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

(10) **DE 695 34 057 T2** 2006.01.12

(12) # Übersetzung der europäischen Patentschrift

(97) **EP 0 725 324 B1**
(21) Deutsches Aktenzeichen: **695 34 057.3**
(96) Europäisches Aktenzeichen: **95 112 325.6**
(96) Europäischer Anmeldetag: **04.08.1995**
(97) Erstveröffentlichung durch das EPA: **07.08.1996**
(97) Veröffentlichungstag der Patenterteilung beim EPA: **09.03.2005**
(47) Veröffentlichungstag im Patentblatt: **12.01.2006**

(51) Int Cl.[8]: ***G06F 3/06*** (2006.01)
G06F 11/10 (2006.01)
G06F 11/20 (2006.01)
G11B 20/18 (2006.01)
G06F 12/12 (2006.01)

(30) Unionspriorität:
382350 **01.02.1995** **US**

(84) Benannte Vertragsstaaten:
DE, FR, GB, IT

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

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

(72) Erfinder:
Burkes, Theresa A., Meridian, Idaho 83642, US;
Diamond, Bryan M., Boise, Idaho 83704, US;
Nelson, Marvin D., Boise, Idaho 83704, US

(54) Bezeichnung: **Verfahren zur Vermeidung der Über-Zuteilung virtueller Kapazität in einem redundanten hierarchischen Datenspeichersystem**

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 Vermeiden einer Überbeanspruchung von virtuellem Speicherraum in einem redundanten hierarchischen Datenspeichersystem. Diese Erfindung bezieht sich ferner auf hierarchische Datenspeichersysteme selbst.

HINTERGRUND DER ERFINDUNG

[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 Plattearray 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; jedoch weist das Paritätsverfahren geringere Leistungsfähigkeits- und Verfügbarkeitscharakteristika als das Spiegelverfahren auf.

[0006] Die US-5,166,939 bezieht sich auf eine Datenspeichervorrichtung ein -verfahren. Die Massendatenspeichervorrichtung umfasst eine Anzahl von physischen Massendatenspeichervorrichtungen, die wirksam miteinander verbunden sind, um auf einer ersten Logikebene als eine oder mehrere logische Redundanzgruppen zu fungieren. Die Breite, Tiefe und der Redundanztyp, z. B. die gespiegelten Daten oder Prüfdaten, jeder logischen Redundanzgruppe sowie die Position redundanter Informationen in derselben sind unabhängig voneinander konfigurierbar, um gewünschte Kapazitäts- und Zuverlässigkeitsanforderungen zu erfüllen. Auf einer zweiten Logikebene werden Blöcke von Massenspeicherdaten zu einer oder mehreren logischen Datengruppen gruppiert. Eine logische Redundanzgruppe kann in mehr als eine derartige Datengruppe unterteilt werden. Die Breite, die Tiefe, die Adressierungssequenz und die Anordnung von Datenblöcken in jeder logischen Datengruppe sind unabhängig voneinander konfigurierbar, um die Massendatenspeichervorrichtung in mehrere logische Massenspeicherbereiche zu unterteilen, die jeweils potentiell unterschiedliche Bandbreiten und Betriebs-RAID-Charakteristika aufweisen. Eine dritte Logikebene zum Interagieren mit einer Anwendungssoftware eines Hostcomputer-Betriebssystems ist ebenfalls vorgesehen. Die Anwendungsebene überlagert die Datengruppen mit logischen Anwendungseinheiten, um zu ermöglichen, dass die Datengruppen – alleine oder in Kombination mit einer oder mehreren Redundanzgruppen – einer Anwendungssoftware gegenüber als eine einzige logische Speichereinheit erscheinen.

[0007] Ferner beschreibt die wissenschaftliche Veröffentlichung („Hybrid Redundancy Direct-Access Storage Device Array with Design Options" IBM TECHNICAL DISCLOSURE BULLETIN, Bd. 37, Nr. 02b, Februar 1994 (02.1994), Seiten 141–148, XP000433790, ARMONK, US) ein Redundanzdirektzugriffsspeichervorrichtungsmischarray, das aus einem Array einer Anzahl von DASDs (DASD = Direct-Access Storage Device, Direktzugriffsspeichervorrichtung), einer intelligenