97) **EP 0 715 249 B1**
(21) Deutsches Aktenzeichen: **695 33 575.8**
(96) Europäisches Aktenzeichen: **95 110 828.1**
(96) Europäischer Anmeldetag: **11.07.1995**
(97) Erstveröffentlichung durch das EPA: **05.06.1996**
(97) Veröffentlichungstag der Patenterteilung beim EPA: **29.09.2004**
(47) Veröffentlichungstag im Patentblatt: **05.01.2006**

(51) Int Cl.[8]: ***G06F 3/06*** (2006.01)
G06F 11/10 (2006.01)
G06F 11/20 (2006.01)
G11B 20/18 (2006.01)

(30) Unionspriorität:
348391 **02.12.1994** **US**

(84) Benannte Vertragsstaaten:
DE, FR, GB

(73) Patentinhaber:
Hewlett-Packard Development Co., L.P., Houston, Tex., US

(72) Erfinder:
Burkes, Theresa A., Meridian, US; Diamond, Bryan M., Boise, US; Nelson, Marvin D., Boise, US

(74) Vertreter:
Schoppe, Zimmermann, Stöckeler & Zinkler, 82049 Pullach

(54) Bezeichnung: **Verfahren der und System zur Speicherplatzreservierung für Datenmigration in einem hierarchischen Datenspeicherungssystem durch dynamische Berechnung maximalen Speicherplatzes für Spiegelredundanz**

Anmerkung: Innerhalb von neun Monaten nach der Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents kann jedermann beim Europäischen Patentamt gegen das erteilte europäische Patent Einspruch einlegen. Der Einspruch ist schriftlich einzureichen und zu begründen. Er gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist (Art. 99 (1) Europäisches Patentübereinkommen).

Die Übersetzung ist gemäß Artikel II § 3 Abs. 1 IntPatÜG 1991 vom Patentinhaber eingereicht worden. Sie wurde vom Deutschen Patent- und Markenamt inhaltlich nicht geprüft.

Beschreibung

GEBIET DER ERFINDUNG

[0001] Diese Erfindung bezieht sich auf Verfahren zum Reservieren eines Speicherraums für eine Datenmigration in einem redundanten hierarchischen Datenspeichersystem. Diese Erfindung bezieht sich ferner auf ein hierarchisches Datenspeichersystem selbst.

[0002] Computersysteme werden in Bezug auf Geschwindigkeit, Zuverlässigkeit und Verarbeitungsfähigkeit ständig verbessert. Folglich sind Computer in der Lage, komplexere und anspruchsvollere Anwendungen zu handhaben. Während jedoch Computer immer besser werden, nehmen auch die Leistungsanforderungen zu, die an Massenspeicher- und Eingabe-/Ausgabevorrichtungen (I-/O-Vorrichtungen) gestellt werden. Es besteht weiterhin ein Bedarf, Massenspeichersysteme zu entwerfen, die in Bezug auf die Leistungsfähigkeit mit sich weiter entwickelnden Computersystemen Schritt halten.

[0003] Diese Erfindung betrifft insbesondere die Massenspeichersysteme des Plattenarray-Typs. Plattenarraydatenspeichersysteme weisen Mehrfach-Speicherplattenlaufwerksvorrichtungen auf, die angeordnet und koordiniert sind, um ein einziges Massenspeichersystem zu bilden. Für derartige Speichersysteme gibt es drei Hauptentwurfskriterien: Kosten, Leistungsfähigkeit und Verfügbarkeit. Am wünschenswertesten ist es, Speichervorrichtungen zu erzeugen, die geringe Kosten pro Megabyte, eine hohe Eingangs-/Ausgangsleistungsfähigkeit und eine hohe Datenverfügbarkeit aufweisen. „Verfügbarkeit" ist die Fähigkeit, auf Daten, die in dem Speichersystem gespeichert sind, zuzugreifen, und die Fähigkeit, im Fall eines Ausfalls einen fortgesetzten Betrieb zu gewährleisten. In der Regel wird eine Datenverfügbarkeit durch die Verwendung einer Redundanz geliefert, wobei Daten, oder Beziehungen zwischen Daten, an mehreren Plätzen gespeichert sind. Für den Fall, dass eine Speicherplatte in dem Plattenarray teilweise oder vollständig ausfällt, können die Benutzerdaten über die auf den verbleibenden Platten gespeicherten redundanten Daten rekonstruiert werden.

[0004] Es gibt zwei übliche Verfahren zum Speichern von redundanten Daten. Gemäß dem ersten bzw. „Spiegel"-Verfahren werden Daten dupliziert und in zwei getrennten Bereichen des Speichersystems gespeichert. Bei einem Plattenarray werden die identischen Daten beispielsweise auf zwei getrennten Platten in dem Plattenarray bereitgestellt. Das Spiegelverfahren weist die Vorteile einer hohen Leistungsfähigkeit und einer hohen Datenverfügbarkeit aufgrund der Duplexspeichertechnik auf. Das Spiegelverfahren ist jedoch auch relativ teuer, da es die Kosten des Speicherns von Daten effektiv verdoppelt.

[0005] Bei dem zweiten bzw. „Paritäts"-Verfahren wird ein Teil des Speicherbereichs verwendet, um redundante Daten zu speichern, die Größe des redundanten Speicherbereichs ist jedoch geringer als der zum Speichern der ursprünglichen Daten verwendete verbleibende Speicherbereich. Bei einem Plattenarray, das fünf Platten aufweist, könnten beispielsweise vier Platten verwendet werden, um Daten zu speichern, wobei die fünfte Platte für das Speichern von redundanten Daten reserviert ist. Das Paritätsverfahren ist vorteilhaft, da es kostengünstiger ist als das Spiegelverfahren, im Vergleich zu dem Spiegelverfahren weist es jedoch auch geringere Leistungsfähigkeits- und Verfügbarkeitscharakteristika auf.

[0006] Die US-Patentschrift 5,166,939 beschreibt eine Massenspeichervorrichtung, die aus einer Mehrzahl von physischen Speichervorrichtungen besteht. Der Vorrichtungssatz ist in eine oder mehrere Redundanzgruppen unterteilt. Jede Redundanzgruppe ist in eine oder mehr Datengruppen unterteilt, von denen jede lediglich eine geringe Anzahl der Laufwerke in der Redundanzgruppe aufweisen kann, wodurch eine hohe Anforderungsrate geliefert wird, oder von denen jede eine große Anzahl von Laufwerken aufweisen kann, wodurch eine hohe Bandbreite bereitgestellt wird. Es ist eine dynamische Konfiguration vorgesehen, um die Konfiguration des Arrays von Speichervorrichtungen von seiner anfänglichen Konfiguration zu verändern, wodurch ermöglicht wird, dass eine neue Datengruppe zu logischen Einheiten hinzugefügt wird, ohne dass es erforderlich wäre, dass die logischen Einheiten entkoppelt werden.

[0007] Das IBM Technical Disclosure Bulletin, Band 33, Nr. 8, Januar 1991, Seiten 270–272, „USING DUAL ACTUATOR SHARED DATA DIRECT ACCESS STORAGE DEVICES IN A REDUNDANT ARRAY", offenbart eine Technik zum Verwenden von DASDs (Dual Actuator Shared Data Direct Access Storage Devices) in einem Plattenarray. Es ist ein Paar von gespiegelten Plattenlaufwerken mit dualen Betätigungsgliedern vorgesehen, die dieselben Daten gemeinsam verwenden. Ferner ist ein paralleles Array von Plattenlaufwerken vorgesehen, die entweder eine vereinfachte Prüfsumme, Parität oder eine andere Form einer Arraytechnologie implementieren. Die meisten Daten werden unter Verwendung einer Rufdatenaufzeichnungsprozedur in das gespiegelte Paar geschrieben. Daten werden über einen gewissen Zeitraum an dem gespiegelten Paar zu-

rückgehalten, bis die gespiegelten Laufwerke beginnen, sich zu füllen. Zu diesem Zeitpunkt werden die ältesten und am wenigsten genutzten Daten zu dem parallelen Array bewegt.

[0008] Die Aufgabe der vorliegenden Erfindung besteht darin, eine verbesserte und flexiblere Leistungsfähigkeit eines Datenspeichersystems, das Spiegel- und Paritätsspeicherbereiche aufweist, zu schaffen.

[0009] Diese Aufgabe wird durch ein Verfahren gemäß Anspruch 1 und ein Datenspeichersystem gemäß Anspruch 12 gelöst.

ZUSAMMENFASSUNG DER ERFINDUNG

[0010] Ein Ausführungsbeispiel der Erfindung liefert ein hierarchisches Datenspeichersystem, das Daten gemäß verschiedener Redundanztechniken speichert, um Leistungsfähigkeit und Zuverlässigkeit zu optimieren. Das Datenspeichersystem umfasst ein Plattenarray, das eine Mehrzahl von Speicherplatten aufweist, und eine Plattenarraysteuerung, die einen Datentransfer zu und von den Platten koordiniert. Die Speicherplatten definieren einen physischen Speicherraum. Das Datenspeichersystem umfasst ferner ein RAID-Verwaltungssystem, das wirksam mit der Plattenarraysteuerung gekoppelt ist, um zwei virtuelle Speicherräume auf den physischen Speicherraum der Speicherplatten abzubilden. Der erste virtuelle Speicherraum bzw. der virtuelle Speicherraum auf RAID-Ebene präsentiert den physischen Speicherraum als Spiegel- und Paritäts-RAID-Bereiche, die Daten gemäß einer RAID-Ebene **1** (Spiegelredundanz) und einer RAID-Ebene **5** (Paritätsredundanz) speichern. Die Spiegel-RAID-Bereiche enthalten mehrere Spiegelzuweisungsblöcke, und die Paritäts-RAID-Bereiche enthalten mehrere Paritätszuweisungsblöcke. Der zweite virtuelle Speicherraum bzw. der virtuelle Speicherraum auf Anwendungsebene präsentiert den physischen Speicherraum als mehrere virtuelle Blöcke. Die virtuellen Blöcke sind über Zeiger, die in einer Virtuelle-Blöcke-Tabelle gehalten werden, den Spiegel- und Paritätszuweisungsblöcken zugeordnet.

[0011] Anfänglich speicherte das RAID-Verwaltungssystem alle Daten in Spiegel-RAID-Bereichen, da eine Spiegelredundanz eine größere Leistungsfähigkeit und Zuverlässigkeit liefert als eine Paritätsredundanz. Alle Daten können in Spiegel-RAID-Bereichen gespeichert werden, bis die zugewiesene Kapazität (d.h. die Menge an Benutzerdaten, die in den virtuellen Blöcken des virtuellen Speicherraums auf Anwendungsebene gespeichert sind) die Hälfte der physischen Gesamtkapazität des Plattenarrays überschreitet. Nachdem diese Schwelle überschritten wurde, beginnt das RAID-Verwaltungssystem, einige Daten in Paritäts-RAID-Bereichen zu speichern. Das RAID-Verwaltungssystem bewegt Daten zwischen den Spiegel- und Paritäts-RAID-Bereichen, so dass die Daten eine Veränderung der Redundanz zwischen RAID-Ebene **1** (d.h. Spiegelredundanz) und RAID-Ebene **5** (d.h. Paritätsredundanz) erfahren, auf eine Weise, die die Leistungsfähigkeit und Zuverlässigkeit optimiert.

[0012] Um zu gewährleisten, dass für diese Bewegung ausreichend Raum reserviert wird, beschränkt das RAID-Verwaltungssystem die Anzahl von virtuellen Blöcken, die in den Spiegel-RAID-Bereichen zugeordnet werden können. Die maximale Anzahl virtueller Blöcke in Spiegel-RAID-Bereichen wird auf der Basis einer Funktion der physischen Kapazität der Speicherplatten, der Anzahl von Speicherplatten und der zugewiesenen Kapazität zum Zeitpunkt jeder Speicheranforderung von dem Benutzer dynamisch berechnet.

Beschreibung der Zeichnungen

[0013] Unter Bezugnahme auf die folgenden beigefügten Zeichnungen, in denen Beispiele gezeigt sind, die den besten Modus zum Praktizieren der Erfindung verkörpern, werden nachfolgend bevorzugte Ausführungsbeispiele der Erfindung beschrieben.

[0014] Fig. 1 ist ein schematisches Blockdiagramm eines Speichersystems gemäß dieser Erfindung.

[0015] Fig. 2 ist eine schematische Veranschaulichung, die eine Datenspeicherung auf RAID-Ebene **1** zeigt;

[0016] Fig. 3 ist eine schematische Veranschaulichung, die eine Datenspeicherung auf RAID-Ebene **5** zeigt; und

[0017] Fig. 4 ist eine schematische Veranschaulichung einer Speicherabbildungsanordnung dieser Erfindung, bei der zwei virtuelle Speicherräume auf einen physischen Speicherraum abgebildet werden.

[0018] Fig. 5 ist ein Flussdiagramm eines Systeminitialisierungsverfahrens zum Reservieren eines Speicher-

raums für eine Datenbewegung gemäß einem Aspekt dieser Erfindung.

[0019] Fig. 6 ist ein Flussdiagramm eines Verfahrens zum Gewährleisten, dass in einem Spiegelspeicher gespeicherte virtuelle Blöcke unter einer maximal zulässigen Anzahl bleiben.

[0020] Fig. 7 ist ein Graph, der eine maximal zulässige Anzahl von in einem Spiegelspeicher gespeicherten virtuellen Blöcken in Bezug auf die Teilmenge der physischen Kapazität, die zum Speichern von Daten verwendet wird, veranschaulicht. Die graphischen Informationen in Fig. 7 beruhen auf einem homogenen Plattenarray und zeigen Fälle für drei unterschiedliche physische Kapazitäten.

[0021] Fig. 8 ist ein Graph, der die Anzahl von sowohl in einem Spiegel- als auch in einem Paritätsspeicher gespeicherten zugewiesenen virtuellen Blöcken in Bezug auf die Teilmenge der für die Datenspeicherung verwendeten physischen Kapazität veranschaulicht. Die graphischen Informationen in Fig. 8 beruhen auf einem heterogenen Plattenarray und zeigen Fälle für drei verschiedene physische Kapazitäten.

[0022] Fig. 1 zeigt ein Datenspeichersystem **10**, das gemäß dieser Erfindung aufgebaut ist. Vorzugsweise ist das Datenspeichersystem **10** ein Plattenarraydatenspeichersystem, das folgende Merkmale umfasst: ein hierarchisches Plattenarray **11**, das eine Mehrzahl von Speicherplatten **12** aufweist, eine Plattenarraysteuerung **14**, die mit dem Plattenarray **11** gekoppelt ist, um einen Transfer von Benutzerdaten zu und von den Speicherplatten **12** zu koordinieren, und ein RAID-Verwaltungssystem **16**.

[0023] Für die Zwecke dieser Offenbarung ist eine „Platte" jegliche nicht-flüchtige, direkt zugreifbare, wiederbeschreibbare Massenspeichervorrichtung, die die Fähigkeit aufweist, ihre eigenen Speicherausfälle zu erfassen. Sie umfasst sowohl sich drehende magnetische und optische Platten als auch Festzustandsplatten, oder nicht-flüchtige elektronische Speicherelemente (z.B. PROMs, EPROMs und EEPROMs). Der Begriff „Plattenarray" umfasst eine Sammlung von Platten, die Hardware, die erforderlich ist, um sie mit einem oder mehreren Hostcomputern zu verbinden, und eine Verwaltungssoftware, die verwendet wird, um den Betrieb der physischen Platten zu steuern und sie der Host-Betriebsumgebung als eine oder mehrere virtuelle Platten zu präsentieren. Eine „virtuelle Platte" ist eine abstrakte Entität, die durch die Verwaltungssoftware in dem Plattenarray verwirklicht ist.

[0024] Der Begriff „RAID" (Redundant Array of Independent Plattes, redundantes Array aus unabhängigen Platten) bedeutet ein Plattenarray, bei dem ein Teil der physischen Speicherkapazität verwendet wird, um redundante Informationen über Benutzerdaten, die auf dem Rest der Speicherkapazität gespeichert sind, zu speichern. Die redundanten Informationen ermöglichen eine Regeneration von Benutzerdaten für den Fall, dass eine der zu dem Array gehörenden Platten oder der Zugriffspfad zu derselben ausfällt. Eine ausführlichere Erläuterung von RAID-Systemen findet sich in einem Buch mit dem Titel The RAIDBook: A Source Book for RAID Technology, das am 9. Juni 1993 von RAID Advisory Board, Lino Lakes, Minnesota, veröffentlicht wurde.

[0025] Die Plattenarraysteuerung **14** ist über einen oder mehrere Schnittstellenbusse **13**, beispielsweise eine Kleincomputer-Schnittstelle (SCSI – small computer system interface) mit dem Plattenarray **11** gekoppelt. Das RAID-Verwaltungssystem **16** ist über ein Schnittstellenprotokoll **15** wirksam mit der Plattenarraysteuerung **14** gekoppelt. Ferner ist das Datenspeichersystem **10** über einen I/O-Schnittstellenbus **17** mit einem Hostcomputer (nicht gezeigt) gekoppelt. Das RAID-Verwaltungssystem **16** kann als separate Komponente verkörpert sein oder kann in der Plattenarraysteuerung **14** oder in dem Hostcomputer konfiguriert sein, um eine Datenverwaltungseinrichtung zum Steuern von Plattenspeicherungs- und von Zuverlässigkeitsebenen sowie zum Transferieren von Daten zwischen verschiedenen Zuverlässigkeits-Speicherstufen bereitzustellen. Diese Zuverlässigkeits-Speicherebenen sind, wie nachfolgend beschrieben wird, vorzugsweise Spiegel- oder Paritäts-Redundanzebenen, können jedoch auch eine Zuverlässigkeits-Speicherebene ohne jegliche Redundanz umfassen.

[0026] Die Plattenarraysteuerung **14** ist vorzugsweise als duale Steuerung implementiert, die aus der Plattenarraysteuerung A **14a** bezeichnet ist, und aus der Plattenarraysteuerung B **14b** bezeichnet ist, besteht. Die dualen Steuerungen **14a** und **14b** verbessern die Zuverlässigkeit, indem sie für den Fall, dass eine Steuerung funktionsuntüchtig wird, eine fortlaufende Sicherung und Redundanz liefern. Diese Erfindung kann jedoch auch mit einer einzigen Steuerung oder mit anderen Architekturen praktiziert werden.

[0027] Das hierarchische Plattenarray **11** kann als unterschiedliche Speicherräume, einschließlich seines physischen Speicherraums und eines oder mehrerer virtueller Speicherräume, charakterisiert werden. Diese verschiedenen Speicheransichten sind durch Abbildungstechniken verbunden. Beispielsweise kann der physische Speicherraum des Plattenarrays in einen virtuellen Speicherraum abgebildet werden, der Speicherbe-

reiche gemäß den verschiedenen Datenzuverlässigkeitsebenen darstellt. Manche Bereiche in dem virtuellen Speicherraum können einer ersten Zuverlässigkeitsspeicherebene zugewiesen sein, beispielsweise Spiegel- oder RAID-Ebene **1**, und andere Bereiche können einer zweiten Zuverlässigkeitsspeicherebene zugewiesen sein, beispielsweise Paritäts- oder RAID-Ebene **5**. Die verschiedenen Abbildungstechniken und virtuellen Plätze, die RAID-Ebenen betreffen, werden nachfolgend ausführlicher beschrieben.

[0028] Das Datenspeichersystem **10** umfasst einen Speicherabbildungsspeicher **21**, der eine durchgängige Speicherung der virtuellen Abbildungsinformationen liefert, die verwendet werden, um verschiedene Speicherräume ineinander abzubilden. Der Speicherabbildungsspeicher befindet sich außerhalb des Plattenarrays und vorzugsweise in der Plattenarraysteuerung **14**. Die Speicherabbildungsinformationen können durch die Steuerung oder das RAID-Verwaltungssystem kontinuierlich oder periodisch aktualisiert werden, während sich die verschiedenen Abbildungskonfigurationen unter den verschiedenen Ansichten ändern.

[0029] Vorzugsweise ist der Speicherabbildungsspeicher **21** als zwei nicht-flüchtige RAMs (Direktzugriffsspeicher) **21a** und **21b** verkörpert, die in jeweiligen Steuerungen **14a** und **14b** angeordnet sind. Ein beispielhafter nicht-flüchtiger RAM (NVRAM) ist ein batteriegepufferter RAM. Ein batteriegepufferter RAM verwendet Energie von einer unabhängigen Batteriequelle, um im Falle eines Verlustes von Leistung an das Datenspeichersystem **10** die Daten über einen bestimmten Zeitraum in dem Speicher beizubehalten. Ein bevorzugter Aufbau ist ein selbstauffrischender batteriegepufferter DRAM (dynamischer RAM).

[0030] Die dualen NVRAMs **21a** und **21b** sorgen für eine redundante Speicherung der Speicherabbildungsinformationen. Die virtuellen Abbildungsinformationen werden dupliziert und gemäß Spiegelredundanztechniken in beiden NVRAMs **21a** und **21b** gespeichert. Auf diese Weise kann der NVRAM **21a** dafür reserviert sein, die ursprünglichen Abbildungsinformationen zu speichern, und der NVRAM **21b** kann dafür reserviert sein, die redundanten Abbildungsinformationen zu speichern. Bei einem alternativen Aufbau kann ein gespiegelter Speicherabbildungsspeicher unter Verwendung eines einzelnen nichtflüchtigen RAMs mit einem ausreichenden Platz, um die Daten im Duplikat zu speichern, konfiguriert sein.

[0031] Wie in Fig. 1 gezeigt ist, weist das Plattenarray **11** mehrere Speicherplattenlaufwerksvorrichtungen **12** auf. Beispielhafte Größen dieser Speicherplatten betragen ein bis drei Gigabytes. Die Speicherplatten können unabhängig voneinander mit mechanischen Buchten verbunden oder von denselben getrennt sein, die eine Schnittstellenbildung mit dem SCSI-Bus **13** liefern. Bei einer Implementierung ist das Datenspeichersystem mit zwölf aktiven mechanischen Buchten entworfen. Vier SCSI-Busse werden verwendet, um eine Schnittstelle zwischen diesen Buchten und der Plattenarraysteuerung **14** zu bilden (d.h. ein Bus pro drei mechanischen Buchten). Falls die aktiven Buchten vollständig beladen sind, weist das Datenspeichersystem eine beispielhafte kombinierte Kapazität von 12 bis 36 Gigabytes auf. Die Plattenarraysteuerung **14** erkennt die Speicherplatten **12** unabhängig davon, in welche Bucht sie eingesteckt sind. Das Datenspeichersystem **10** ist entworfen, um ein „heißes Einstecken" zusätzlicher Platten in verfügbare mechanische Buchten in dem Plattenarray, während das Plattenarray in Betrieb ist, zu ermöglichen.

[0032] Die Speicherplatten **12** in dem Plattenarray **11** können zu Erläuterungszwecken vom Konzept her so erklärt werden, dass sie in einer Spiegelgruppe **18** aus mehreren Platten **20** und einer Paritätsgruppe **22** aus mehreren Platten **24** angeordnet sind. Die Spiegelgruppe **18** stellt eine erste Speicherposition oder einen ersten RAID-Bereich des Plattenarrays dar, die bzw. der Daten gemäß einer ersten bzw. Spiegelredundanzebene speichert. Diese Spiegelredundanzebene wird auch als RAID-Ebene **1** betrachtet. Die RAID-Ebene **1** bzw. oder ein Spiegeln von Platten bietet die höchste Datenzuverlässigkeit, indem sie insofern einen Eins-Zu-Eins-Schutz bietet, als jedes Datenbit dupliziert und in dem Datenspeichersystem gespeichert wird. Die Spiegelredundanz ist schematisch durch die drei Paare von Platten **20** in Fig. 1 dargestellt. Ursprüngliche Daten können auf einem ersten Satz von Platten **26** gespeichert werden, während dupliziert, redundante Daten auf dem gepaarten zweiten Satz von Platten **28** gespeichert werden.

[0033] Fig. 2 veranschaulicht die Speicherung von Daten gemäß der RAID-Ebene **1** noch ausführlicher. Die vertikalen Spalten stellen einzelne Platten dar, von denen Platten 0, 1, 2 und 3 veranschaulicht sind. Der in diesem Plattenarray aus vier Platten enthaltene physische Speicherplatz kann zu mehreren Streifen konfiguriert sein, wie sie durch die horizontalen Zeilen veranschaulicht sind. Ein „Streifen" erstreckt sich über die Speicherplatten und ist aus zahlreichen gleich großen Segmenten an Speicherplatz gebildet, wobei ein Segment jeder Platte in dem Array zugeordnet ist. Das heißt, dass ein Segment der Abschnitt eines Streifens ist, der sich auf einer einzelnen Platte befindet. Jeder Streifen hält eine vorbestimmte Datenmenge, die über die Speicherplatten verteilt ist. Manche Segmente eines Streifens werden für ursprüngliche Daten verwendet, während andere Segmente für redundante Daten verwendet werden.

[0034] Bei diesem Beispiel einer Spiegelredundanz (RAID-Ebene **1**) werden Daten, die auf der Platte 0 in dem Segment 0 des Streifens 0 gespeichert sind, dupliziert und auf der Platte 1 in dem Segment 0' des Streifens 0 gespeichert. Desgleichen werden Daten, die auf der Platte 2 in dem Segment 5 des Streifens 2 gespeichert sind, zu dem Segment 5' des Streifens 2 auf der Platte 3 gespiegelt. Auf diese Weise wird jede Dateneinheit dupliziert und auf den Platten gespeichert. Der Redundanzentwurf der Fig. 2 ist zu Erläuterungszwecken vorgesehen. Die redundanten Daten müssen nicht sauber in demselben Streifen plaziert werden, wie gezeigt ist. Beispielsweise könnten Daten, die auf der Platte 0 in dem Segment 2 des Streifens 1 gespeichert sind, dupliziert werden und auf der Platte 3 in dem Segment T' des Streifens S plaziert werden.

[0035] Unter erneuter Bezugnahme auf Fig. 1 stellt die Paritätsgruppe **22** der Platten **24** eine zweite Speicherposition bzw. einen zweiten RAID-Bereich dar, in der bzw. dem Daten gemäß einer zweiten Redundanzebene, z.B. RAID-Ebene **5**, gespeichert werden. In dieser erläuternden Veranschaulichung von sechs Platten sind ursprüngliche Daten auf der fünften Platte **30** gespeichert, und redundante „Paritäts"-Daten sind auf der sechsten Platte **32** gespeichert.

[0036] Fig. 3 zeigt einen Paritäts-RAID-Bereich-Entwurf etwas ausführlicher. Ähnlich dem Entwurf des Spiegel-RAID-Bereichs der Fig. 2 kann der physische Speicherraum der Platten 0, 1, 2, 3 zu mehreren gleich großen Streifen konfiguriert sein. Bei diesem veranschaulichten Beispiel werden Daten gemäß der RAID-Ebene **5** gespeichert, und die redundanten Daten werden in den Segmenten gespeichert, auf die durch den Buchstaben P Bezug genommen wird. Die redundanten P-Segmente speichern die Parität der anderen Segmente in dem Streifen. In dem Streifen 0 speichert das redundante P-Segment auf der Platte 3 beispielsweise die Parität der Platten 0, 1 und 2. Die Parität für jeden Streifen wird durch eine bestimmte Funktion berechnet, beispielsweise eine Exklusiv-ODER-Funktion, die durch das Symbol „⊕" dargestellt wird. Die Paritäten für die ersten vier Streifen (wobei die tiefgestellte Zahl den entsprechenden Streifen darstellt) lauten wie folgt:

P_0 = Segment 0 ⊕ Segment 1 ⊕ Segment 2
= Platte 0 ⊕ Platte 1 ⊕ Platte 2

P_1 = Segment 3 ⊕ Segment 4 ⊕ Segment 5
= Platte 0 ⊕ Platte 1 ⊕ Platte 3

P_2 = Segment 6 ⊕ Segment 7 ⊕ Segment 8
= Platte 0 ⊕ Platte 2 ⊕ Platte 3

P_3 = Segment 9 ⊕ Segment 10 ⊕ Segment 11
= Platte 1 ⊕ Platte 2 ⊕ Platte 3

[0037] Eine Paritätsredundanz ermöglicht eine Regeneration von Daten, die auf einer der Platten nicht mehr zur Verfügung stehen. Wenn beispielsweise die Daten in dem Segment 5 nicht mehr zur Verfügung stehen, kann ihr Inhalt aus den Segmenten 3 und 4 und den Paritätsdaten in dem Segment P ermittelt werden. Eine Paritätsspeicherung ist kostengünstiger als eine Spiegelspeicherung, ist jedoch auch weniger zuverlässig und weist eine geringere Leistungsfähigkeit auf.

[0038] Die Plattenanordnung der Fig. 1 ist zu Konzeptionszwecken vorgesehen. In der Praxis weist das Plattenarray **11** einfach eine Mehrzahl von Platten **12** auf, die in der Lage sind, Daten gemäß einer Spiegel- und Paritätsredundanz zu speichern. Aus dem durch alle Platten **12** gelieferten, zur Verfügung stehenden Speicherraum wird ein Abschnitt dieses Speicherraums einer Spiegelredundanz zugewiesen, und ein weiterer Abschnitt wird einer Paritätsredundanz zugewiesen. Vorzugsweise sind die Platten **12** konfiguriert, um mehrere gleich große Speicherregionen (die bei Fig. 4 mit dem Bezugszeichen **35** benannt sind) zu enthalten, wobei einzelne Regionen mehrere Segmente aufweisen. Die Regionen sind gruppiert, um bei einer virtuellen Ansicht des Speicherraums RAID-Bereiche zu bilden. Zusätzlich wird eine weitere (durch einen Host definierte) Ansicht eines Speicherraums dem Benutzer oder Host präsentiert, so dass die RAID-Bereiche und die Datenredundanz-Speichertechniken für den Benutzer oder Host transparent sind. Diese Merkmale werden nachstehend unter Bezugnahme auf Fig. 4 ausführlicher erläutert.

[0039] Das Datenspeichersystem **10** verwaltet die „Migration bzw. Bewegung" von Daten zwischen Spiegel- und Paritätsspeicherschemata. Die Verwaltung beider Redundanztypen wird durch das RAID-Verwaltungssystem **16** (Fig. 1) koordiniert. Das RAID-Verwaltungssystem **16** verwaltet die zwei unterschiedlichen Typen von RAID-Bereichen in dem Plattenarray als Speicherhierarchie, wobei die Spiegel-RAID-Bereiche ähnlich einem Cache für die Paritäts-RAID-Bereiche agieren. Das RAID-Verwaltungssystem **16** verschiebt und organisiert die

Daten zwischen dem Spiegel- und dem Paritäts-RAID-Bereich gemäß einem definierten Leistungsfähigkeitsprotokoll, und verwaltet die Daten ansonsten auf ebensolche Weise. Der Vorgang des Bewegens von Daten zwischen dem Spiegel- und dem Paritäts-RAID-Bereich wird als „Migration" bezeichnet.

[0040] Das Datenspeichersystem **10** versucht, die für die Leistungsfähigkeit kritischeren Daten in die Spiegel-RAID-Bereiche zu plazieren, da dies die höchste Leistungsfähigkeit und Zuverlässigkeit liefert. Die Leistungsfähigkeitsprotokolle, die durch das RAID-Verwaltungssystem **16** implementiert sind, umfassen eine von zwei bevorzugten Migrationsrichtlinien. Gemäß der ersten Migrationsrichtlinie, die als „Zugriffshäufigkeit" bekannt ist, werden die Daten in dem hierarchischen Plattenarray, auf die am häufigsten zugegriffen wird, in dem Spiegel-RAID-Bereich **18** beibehalten. Daten, auf die weniger häufig zugegriffen wird, werden in dem Paritäts-RAID-Bereich **22** gehalten. Gemäß einer zweiten Migrationsrichtlinie, die als „Zugriffsaktualität" bekannt ist, werden die Daten, deren Zugriff am wenigsten lange zurückliegt, in dem Spiegel-RAID-Bereich **18** beibehalten, während die Daten, deren Zugriff länger zurückliegt, in dem Paritäts-RAID-Bereich **22** gespeichert werden. Es können auch andere Leistungsfähigkeitsprotokolle verwendet werden. Idealerweise sind solche Protokolle auf der Basis der spezifischen Computeranwendung und der Erfordernisse des Benutzers definiert.

[0041] Zusätzlich „stimmt" das RAID-Verwaltungssystem **16** die Speicherressourcen eines Datenspeichersystems gemäß einer Funktion von zwei Parametern „ab". Größe der physischen Speicherkapazität und Größe der vorhandenen Menge an Benutzerdaten, die in dem Datenspeichersystem gespeichert sind. Anfänglich werden alle Daten in Spiegel-RAID-Bereichen gespeichert, da dies die höchste Leistungsfähigkeit und Zuverlässigkeit liefert. Während mehr Daten zu dem Datenspeichersystem hinzugefügt werden, werden die Daten zwischen Spiegel-RAID-Bereichen und Paritäts-RAID-Bereichen bewegt, um die Leistungsfähigkeit und Zuverlässigkeit zu optimieren. Während sich das Datenspeichersystem an eine volle Kapazität annähert, werden im Rahmen des Bemühens, alle durch den Benutzer gestellten Anforderungen zu erfüllen und trotzdem durch Redundanz eine Zuverlässigkeit zu liefern, immer mehr Daten in Paritäts-RAID-Bereiche bewegt. Dementsprechend liefert das Datenspeichersystem dieser Erfindung eine maximale Flexibilität und Anpassung. Es erfordert nicht, dass der Benutzer einen spezifischen Speicherbetriebszustand auswählt, sondern kann sich statt dessen an jegliche durch den Benutzer an das System gestellte Forderung anpassen.

[0042] Fig. 4 veranschaulicht ein Speicherabbilden des verfügbaren Speicherraums des Datenspeichersystems **10** als mehrere Lagen eines abgebildeten virtuellen Speicherraums. Jedes vertikal verlängerte Rechteck in dem Diagramm stellt eine Ansicht des physischen Speicherraums dar. Bei diesem Diagramm wird auf den physischen Speicherraum **34** durch zwei virtuelle Speicheransichten **40** und **50** Bezug genommen. Der physische Speicherraum **34** wird durch vier Platten (z.B. Platten **12** in Fig. 1) dargestellt, die durch die Bezugszeichen 0, 1, 2 und 3 benannt sind. Die den Platten zugeordneten vier Rechtecke stellen eine Ansicht des physischen Speicherraums dar, bei der die Platten 1, 2 und 3 eine ungefähr gleiche Speicherkapazität aufweisen und die Platte 0 eine etwas geringere Speicherkapazität aufweist. Beispielhafte Speicherkapazitäten für derartige Platten betragen 1 bis 3 Gigabytes. Der Speicherraum **34** ist in Bereiche A0, A1, A2 usw. unterteilt. Einzelne Bereiche enthalten zahlreiche Streifen, im Bereich A0 z.B. Streifen O–Q. Einzelne Bereiche enthalten auch zahlreiche Regionen **35**. Die Regionen **35** bestehen vorzugsweise aus einer ausgewählten Anzahl von Segmenten einer gleichmäßigen Größe auf jeder Speicherplatte, so dass die Regionen über das gesamte Plattenarray eine gleiche Größe aufweisen. Eine beispielhafte Größe einer Region **35** ist ein Megabyte.

[0043] Der Speicherraum der Platten wird in eine erste virtuelle Zwischenansicht **40** der RAID-Ebene des physischen Speicherraums **34** abgebildet. Diese erste virtuelle Ansicht ist vom Konzept her ein Satz von RAID-Bereichen, die auf eine zweite Anwendungsansicht abgebildet werden können, die einen angrenzend adressierbaren Speicherraum darstellt. Die physische Konfiguration und RAID-Ansichten des Speicherraums entziehen sich der Anwendungsansicht.

[0044] Der RAID-Bereich-Speicherraum **40** ist die Speicheransicht, die den Spiegel- und Paritätsspeicherraum identifiziert. Beispielsweise kann ein RAID-Bereich **42** einen Spiegel-RAID-Bereich von M Zuweisungsblöcken **43** darstellen, während der RAID-Bereich **44** einen Paritäts-RAID-Bereich von N Zuweisungsblöcken **45** darstellen kann. Die Zuweisungsblöcke **43** und **45** weisen vorzugsweise eine gleiche Größe auf, wobei eine beispielhafte Größe 64 Kilobytes beträgt. Diese RAID-Bereiche beziehen sich auf entsprechende physische Bereiche A0, A1, A2 usw. auf dem physischen Speicherraum **34**. Beispielsweise können sechzehn 64K-Zuweisungsblöcke **43** oder **45** an der virtuellen RAID-Ansicht auf eine einzige 1M-Region **35** abgebildet werden.

[0045] Die Spiegel- und Paritäts-RAID-Bereiche können den gesamten Speicherraum **34** des Plattenarrays einnehmen, müssen aber nicht. Dementsprechend kann während bestimmter Anwendungen ein unbenutzter und unbezeichneter Speicherraum vorliegen, der keinem bestimmten RAID-Bereich entspricht. Ein solcher

Speicherraum kann jedoch in einen Spiegel- oder Paritäts-RAID-Bereich umgewandelt werden. Ferner ist zu beachten, dass die RAID-Bereiche in der Abbildung in benachbarte Bereiche auf dem Plattenarray abgebildet werden, wobei jede einem RAID-Bereich zugeordnete Region an derselben physischen Adresse auf jeder Speicherplatte angeordnet ist. Die RAID-Bereiche können alternativ auch zu nicht-benachbarten Bereichen auf dem Plattenarray abgebildet werden.

[0046] Der in den RAID-Bereichen verfügbare Speicherraum wird in eine zweite virtuelle Vorderes-Ende-Ansicht **50** der Anwendungsebene abgebildet, die eine Speicheransicht ist, wie sie durch den Benutzer oder das Hostanwendungsprogramm definiert und demselben präsentiert wird. Bei einer Betrachtung durch den Benutzer oder das Hostanwendungsprogramm kann die virtuelle Ansicht **50** der Anwendungsebene eine einzige große Speicherkapazität darstellen, die den verfügbaren Speicherraum auf den Speicherplatten **12** angibt. Der virtuelle Speicherraum **50** präsentiert eine Ansicht eines linearen Satzes von gleich großen virtuellen Speicherblöcken **52** und **53**, die einzeln als 0, 1, 2, ... J – 1, J, J + 1, ..., L – 1, L, L + 1, ..., usw. bezeichnet werden. Die virtuellen Blöcke **52** und **53** weisen vorzugsweise dieselbe Größe auf wie die Zuweisungsblöcke in dem RAID-Bereich **40**, wobei eine beispielhafte Größe 64 Kilobytes beträgt. Der virtuelle Blockspeicherraum **50** wird in der durch die RAID-Bereiche **40** dargestellten Ansicht durch eine Tabelle von Bezugnahmen oder Zeigern (wie sie durch Pfeile **54** dargestellt sind) auf Zuweisungsblöcke, dargestellt. Die virtuellen Blöcke **52** und **53** bei der virtuellen Anwendungsansicht **50** sind somit über die in der virtuellen Blocktabelle beibehaltenen Zeiger den Zuweisungsblöcken **43** und **45** an der virtuellen RAID-Ansicht **50** zugeordnet. Es gibt mindestens zwei Arten von RAID-Bereichen, auf die von der virtuellen Blocktabelle verwiesen werden kann: Spiegel und Parität.

[0047] Das RAID-Verwaltungssystem **16** kann die Konfiguration der RAID-Bereiche über den physischen Speicherraum dynamisch verändern. Die Anzahl von RAID-Bereichen für jeden Typ kann je nach der Menge von Benutzerdaten, die in dem System gespeichert sind, und je nach der Größe des physischen Plattenspeicherraums erhöht oder verringert werden. Folglich befinden sich das Abbilden der RAID-Bereiche in der virtuellen Ansicht **40** der RAID-Ebene auf die Platten und das Abbilden der virtuellen Front-End-Ansicht **50** auf die RAID-Ansicht **40** allgemein in einem Veränderungszustand. Der Speicherabbildungsspeicher in den NVRAMs **21a** und **21b** (Fig. 1) behält die aktuellen Abbildungsinformationen bei, die durch das RAID-Verwaltungssystem **16** verwendet werden, um die RAID-Bereiche auf die Platten abzubilden, und behält die Informationen bei, die verwendet werden, um zwischen die zwei virtuellen Ansichten abzubilden. Während das RAID-Verwaltungssystem die Abbildungen der RAID-Ebene dynamisch verändert, aktualisiert es auch die Abbildungsinformationen in dem Speicherabbildungsspeicher, um die Veränderungen zu reflektieren.

[0048] Die Migrationsoperation des Speichersystems **10** wird nun unter Bezugnahme auf Fig. 1 und Fig. 4 beschrieben.

[0049] Für die Zwecke einer fortführenden Erläuterung nehmen virtuelle Blöcke **53** des virtuellen Speicherraums **50** der Anwendungsebene Bezug auf zugeordnete Zuweisungsblöcke **45** in dem Paritäts-RAID-Bereich **44**, die in dem Bereich A1 des physischen Speicherraums **34** gespeichert sind. Derartige virtuelle Blöcke **53** können als „virtuelle Paritätsblöcke" bezeichnet werden, während die zugeordneten Zuweisungsblöcke **45** als „Paritätszuweisungsblöcke" bezeichnet werden. Desgleichen nehmen virtuelle Blöcke **52** Bezug auf zugeordnete Zuweisungsblöcke **43** in dem Spiegel-RAID-Bereich **42**, die in dem Bereich A0 des physischen Speicherraums **34** gespeichert sind. Derartige virtuelle Blöcke **52** können hierin als „virtuelle Spiegelblöcke" bezeichnet werden, während die zugeordneten Zuweisungsblöcke **43** als „Spiegelzuweisungsblöcke" bezeichnet werden.

[0050] Um Daten von einem RAID-Bereich zu einem anderen zu bewegen, wird allgemein ein virtueller Block, der einem Zuweisungsblock eines Typs der ersten RAID-Ebene (z.B. Spiegel oder Ebene **1**) zugeordnet ist, ausgewählt. Anschließend wird ein nicht benutzter Zuweisungsblock, der einen Typ der zweiten RAID-Ebene (beispielsweise Parität oder Ebene **5**) darstellt, lokalisiert. Falls ein nicht benutzter Zuweisungsblock nicht lokalisiert werden kann, wird einer erzeugt. Als nächstes werden Daten von dem Zuweisungsblock, der zuvor dem ausgewählten virtuellen Block zugeordnet war, an den nicht benutzten Zuweisungsblock transferiert, der bewirkt, dass die Daten einer Redundanzebenenänderung unterzogen werden. Beispielsweise werden Daten, die einst gemäß der Spiegelredundanz gespeichert wurden, nun gemäß einer Paritätsredundanz gespeichert, oder umgekehrt. Als abschließenden Schritt wird die **Abb.** 54 des virtuellen Speicherraums **50** der Anwendungsebene auf den virtuellen Speicherraum **40** der RAID-Ebene modifiziert und aktualisiert, um die Datenverschiebung widerzuspiegeln. Der ausgewählte virtuelle Block, der zuvor einem Zuweisungsblock des Typs der ersten RAID-Ebene zugeordnet war, nimmt nun über einen aktualisierten Zeiger Bezug auf einen Zuweisungsblock des Typs der zweiten RAID-Ebene, der die migrierten Daten enthält. Jegliche Abbildungsveränderung, die während dieser Transaktion auftritt, würde in dem Speicherabbildungsspeicher **21** aktualisiert wer-

den.

[0051] Die fortdauernde Erörterung liefert eine ausführlichere Erklärung des Bewegens von Daten zwischen Spiegel- und Paritätsspeicherbereichen gemäß bevorzugten Verfahren und Sequenzen dieser Erfindung. Zum Bewegen von Daten von einem Paritäts- zu einem Spiegelspeicher wird die folgende Sequenz verwendet:

1. Das RAID-Verwaltungssystem lokalisiert einen unbenutzten Spiegelzuweisungsblock **43** in einem Spiegel-RAID-Bereich **42**.
2. Falls keiner gefunden wird, erzeugt das RAID-Verwaltungssystem einen Spiegelzuweisungsblock (nachfolgend erläutert).
3. Das RAID-Verwaltungssystem hebt neue Speicheranforderungen an den zu bewegenden virtuellen Block zeitweilig auf.
4. Das RAID-Verwaltungssystem wartet, bis alle aktiven Datenspeicherungsanforderungen an den virtuellen Block abgeschlossen sind.
5. Die Daten von dem Paritätszuweisungsblock **45**, der dem virtuellen Block **53** zugewiesen ist, werden in einen vorübergehenden Speicherpuffer eingelesen.
6. Die Daten werden anschließend in den bei Schritt **2** gewählten Spiegelzuweisungsblock **43** geschrieben.
7. Die virtuelle Blocktabelle wird modifiziert, um auf die neue Position der Daten in dem Spiegelzuweisungsblock **43** zu verweisen.
8. Die zeitweilig aufgehobenen Speicheranforderungen werden wiederaufgenommen.

[0052] Gemäß der obigen Prozedur bewegte sich ein virtueller Block **53** in der virtuellen Ansicht **50** der Anwendungsebene von einer Paritäts- zu einer Spiegelspeicherung. Relativ zu der virtuellen Zwischenansicht **40** bewegten sich Daten von einem Paritätszuweisungsblock **45** in dem Paritäts-RAID-Bereich **44** zu einem Spiegelzuweisungsblock **43** in dem Spiegel-RAID-Bereich **42**. In dem physischen Speicherraum bewegten sich Daten von dem Bereich A1 zum Bereich A0.

[0053] Falls kein unbenutzter Spiegelzuweisungsblock lokalisiert werden kann (Schritt 1 oben), versucht das RAID-Verwaltungssystem die folgende bevorzugte Sequenz von drei Techniken. Erstens versucht das RAID-Verwaltungssystem, einen unbenutzten (und somit unbenannten) RAID-Bereich zu lokalisieren, der in einen Spiegel-RAID-Bereich umgewandelt werden kann, ohne die Systemschwelle eines unbenutzten Speichers der RAID-Ebene zu verletzen, die benötigt wird, um zu garantieren, dass eine Migration immer stattfinden kann. Falls dies fehlschlägt und das System mehr als die reservierte Menge an unbenutztem Speicher der RAID-Ebene aufweist, bewegt das System Daten innerhalb des Paritätsspeichers, um unbenutzte Paritätszuweisungsblöcke zu unbenutzten RAID-Bereichen zu sammeln. Falls diese Migration einen unbenutzten RAID-Bereich ergibt, der, wie oben erwähnt, in einen Spiegel-RAID-Bereich umgewandelt werden kann, wandelt das System ihn in einen Spiegel-RAID-Bereich um. Andernfalls bewegt das System alternativ dazu Daten von einem Spiegel- zu einem Paritätsspeicher, verdichtet den Spiegelspeicher und wandelt einen unbenutzten Speicher der RAID-Ebene in einen Paritätsspeicher um, bis das System den unbenutzten Speicher der RAID-Ebene ausreichend für die Lokalisierung eines unbenutzten Spiegelzuweisungsblocks oder eine Umwandlung eines unbenutzten RAID-Bereichs in einen Spiegel-RAID-Bereich erhöht. Da Spiegelzuweisungsblöcke mehr physischen Speicherraum belegen als Paritätszuweisungsblöcke, führt diese letzte Technik zu einer Nettoerhöhung des Umfangs an unbenutztem Speicher der RAID-Ebene.

[0054] Das Erzeugungs-/Umwandlungsprotokoll, das verwendet wird, um unbenutzte Spiegelzuweisungsblöcke zu lokalisieren und einzurichten, ist vorteilhaft, da es dem RAID-Verwaltungssystem ermöglicht, die Speicherzuweisung gemäß der Menge an Benutzerdaten und der Größe des physischen Speicherraums zwischen Paritäts- und Spiegelbereichen selektiv einzustellen. Da die Datennutzung und die Speicherkapazität variieren, verwendet das RAID-Verwaltungssystem eine oder mehrere der obigen drei Techniken, um die Datenmenge, die in einem Spiegelspeicher gehalten wird, zu maximieren.

[0055] Das RAID-Verwaltungssystem versucht, die Situation zu vermeiden, bei der eine Speicheranforderung warten muss, bis die Platzschaffungssequenz einen unbenutzten Spiegelzuweisungsblock ergibt, indem sie während einer Leerlaufzeit unbenutzte RAID-Bereiche erzeugt. In manchen Situationen können Speicheranforderungen während der Platzschaffungssequenz jedoch zeitweilig aufgehoben werden. Das RAID-Verwaltungssystem konfiguriert den virtuellen Blockspeicherraum derart, dass der virtuelle Platz geringer ist als die RAID-Ansicht. Dies gewährleistet, dass für eine Migration oder andere Zwecke ein freier Platz reserviert wird, der gleich zumindest einem RAID-Bereich ist. Auf diese Weise ergibt die Sequenz von Techniken immer einen unbenutzten Spiegelzuweisungsblock.

[0056] Um Daten von einem Spiegel- zu einem Paritätsspeicher zu bewegen, wird die folgende Sequenz ver-

wendet:

1. Das RAID-Verwaltungssystem wählt einen virtuellen Block aus **52** aus, um gemäß einer Migrationsrichtlinie, beispielsweise Zugriffsaktualität oder Zugriffshäufigkeit, von einem Spiegel- zu einem Paritätsspeicher zu bewegen.
2. Das RAID-Verwaltungssystem lokalisiert einen unbenutzten Paritätszuweisungsblock **45** in einem Paritäts-RAID-Bereich **44**.
3. Falls kein solcher Block gefunden werden kann, wird ein für eine Migration reservierter Platz gemäß den oben beschriebenen Erzeugungstechniken in einen Paritäts-RAID-Bereich umgewandelt.
4. Neue Speicheranforderungen an den zu bewegenden virtuellen Block werden zeitweilig aufgehoben.
5. Das RAID-Verwaltungssystem wartet, bis alle aktiven Speicheranforderungen an den virtuellen Block abgeschlossen sind.
6. Daten werden von dem Spiegelzuweisungsblock **43**, der dem virtuellen Block **52** zugeordnet ist, in einen vorübergehenden Speicherpuffer gelesen.
7. Die Daten werden in den gewählten Paritätszuweisungsblock **45** geschrieben.
8. Die virtuelle Blocktabelle wird modifiziert, um auf die neue Position der Daten in dem Paritätszuweisungsblock **45** zu verweisen.
9. Datenanforderungen an den virtuellen Block werden wiederaufgenommen.

[0057] Die obigen zwei aufgeführten Sequenzen liefern Beispiele dessen, wie das Speichersystem dieser Erfindung arbeiten kann, um Daten zwischen zwei verschiedenen Redundanzebenen zu bewegen.

[0058] Um die Datenbewegung zu bewerkstelligen, besteht ein Ziel des Datenspeichersystems darin, zu gewährleisten, dass eine ausreichende Menge an Speicherraum für die Migration bzw. Bewegung reserviert ist. Ferner besteht ein weiteres Ziel darin, möglichst viele Benutzerdaten in dem Spiegelspeicher zu halten, da die Spiegelredundanz die höchste Leistungsfähigkeit und Zuverlässigkeit bietet. Jedoch existiert lediglich eine endliche Menge an physischem Speicherraum in dem Plattenarray. Virtuelle Blöcke in dem Spiegelspeicher beginnen die physische Kapazität einzunehmen, wenn Daten in dem Plattenarray gespeichert werden. Der Begriff „physische Kapazität" bezieht sich auf die Speicherkapazität, die zum Speichern von Benutzerdaten und ihren Redundanzinformationen zur Verfügung steht.

[0059] Eine Spiegelspeicherung verstopft effektiv das Plattenarray, wenn sie die Hälfte der physischen Kapazität minus den reservierten Speicherraum einnimmt. An diesem Punkt werden alle verwendeten virtuellen Blöcke in einer Spiegelspeicherung gehalten. Falls die Spiegelspeicherung so viel des physischen Speicherraums einnimmt, dass der Umfang an ungenutztem RAID-Raum geringer ist als der eines RAID-Bereichs, kann das Datenspeichersystem eine Migration bzw. Wanderung einzelner virtueller Blöcke von Spiegel zu Paritäts-RAID-Bereichen nicht mehr garantieren. Die nächste Clientdatenanforderung, die einen zuvor ungenutzten virtuellen Block verwendet, wird fehlschlagen, da der Umfang an ungenutztem Raum geringer ist als der eines RAID-Bereiches und nicht in eine Paritätsspeicherung zum Bewegen einzelner virtueller Blöcke aus der Spiegelspeicherung heraus umgewandelt werden kann. Diese Situation muss vermieden werden.

[0060] Das RAID-Verwaltungssystem **16** weist deshalb eine ausreichende Menge an reserviertem Speicherraum für die Datenmigration zu. Dies erfolgt durch eine Kombination zweier Techniken. Erstens reserviert das RAID-Verwaltungssystem logisch eine Menge an Speicherraum, der zumindest zu einem Spiegel-RAID-Bereich äquivalent ist. Zweitens begrenzt das RAID-Verwaltungssystem die Anzahl von Blöcken, die einer Spiegelspeicherung innerhalb des Datenspeichersystems zugewiesen werden können. Die maximale Anzahl virtueller Blöcke in einer Spiegelspeicherung wird auf der Grundlage einer Funktion der physischen Kapazität der Speicherplatten, der Anzahl der Speicherplatten und der zugewiesenen Kapazität zum Zeitpunkt jeder Speicherungsanforderung von dem Benutzer dynamisch berechnet.

[0061] Fig. 5 liefert ein Flussdiagramm eines Initialisierungsverfahrens dieser Erfindung zum Reservieren von Speicherraum für eine Datenmigration. Die Schritte sind unter zusätzlicher Bezugnahme auf Fig. 1 und Fig. 4 beschrieben.

[0062] Die ersten zwei Schritte **100** und **102** richten die zweistufige virtuelle Speicherraumumgebung für das hierarchische Datenspeichersystem ein. Bei Schritt **100** bildet das RAID-Verwaltungssystem **16** den durch das Plattenarray **12** bereitgestellten physischen Speicherraum in einen virtuellen Speicherraum **40** auf RAID-Ebene ab. Dieser virtuelle Speicherraum auf RAID-Ebene präsentiert den physischen Speicherraum als Spiegel- und Paritäts-RAID-Bereiche. Die Spiegel-RAID-Bereiche enthalten mehrere Spiegelzuweisungsblöcke, die Daten gemäß einer Spiegelredundanz oder RAID-Ebene **1** speichern. Die Paritäts-RAID-Bereiche enthalten mehrere Paritätszuweisungsblöcke, die Daten gemäß einer Paritätsredundanz oder RAID-Ebene **5** speichern.

Bei Schritt **102** bildet das RAID-Verwaltungssystem den virtuellen Speicherraum **40** auf RAID-Ebene in einen virtuellen Speicherraum **50** auf Anwendungsebene ab, der den physischen Speicherraum als mehrere virtuelle Blöcke präsentiert. Die virtuellen Blöcke werden über die in der Virtuelle-Blöcke-Tabelle aufrechterhaltenen Zeiger entsprechenden Zuweisungsblöcken in RAID-Bereichen des virtuellen Speicherraums auf RAID-Ebene zugeordnet. Wie oben unter Bezugnahme auf Fig. 4 erläutert wurde, können die virtuellen Blöcke nach Wunsch zwischen Spiegel- und Paritätsspeicherung bewegt werden.

[0063] Während des anfänglichen Abbildens bildet das RAID-Verwaltungssystem absichtlich weniger als die volle physische Kapazität des physischen Speicherraums in den ersten und den zweiten virtuellen Speicherraum ab. Auf diese Weise betragen die jeweiligen Kapazitäten des ersten und des zweiten virtuellen Speicherraums jeweils weniger als die gesamte physische Kapazität. Dies ermögliche eine logische Reservierung einer Menge an Raum, der nicht zum Speichern von Daten an den Benutzer vergeben wird (Schritt **104**). Vorzugsweise wird eine Menge an Raum reserviert, die äquivalent zu zumindest einem RAID-Bereich ist. Dieser Raum ist üblicherweise auf nicht-angrenzende Weise auf die Speicherplatten verteilt. Wenn er für eine Migration benötigt wird, wird der Raum zu ungenutzten RAID-Bereichen zusammengefasst.

[0064] Bei Schritt **106** wird die physische Kapazität (PhyCap) des physischen Speicherraums in dem Plattenarray festgestellt. Dies ist ein feststehender Wert für eine gegebene Plattenarraykonfiguration minus den reservierten Raum. Wenn das Plattenarray beispielsweise zwölf Ein-Gigabyte-Speicherplatten aufweist, wäre die physische Kapazität im Wesentlichen zwölf Gigabytes (minus die Menge bzw. den Umfang eines RAID-Bereichs). Das Datenspeichersystem leitet diesen Wert intern ab, indem es das Vorliegen und die Größe jeder Speicherplatte erfasst. Wenn eine Änderung der physischen Konfiguration auftritt, zum Beispiel wenn Speicherplatten in das System eingesteckt oder aus demselben entfernt werden, liefert das Datenspeichersystem einen neuen Wert für die physische Kapazität (PhyCap).

[0065] Es ist wünschenswert, dass jederzeit zumindest eine gewisse Spiegelspeicherung aufrechterhalten wird, um während des Cachespeicherns wichtiger Daten eine höhere Leistungsfähigkeit zu liefern. Das RAID-Verwaltungssystem legt einen willkürlichen Prozentsatz von Benutzerdaten fest, die in einer Spiegelspeicherung gespeichert werden sollen (Schritt **108**). Beispielsweise werden bis zu 10% der Benutzerdaten in einer Spiegelspeicherung behalten. Eine Speicherung in Spiegel-RAID-Bereichen ist vorteilhaft, da eine Spiegelredundanz eine höhere Leistungsfähigkeit und Zuverlässigkeit liefert als eine Paritätsredundanz.

[0066] Während dieser Systeminitialisierung sind alle virtuellen Blöcke in dem virtuellen Speicherraum **50** auf Anwendungsebene nicht zugewiesen. Die zugewiesene Kapazität wird auf Null eingestellt. Der Begriff „zugewiesene Kapazität" bedeutet die Menge an Benutzerdaten, die in den virtuellen Blöcken in dem virtuellen Speicherraum **50** auf Anwendungsebene gespeichert sind. Bei Schritt **110** berechnet das RAID-Verwaltungssystem eine maximal zulässige Anzahl virtueller Blöcke, die dem Benutzer zur Verfügung stehen, wenn gegeben ist, dass ein minimaler Prozentsatz an Benutzerdaten in Spiegelspeicherung gehalten werden soll.

[0067] Fig. 6 zeigt ein Flussdiagramm eines Verfahrens zum Gewährleisten, dass eine Spiegelspeicherung niemals die physische Speicherungskapazität einnimmt. Bei Schritt **112** werden virtuelle Datenblöcke erzeugt oder bewegt. Anfänglich, wenn Daten zuerst in dem System gespeichert werden, werden virtuelle Blöcke in dem virtuellen Speicherraum **50** auf Anwendungsebene in Spiegel-RAID-Bereichen zugewiesen. Das heißt, dass die virtuellen Blöcke über die Zeiger in der Virtuelle-Blöcke-Tabelle Spiegelzuweisungsblöcken zugeordnet werden. Die virtuellen Blöcke werden auf der Basis von Anforderungen seitens des Benutzers nach Bedarf zugewiesen. Vorzugsweise werden alle Daten in virtuellen Blöcken, die den Spiegel-RAID-Bereichen zugeordnet sind, gespeichert, bis die zugewiesene Kapazität die Hälfte der physischen Gesamtkapazität des Plattenarrays überschreitet.

[0068] Während Daten weiterhin in dem Plattenarray **11** gespeichert werden, zählt das RAID-Verwaltungssystem **16** weiterhin die virtuellen Blöcke, die derzeit Benutzerdaten speichern. Diese Zählung, die ohne weiteres anhand der Kapazität oder Größe dargestellt wird, stellt die zugewiesene Kapazität (AllCap) zu einem beliebigen gegebenen Zeitpunkt dar (Schritt **114**). Bei dem Entscheidungsblock **116** wird die zugewiesene Kapazität (AllCap) mit der physischen Kapazität (PhyCap) des Plattenarrays verglichen. Falls die zugewiesene Kapazität geringer als die oder gleich der Hälfte der physischen Kapazität ist (d.h. der „Nein"-Zweig von Schritt **116**), werden Daten weiterhin lediglich in virtuellen Blöcken in Spiegelspeicherung gespeichert (Schritt **118**).

[0069] Wenn die zugewiesene Kapazität größer als die oder gleich der physischen Kapazität ist (d.h. der „Ja"-Zweig aus Schritt **116**), berechnet das RAID-Verwaltungssystem die maximal zulässige Anzahl virtueller Blöcke in Spiegelspeicherung für die vorliegende Situation (Schritt **120**). Diese Berechnung ist eine Funktion

der zugewiesenen Kapazität (AllCap) der physischen Kapazität (PhyCap) des Plattenarrays und der Anzahl von Platten in dem Plattenarray (n). Dieser maximale Wert wird dynamisch berechnet, wenn sich die Bedingungen ändern. Das heißt, dass sich eine Variable – die zugewiesene Kapazität – üblicherweise in einem Zustand ständiger Veränderung befindet, und somit wird die maximal zulässige Anzahl virtueller Blöcke in einer Spiegelspeicherung neu berechnet, wenn sich diese Variable ändert.

[0070] Die Berechnung verwendet etwas andere Funktionen, je nachdem, ob das Plattenarray homogen oder heterogen ist. Ein „homogenes" Plattenarray besteht aus Speicherplatten gleicher Größe. Ein „heterogenes" Plattenarray besteht aus Speicherplatten unterschiedlicher Größen. Für das homogene Plattenarray berechnet das RAID-Verwaltungssystem die maximale Anzahl virtueller Blöcke in einer Spiegelspeicherung unter Verwendung der folgenden Gleichung [1]:

$$\max = AllCap - 2\left(\frac{n-1}{n-2}\right)\left(AllCap - 1/2\,PhyCap\right)$$

wobei „n" gleich der Anzahl von Speicherplatten in dem Array ist, „AllCap" die zugewiesene Kapazität darstellt und „PhyCap" die physische Kapazität des Plattenarrays darstellt.

[0071] Für ein heterogenes Plattenarray unterteilt das RAID-Verwaltungssystem das Plattenarray zunächst in Darstellungen homogen-artiger Plattenarrays. Beispielsweise sei angenommen, dass das heterogene Plattenarray aus vier Ein-Gigabyte-Speicherplatten und zwei Drei-Gigabyte-Speicherplatten besteht. Das RAID-Verwaltungssystem könnte diese Konfiguration in zwei Darstellungen homogener Plattenarrays unterteilen, bei denen eine Darstellung ein Gigabyte an Speicherung auf allen sechs Platten umfasst und die andere Darstellung die verbleibenden zwei Gigabytes der Speicherung auf den zwei Drei-Gigabyte-Platten umfasst. Das RAID-Verwaltungssystem berechnet anschließend die maximale Anzahl virtueller Blöcke in einer Spiegelspeicherung für das heterogene Plattenarray unter Verwendung folgender Gleichung [2]:

$$\max = AllCap - \sum_i 2\frac{n_i - 1}{n_i - 2}\left(AllCap_i - 1/2\,PhyCap_i\right)$$

wobei i die Anzahl von Darstellungen homogener Plattenarrays ist, die aus dem heterogenen Plattenarray erzeugt werden können.

[0072] Das RAID-Verwaltungssystem schränkt eine Zuweisung einer über die maximal zulässige Anzahl hinausgehenden Anzahl virtueller Blöcke zur Spiegelspeicherung ein. Bei Schritt **122** wird bestimmt, ob eine Neuzuweisung virtueller Blöcke oder eine Bewegung virtueller Blöcke von Paritäts- zu Spiegelspeicherung die maximal zulässige Anzahl virtueller Blöcke in Spiegelspeicherung überschreitet. Falls dies nicht der Fall ist (d.h. der „Nein"-Zweig des Schritts **122**), können Daten in den virtuellen Blöcken gespeichert werden, die der Spiegelspeicherung zugeordnet sind (Schritt **118**). Wenn dagegen die maximal zulässige Anzahl überschritten wird (d.h. der „Ja"-Zweig aus Schritt **122**), bewegt das RAID-Verwaltungssystem virtuelle Blöcke von der Spiegelzur Paritätsspeicherung. Das heißt, dass Daten von Spiegelzuweisungsblöcken zu Paritätszuweisungsblöcken in dem virtuellen Speicherraum **40** auf RAID-Ebene bewegt werden und die Zeiger in der Virtuelle-Blöcke-Tabelle aktualisiert werden, so dass die virtuellen Blöcke in dem virtuellen Speicherraum **50** auf Anwendungsebene effektiv von der Spiegelspeicherung zur Paritätsspeicherung wandern.

[0073] Das RAID-Verwaltungssystem kann unter Verwendung des reservierten Raums, dessen Vorliegen durch ein Einschränken der Anzahl virtueller Blöcke in Spiegelspeicherung gewährleistet wird, Daten von Spiegelzuweisungsblöcken zu Paritätszuweisungsblöcken bewegen. Durch dynamisches Berechnen der maximalen Anzahl virtueller Blöcke hält das RAID-Verwaltungssystem zusätzlich die Anzahl virtueller Blöcke in Spiegelspeicherung zwischen dem berechneten Maximum und dem voreingestellten Minimum.

[0074] Durch Beschränken der Anzahl virtueller Blöcke in Spiegelspeicherung verhindert das System, dass der physische Speicherraum vollständig durch derartige virtuelle Blöcke eingenommen wird. Diese Praxis verhindert effektiv, dass die ungenutzte physische Kapazität unter den logisch reservierten Umfang fällt, um dadurch zu gewährleisten, dass eine Datenbewegung zwischen Paritäts- und Spiegel-RAID-Bereichen immer verfügbar ist.

[0075] Wenn mehr Daten in dem Plattenarray gespeichert werden, nähert sich die zugewiesene Kapazität der physischen Kapazität an. Um die Benutzeranforderungen bezüglich eines Handhabens großer Datenvolumina zu berücksichtigen, fährt das RAID-Verwaltungssystem fort, Daten von Spiegel-RAID-Bereichen zu Pari-

täts-RAID-Bereichen zu bewegen. Folglich nimmt das Verhältnis von Daten, die in Spiegel-RAID-Bereichen gespeichert sind, zu Daten, die in Paritäts-RAID-Bereichen gespeichert sind, allmählich ab.

[0076] Fig. 7 und Fig. 8 veranschaulichen die praktischen Effekte des Verfahrens dieser Erfindung. Fig. 7 stellt den Fall, bei dem ein homogenes Plattenarray drei unterschiedliche Größen einer physischen Kapazität: 3, 6 und 12 Gigabytes, aufweist, graphisch dar. Die Darstellung in Form eines umgekehrten „V" für jedes Szenario stellt die maximal zulässige Anzahl virtueller Blöcke in Spiegelspeicherung, die durch die obige Gleichung [1] berechnet wurde, in Bezug auf die zunehmende Menge an Daten, die in dem Plattenarray gespeichert werden, dar. Bei jedem Szenario wächst die maximal zulässige Anzahl virtueller Blöcke in Spiegelspeicherung von Null auf die Hälfte der physischen Speicherkapazität. Anfänglich werden Daten lediglich in Spiegelspeicherung gespeichert. Nachdem die Menge an Benutzerdaten diese Hälfte-Schwelle (wie sie durch die Bruchteilmenge 0,5 angegeben ist) erreicht und überschreitet, beginnt die maximal zulässige Anzahl virtueller Blöcke in Spiegelspeicherung abzunehmen. In dem Maße, wie sich das Volumen der Datenspeicherung auf dem Plattenarray an die volle Kapazität annähert, werden immer weniger virtuelle Blöcke einer Spiegelspeicherung zugewiesen. Es sei erwähnt, dass, während die virtuellen Blöcke in Spiegelspeicherung zahlenmäßig abnehmen, mehr virtuelle Blöcke in Paritätsspeicherung verwendet werden, um Daten zu speichern. Dies geht deutlicher aus Fig. 8 hervor.

[0077] Fig. 8 stellt den Fall, bei dem ein heterogenes Plattenarray drei unterschiedliche Größen der physischen Kapazität: 6, 12 und 18 Gigabytes aufweist, graphisch dar. Es gibt drei Kurvenpaare, die diesen Drei-Größen-Szenarios zugeordnet sind, wie durch die Legende am oberen Ende des Graphen angegeben ist. Die negativ geneigten Kurven stellen des Speicherungsumfang dar, der virtuelle Blöcke in Spiegelspeicherung verwendet, und die positiv geneigten Kurven stellen die Speicherungsmenge dar, die virtuelle Blöcke in Paritätsspeicherung verwendet. Man beachte, dass die Ursprungskoordinate für die x-Achse 0,5 beträgt, was bedeutet, dass das System an diesem Punkt bereits halb gefüllt ist. Während das Plattenarray darin fortfährt, sich über die Hälfte der Kapazität hinaus mit Daten zu füllen, nimmt die maximal zulässige Anzahl virtueller Blöcke in Spiegelspeicherung stetig bis auf effektiv Null ab (obwohl es wünschenswert ist, zumindest einen geringen Prozentsatz virtueller Blöcke in Spiegelspeicherung zu halten). Zur selben Zeit nimmt die Anzahl virtueller Blöcke in Paritätsspeicherung invers auf die volle Kapazität hin zu.

[0078] Diese Graphen fassen die bevorzugten Speicherungs- und Migrationspolitik dieser Erfindung präzise zusammen. Anfänglich werden Daten gänzlich in Spiegelspeicherung gespeichert. Bei halber Kapazität beginnt das hierarchische Datenspeichersystem, unter Verwendung einer Datenbewegung zwischen den Typen die Daten sowohl in Spiegel- als auch in Paritätsspeicherung zu speichern. Während sich das Datenspeichersystem weiterhin anfüllt, werden proportional mehr Daten zur Paritätsspeicherung bewegt, um das zunehmende Datenvolumen zu berücksichtigen, während die Spiegelspeicherung abnimmt. Auf jeder dieser Stufen wird die maximal zulässige Anzahl virtueller Blöcke in Spiegelspeicherung so berechnet, dass die Anzahl existierender virtueller Blöcke in Spiegelspeicherung unter diesem Wert konsequent begrenzt ist, um eine Reservierung ausreichenden Raums für die Datenbewegung zu gewährleisten.

[0079] Die folgende Erörterung liefert typische Anwendungen der Verfahren während des Betriebs des hierarchischen Datenspeichersystems. Bei jeder Anwendung berechnet das RAID-Verwaltungssystem die maximal zulässige Anzahl virtueller Blöcke in Spiegelspeicherung zur Analyse von Migrationsmöglichkeiten. Jede Anwendung geht davon aus, dass mehr als die Hälfte des physischen Speicherraums genutzt wird.

[0080] Anwendung 1: Der Benutzer fordert an, auf einen virtuellen Block zuzugreifen, der derzeit einer Paritätsspeicherung zugeordnet ist. Diese Anforderung bewirkt eine Bewegung des virtuellen Blocks von Paritäts- zu Spiegelspeicherung. Die maximal zulässige Anzahl virtueller Blöcke in Spiegelspeicherung wird berechnet. Falls die tatsächliche Anzahl existierender virtueller Blöcke in Spiegelspeicherung gleich der maximal zulässigen Anzahl ist, werden diejenigen virtuellen Blöcke in Spiegelspeicherung, deren Zugriff am längsten zurückliegt (oder auf die am seltensten zugegriffen wird), zu Paritätsspeicherung bewegt. Dies verringert die Anzahl virtueller Blöcke in Spiegelspeicherung unter die maximal zulässige Anzahl. Danach kann auf den angeforderten virtuellen Block zugegriffen werden.

[0081] Anwendung 2: Der Benutzer fordert an, Daten in einen zuvor ungenutzten virtuellen Block zu schreiben. Das RAID-Verwaltungssystem berechnet wiederum die maximal zulässige Anzahl virtueller Blöcke in Spiegelspeicherung und vergleicht das Ergebnis mit der tatsächlichen Anzahl existierender virtueller Blöcke in Spiegelspeicherung. Falls die tatsächliche Anzahl die maximal zulässige Anzahl überschreitet, werden virtuelle Blöcke von Spiegelspeicherung zu Paritätsspeicherung bewegt, um in Spiegelspeicherung Raum frei zu machen, um die neuen Daten zu empfangen.

[0082] Anwendung 3: Der Benutzer fügt zu dem System eine Speicherplatte hinzu, um den Umfang an physischer Kapazität zu erhöhen. Wenn der physische Speicherraum von der neuen Speicherplatte integriert wird, berechnet das RAID- Verwaltungssystem eine neue maximale Anzahl auf der Grundlage der vergrößerten Kapazität. Da gewünscht ist, so viele Daten wie möglich in Spiegel-RAID-Bereichen zu speichern, werden diejenigen virtuellen Blöcke, die in jüngerer Zeit verwendet wurden (oder auf die häufiger zugegriffen wird), von Paritätsspeicherung zu Spiegelspeicherung bewegt, bis die tatsächliche Anzahl existierender virtueller Blöcke in Spiegelspeicherung die maximal zulässige Anzahl erreicht.

[0083] Die Erfindung weist viele Vorteile auf. Erstens liefert sie einen geradlinigen Lösungsansatz, eine Spiegelspeicherungszuweisung zu begrenzen. Dadurch wird die dynamischste Quelle der Überlastung in dem System beseitigt. Zweitens muss das System nicht entworfen sein, um verschachtelte Transaktionen zu unterstützen, die eine spezifische Position oder Konfiguration von Raum, der für eine Migration reserviert ist, garantieren. Drittens ermöglicht das System eine maximale Flexibilität für Speicherungstypgrößen durch eine ohne weiteres erfolgende Anpassung an homogene oder heterogene Plattenarrays. Viertens optimiert das System die Leistungsfähigkeit und Zuverlässigkeit, indem es ständig möglichst viele Daten in Spiegelspeicherung hält.

[0084] Vorschriftsgemäß wurde die Erfindung in einer Sprache beschrieben, die mehr oder weniger spezifisch für strukturelle und methodische Merkmale ist. Man sollte jedoch verstehen, dass die Erfindung nicht auf die gezeigten und beschriebenen spezifischen Merkmale beschränkt ist, da die hierin offenbarten Einrichtungen bevorzugte Formen der Umsetzung der Erfindung umfassen.

Patentansprüche

1. Ein Verfahren zum Reservieren eines Raums in einem hierarchischen Datenspeichersystem; wobei das Datenspeichersystem (**10**) ein Plattenarray (**11**) mehrerer Speicherungsplatten (**12**) umfasst, die einen physischen Speicherraum (**34**) definieren, wobei der physische Speicherraum in einen virtuellen Speicherraum (**40**) auf RAID-Ebene, der Spiegel- und Paritäts-RAID-Bereiche aufweist, abgebildet wird, wobei der virtuelle Speicherraum auf RAID-Ebene in einen virtuellen Speicherraum (**50**) auf Anwendungsebene abgebildet wird, der mehrere virtuelle Blöcke aufweist, die dem Spiegel- und dem Paritäts-RAID-Bereich zugeordnet sind; wobei das Verfahren folgende Schritte umfasst:
(a) Bereitstellen einer Anzahl (n) der Speicherungsplatten (**12**) in dem Plattenarray (**11**);
(b) Zuweisen von virtuellen Blöcken in dem virtuellen Speicherraum (**50**) auf Anwendungsebene in Spiegel-RAID-Bereichen zum Speichern von Daten gemäß einer Spiegelredundanz;
(c) Speichern von Daten in den virtuellen Blöcken und Aufrechterhalten eines Zählwerts der virtuellen Blöcke;
(d) Bereitstellen einer physischen Kapazität des physischen Speicherraums (**34**) in dem Plattenarray (**11**);
(e) Bestimmen einer zugewiesenen Kapazität, die eine Gesamtanzahl von virtuellen Blöcken darstellt, die derzeit zum Speichern von Daten verwendet werden, auf der Basis des Zählwerts der virtuellen Blöcke;
(f) dynamisches Berechnen einer maximal zulässigen Anzahl von virtuellen Blöcken in Spiegel-RAID-Bereichen als Funktion der physischen Kapazität und der zugewiesenen Kapazität; und
Beschränken einer Zuweisung von mehr virtuellen Blöcken in Spiegel-RAID-Bereichen als die berechnete maximal zulässige Anzahl von virtuellen Blöcken in Spiegel-RAID-Bereichen, um dadurch eine Raumreservierung für eine Datenwanderung zwischen Paritäts- und Spiegel-RAID-Bereichen zu gewährleisten.

2. Ein Verfahren gemäß Anspruch 1, das ferner die folgenden zusätzlichen Schritte umfasst: Zuweisen der virtuellen Blöcke in dem virtuellen Speicherraum (**50**) auf Anwendungsebene in Paritäts-RAID-Bereichen zum Speichern von Daten gemäß einer Paritätsredundanz; und Bewegen virtueller Blöcke zwischen einer Spiegel- und einer Paritätsspeicherung unter Verwendung eines reservierten Raums, der durch ein Einschränken einer Zuweisung von virtuellen Blöcken in Spiegel-RAID-Bereichen gewährleistet ist.

3. Ein Verfahren gemäß Anspruch 1, das ferner den zusätzlichen Schritt des anfänglichen Reservierens einer Menge an physischem Speicherraum, der zu zumindest einem Spiegel-RAID-Bereich äquivalent ist, umfasst.

4. Ein Verfahren gemäß Anspruch 1, das ferner die folgenden zusätzlichen Schritte umfasst: in dem Fall, dass die Anzahl von virtuellen Blöcken in Spiegel-RAID-Bereichen geringer ist als die maximal zulässige Anzahl, Speichern von Daten lediglich in den virtuellen Blöcken in Spiegel-RAID-Bereichen; und in dem Fall, dass die Anzahl von virtuellen Blöcken in Spiegel-RAID-Bereichen die maximal zulässige Anzahl überschreitet, Zuweisen der virtuellen Blöcke in dem virtuellen Speicherraum (**50**) auf Anwendungsebene in Paritäts-RAID-Bereichen zum Speichern von Daten gemäß einer Paritätsredundanz; und Speichern von Daten in virtuellen Blöcken sowohl in den Spiegel- als auch in den Paritäts-RAID-Bereichen.

5. Ein Verfahren gemäß Anspruch 1, das ferner die folgenden zusätzlichen Schritte umfasst: Zuweisen der virtuellen Blöcke in dem virtuellen Speicherraum (**50**) auf Anwendungsebene in Paritäts-RAID-Bereichen zum Speichern von Daten gemäß einer Paritätsredundanz; Speichern von Daten in virtuellen Blöcken sowohl in den Spiegel- als auch den Paritäts-RAID-Bereichen; und, während mehr Daten in dem Plattenarray gespeichert werden, so dass die zugewiesene Kapazität die physische Kapazität annähernd erreicht, Bewegen virtueller Blöcke von Spiegel-RAID-Bereichen zu Paritäts-RAID-Bereichen, um einen abnehmenden Anteil von Daten, die in Spiegel-RAID-Bereichen gespeichert sind, hin zu Daten, die in Paritäts-RAID-Bereichen gespeichert sind, zu erzeugen.

6. Ein Verfahren gemäß Anspruch 5, das ferner den zusätzlichen Schritt des Aufrechterhaltens einer minimalen Anzahl von virtuellen Blöcken in Spiegel-RAID-Bereichen umfasst.

7. Ein Verfahren gemäß Anspruch 1, das ferner die folgenden zusätzlichen Schritte umfasst: Vergleichen der zugewiesenen Kapazität mit der physischen Kapazität; und Berechnen der maximal zulässigen Anzahl bei Schritt (f) lediglich in dem Fall, dass die zugewiesene Kapazität größer als oder gleich einer Hälfte der physischen Kapazität ist.

8. Ein Verfahren gemäß Anspruch 1, das ferner folgenden Schritt umfasst:
falls die zugewiesene Kapazität größer als oder gleich einer Hälfte der physischen Kapazität ist, Durchführen der folgenden Schritte:
Berechnen einer maximal zulässigen Anzahl von virtuellen Blöcken in Spiegel-RAID-Bereichen als Funktion der physischen Kapazität und der zugewiesenen Kapazität; und
Einschränken der Zuweisung von mehr virtuellen Blöcken in Spiegel-RAID-Bereichen als die berechnete maximal zulässige Anzahl von virtuellen Blöcken in Spiegel-RAID-Bereichen;
Zuweisen der virtuellen Blöcke in dem virtuellen Speicherraum (**50**) auf Anwendungsebene in Paritäts-RAID-Bereichen zum Speichern von Daten gemäß einer Paritätsredundanz; und
Speichern von Daten in virtuellen Blöcken sowohl in den Spiegel- als auch den Paritäts-RAID-Bereichen.

9. Ein Verfahren gemäß Anspruch 8, bei dem das Abbilden des physischen Speicherraums, das in einen virtuellen Speicherraum (**40**) auf RAID-Ebene erfolgt, der Spiegel- und Paritäts-RAID-Bereiche aufweist, und bei dem das Abbilden des virtuellen Speicherraums auf RAID-Ebene, das in einen virtuellen Speicherraum (**50**) auf Anwendungsebene erfolgt, der mehrere virtuelle Blöcke aufweist, die den Spiegel- und Paritäts-RAID-Bereichen zugeordnet sind, virtuelle Speicherräume auf RAID-Ebene und Anwendungsebene ergibt, die jeweilige Kapazitäten aufweisen, die geringer sind als die Kapazität des physischen Speicherraums, um eine Menge an Raum, die nicht für eine Verwendung durch einen Benutzer gedacht ist, logisch zu reservieren.

10. Ein Verfahren gemäß Anspruch 8, das ferner den zusätzlichen Schritt des Aufrechterhaltens einer minimalen Anzahl von virtuellen Blöcken in Spiegel-RAID-Bereichen umfasst.

11. Ein Verfahren gemäß Anspruch 8, bei dem, während mehr Daten in dem Plattenarray (**11**) gespeichert werden, derart, dass die zugewiesene Kapazität an die physische Kapazität annähernd erreicht, virtuelle Blöcke von Spiegel-RAID-Bereichen zu Paritäts-RAID-Bereichen bewegt werden, um einen abnehmenden Anteil von Daten, die in Spiegel-RAID-Bereichen gespeichert sind, hin zu Daten, die in Paritäts-RAID-Bereichen gespeichert sind, zu erzeugen.

12. Ein hierarchisches Datenspeichersystem, das folgende Merkmale aufweist: ein Plattenarray (**11**), das eine Anzahl von Speicherplatten (**12**) aufweist, die einen physischen Speicherraum (**34**) definieren; eine Plattenarraysteuerung (**40**), die mit dem Plattenarray (**11**) gekoppelt ist, zum Koordinieren eines Datentransfers zu und von den Speicherplatten (**12**); ein RAID-Verwaltungssystem (**16**), das wirksam mit der Plattenarraysteuerung (**14**) gekoppelt ist, zum Abbilden eines virtuellen Speicherraums (**40**) auf RAID-Ebene in den physischen Speicherraum (**34**) der Speicherplatten; wobei der virtuelle Speicherraum (**40**) auf RAID-Ebene mehrere Spiegel- und Paritäts-RAID-Bereiche aufweist; wobei das RAID-Verwaltungssystem (**16**) virtuelle Blöcke in Spiegel-RAID-Bereichen zum Speichern von Daten gemäß einer Spiegelredundanz und virtuelle Blöcke in Paritäts-RAID-Bereichen zum Speichern von Daten gemäß einer Paritätsredundanz zuweist, wobei das RAID-Verwaltungssystem (**16**) einen Zählwert der virtuellen Blöcke aufrechterhält;
wobei das RAID-Verwaltungssystem (**16**) konfiguriert ist, um eine maximal zulässige Anzahl von virtuellen Blöcken in den Spiegel-RAID-Bereichen als Funktion der physischen Kapazität und der zugewiesenen Kapazität auf der Basis des Zählwerts von virtuellen Blöcken dynamisch zu berechnen; und
wobei das RAID-Verwaltungssystem (**16**) konfiguriert ist, um eine Zuweisung von mehr virtuellen Blöcken in Spiegel-RAID-Bereichen als die berechnete maximal zulässige Anzahl von virtuellen Blöcken in Spie-

gel-RAID-Bereichen einzuschränken, um dadurch eine Raumreservierung für eine Datenbewegung zwischen Paritäts- und Spiegel-RAID-Bereichen zu gewährleisten.

13. Ein hierarchisches Datenspeichersystem gemäß Anspruch 12, bei dem das Plattenarray (**11**) homogen ist, wobei es aus mehreren Plattenspeichern einer gleichen Größe besteht; und bei dem das RAID-Verwaltungssystem (**16**) konfiguriert ist, um die maximal zulässige Anzahl von virtuellen Blöcken in den Spiegel-RAID-Bereichen gemäß der folgenden Gleichung

$$\mathrm{max} = \mathrm{AllCap} - 2\left(\frac{n-1}{n-2}\right)\left(\mathrm{AllCap} - 1/2\ \mathrm{PhyCap}\right),$$

bei der „n" gleich der Anzahl von Speicherplatten in dem Plattenarray ist, „AllCap" die zugewiesene Kapazität darstellt und „PhyCap" die physische Kapazität des Plattenarrays darstellt, dynamisch zu berechnen.

14. Ein hierarchisches Datenspeichersystem gemäß Anspruch 12, bei dem das Plattenarray (**11**) heterogen ist und aus Speicherplatten verschiedener Größen besteht; und bei dem das RAID-Verwaltungssystem (**16**) konfiguriert ist, um die maximal zulässige Anzahl von virtuellen Blöcken in den Spiegel-RAID-Bereichen gemäß der folgenden Gleichung

$$\mathrm{max} = \mathrm{AllCap} - \sum_i 2\,\frac{n_i - 1}{n_i - 2}\left(\mathrm{AllCap}_i - 1/2\ \mathrm{PhyCap}\right),$$

bei der „n" gleich der Anzahl von Speicherplatten in dem Plattenarray ist, „AllCap" die zugewiesene Kapazität darstellt, „PhyCap" die physische Kapazität des Plattenarrays darstellt und i die Anzahl von homogenen Plattenarraydarstellungen ist, die aus dem heterogenen Plattenarray hergestellt werden können, dynamisch zu berechnen.

Es folgen 7 Blatt Zeichnungen

Anhängende Zeichnungen

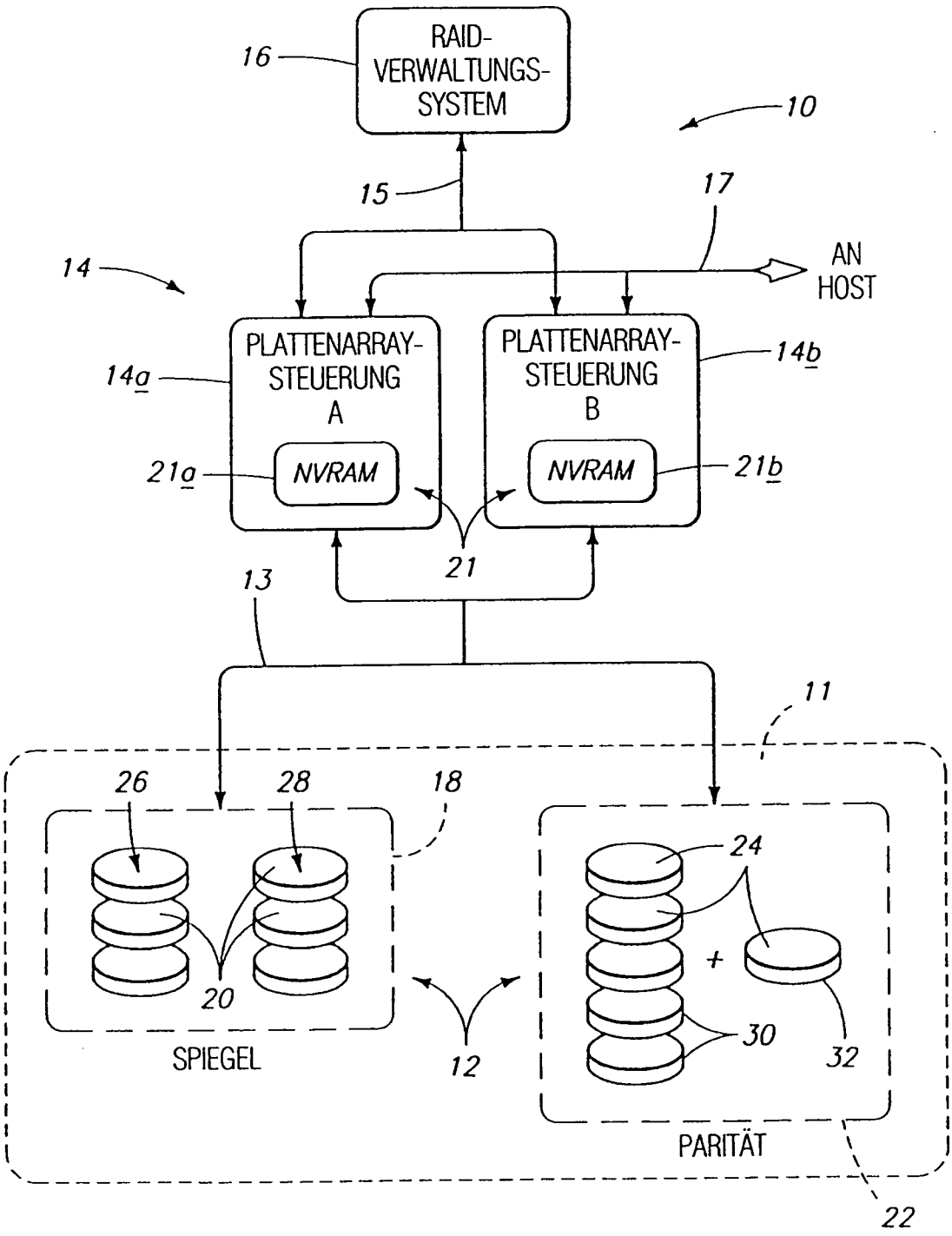


Fig 1

PLATTEN				STREIFEN-NUMMER
0	1	2	3	
0	0'	1	1'	0
2	2'	3	3'	1
4	4'	5	5'	2
6	6'	7	7'	3
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*
T–1	T–1'	T	T'	S

Fig 2

PLATTEN				STREIFEN-NUMMER
0	1	2	3	
0	1	2	P	0
3	4	P	5	1
6	P	7	8	2
P	9	10	11	3
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*
P	R–2	R–1	R	Q

Fig 3

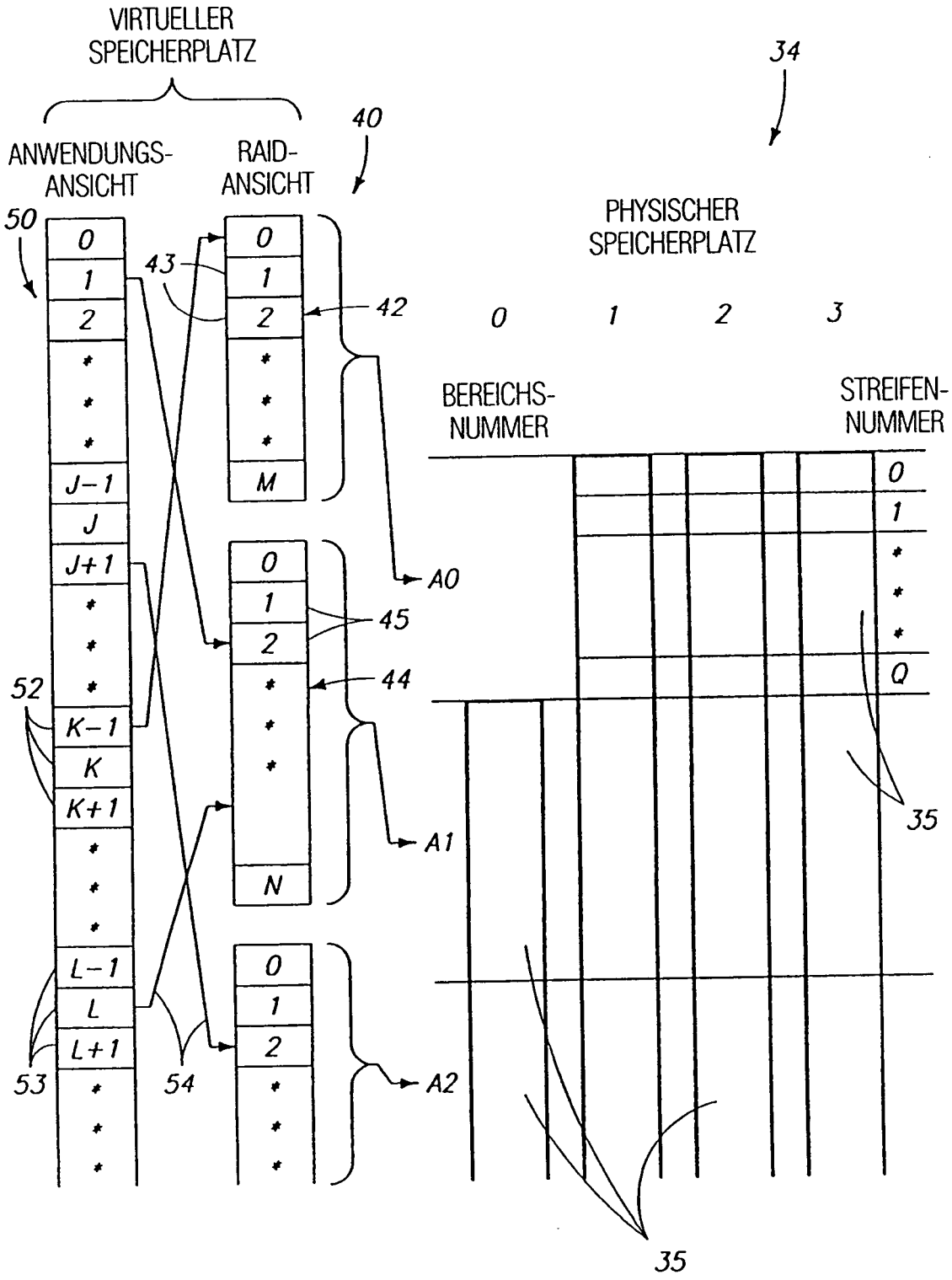
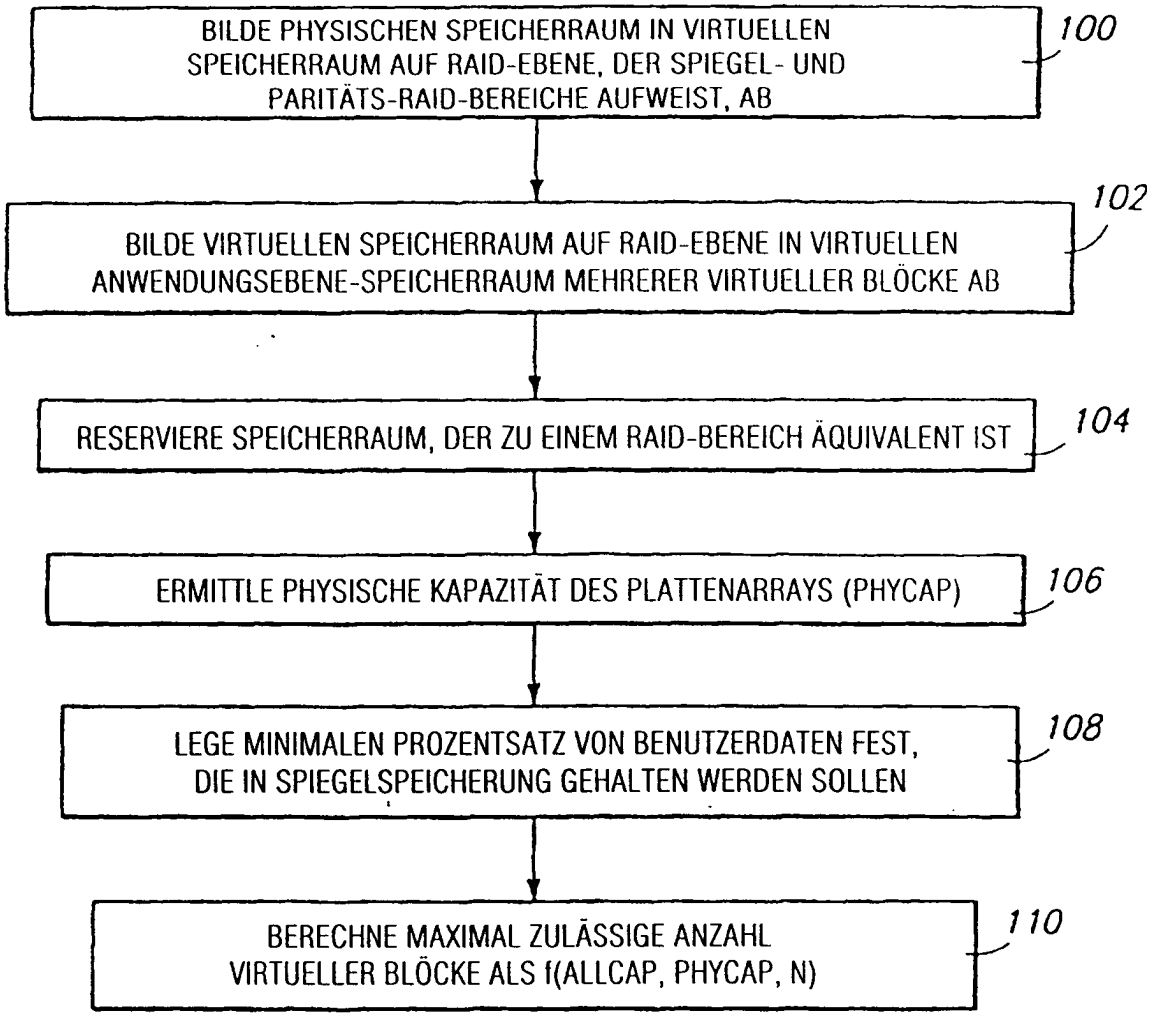
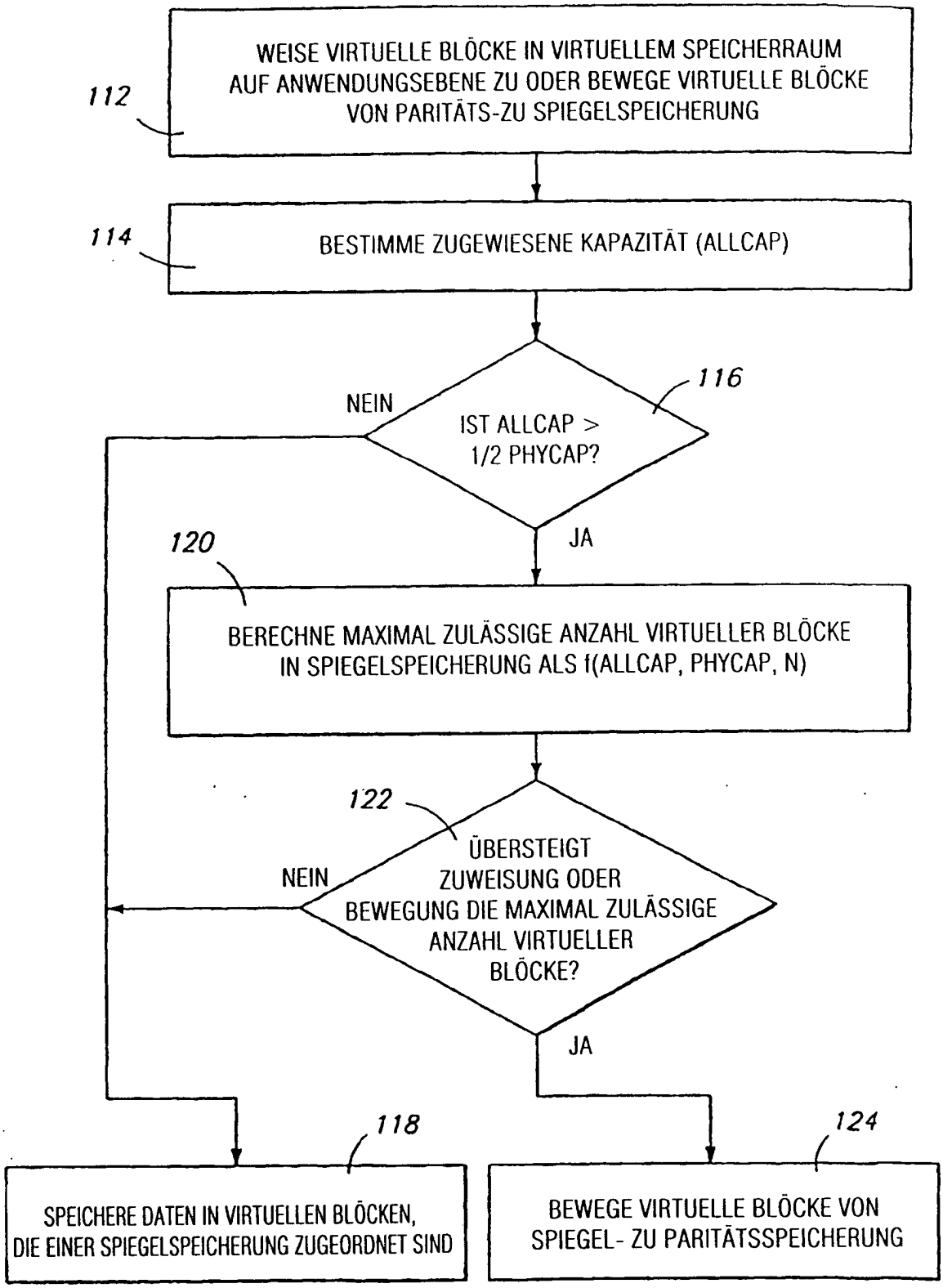

VIRTUELLER SPEICHERPLATZ
ANWENDUNGS-ANSICHT
RAID-ANSICHT
40
34
50
PHYSISCHER SPEICHERPLATZ
0
1
2
3
BEREICHS-NUMMER
STREIFEN-NUMMER
43
42
45
44
52
53
54
A0
A1
A2
35
35

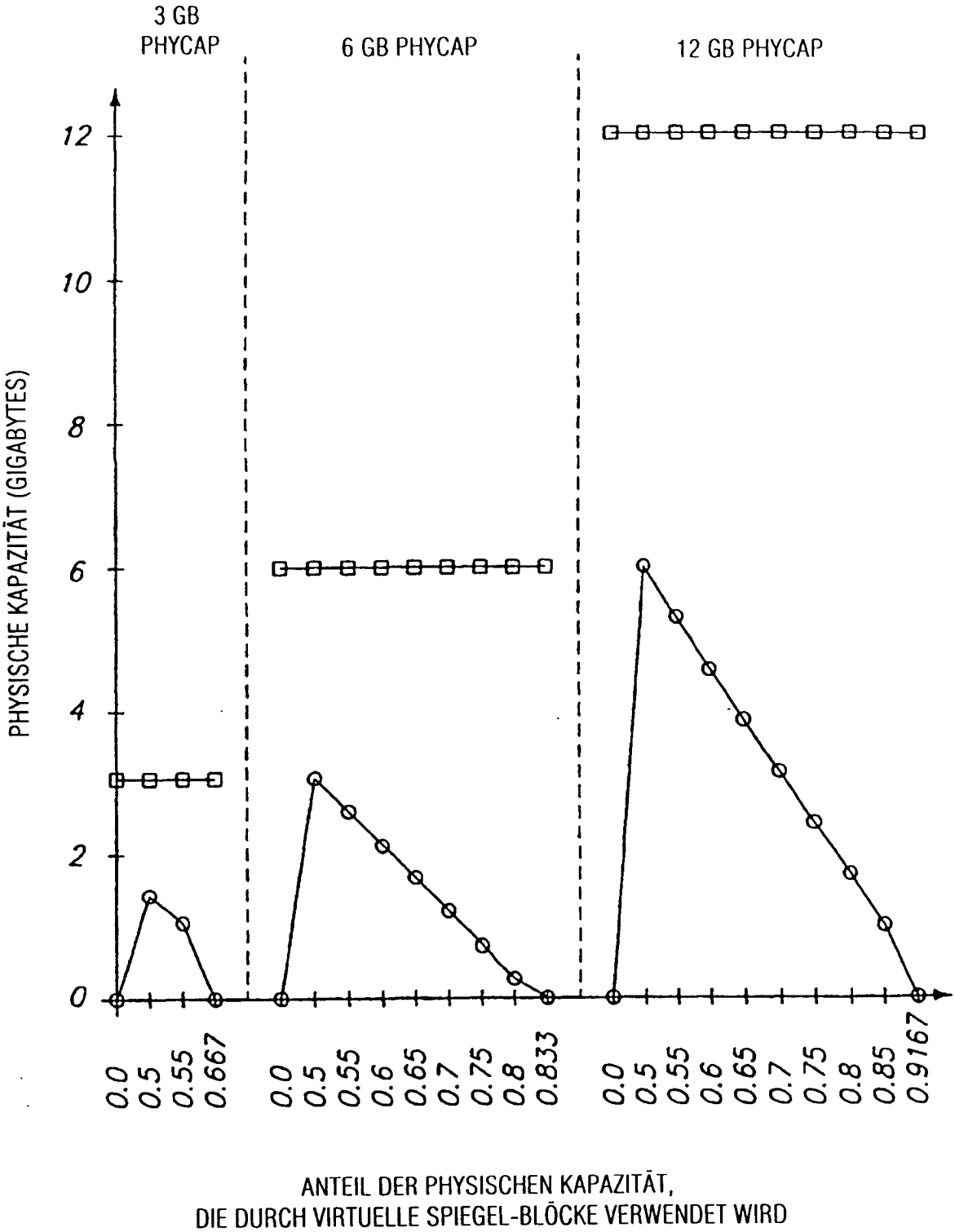
Fig 4


BILDE PHYSISCHEN SPEICHERRAUM IN VIRTUELLEN SPEICHERRAUM AUF RAID-EBENE, DER SPIEGEL- UND PARITÄTS-RAID-BEREICHE AUFWEIST, AB
100
BILDE VIRTUELLEN SPEICHERRAUM AUF RAID-EBENE IN VIRTUELLEN ANWENDUNGSEBENE-SPEICHERRAUM MEHRERER VIRTUELLER BLÖCKE AB
102
RESERVIERE SPEICHERRAUM, DER ZU EINEM RAID-BEREICH ÄQUIVALENT IST
104
ERMITTLE PHYSISCHE KAPAZITÄT DES PLATTENARRAYS (PHYCAP)
106
LEGE MINIMALEN PROZENTSATZ VON BENUTZERDATEN FEST, DIE IN SPIEGELSPEICHERUNG GEHALTEN WERDEN SOLLEN
108
BERECHNE MAXIMAL ZULÄSSIGE ANZAHL VIRTUELLER BLÖCKE ALS f(ALLCAP, PHYCAP, N)
110

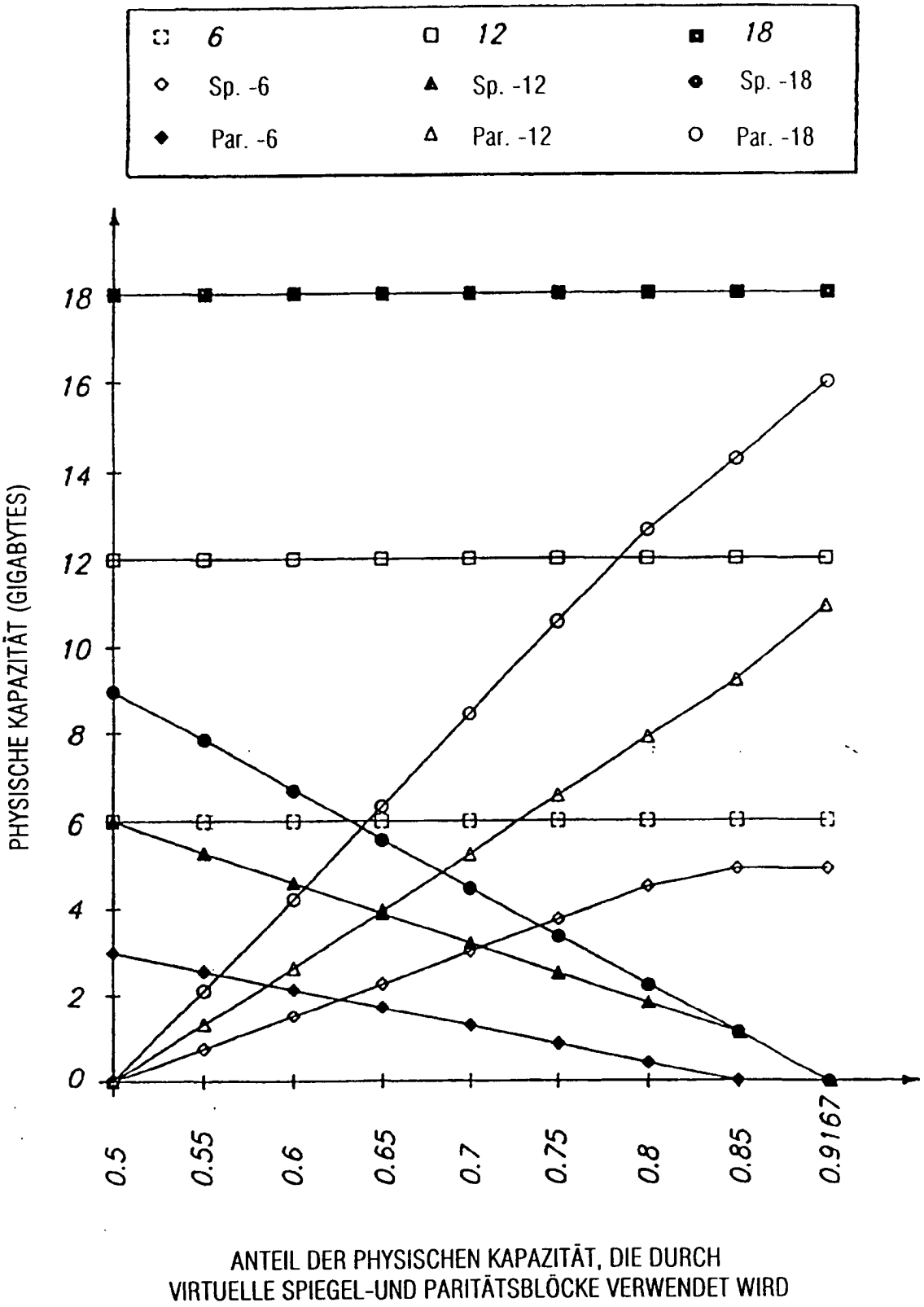
FIGUR 5


WEISE VIRTUELLE BLÖCKE IN VIRTUELLEM SPEICHERRAUM AUF ANWENDUNGSEBENE ZU ODER BEWEGE VIRTUELLE BLÖCKE VON PARITÄTS-ZU SPIEGELSPEICHERUNG
112
BESTIMME ZUGEWIESENE KAPAZITÄT (ALLCAP)
114
IST ALLCAP > 1/2 PHYCAP?
116
NEIN
JA
BERECHNE MAXIMAL ZULÄSSIGE ANZAHL VIRTUELLER BLÖCKE IN SPIEGELSPEICHERUNG ALS f(ALLCAP, PHYCAP, N)
120
ÜBERSTEIGT ZUWEISUNG ODER BEWEGUNG DIE MAXIMAL ZULÄSSIGE ANZAHL VIRTUELLER BLÖCKE?
122
NEIN
JA
SPEICHERE DATEN IN VIRTUELLEN BLÖCKEN, DIE EINER SPIEGELSPEICHERUNG ZUGEORDNET SIND
118
BEWEGE VIRTUELLE BLÖCKE VON SPIEGEL- ZU PARITÄTSSPEICHERUNG
124

FIGUR 6


3 GB PHYCAP
6 GB PHYCAP
12 GB PHYCAP
PHYSISCHE KAPAZITÄT (GIGABYTES)
12
10
8
6
4
2
0
0.0 0.5 0.55 0.667
0.0 0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.833
0.0 0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85 0.9167
ANTEIL DER PHYSISCHEN KAPAZITÄT,
DIE DURCH VIRTUELLE SPIEGEL-BLÖCKE VERWENDET WIRD

FIGUR 7


6
12
18
Sp. -6
Sp. -12
Sp. -18
Par. -6
Par. -12
Par. -18
PHYSISCHE KAPAZITÄT (GIGABYTES)
0
2
4
6
8
10
12
14
16
18
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9167
ANTEIL DER PHYSISCHEN KAPAZITÄT, DIE DURCH
VIRTUELLE SPIEGEL-UND PARITÄTSBLÖCKE VERWENDET WIRD

FIGUR 8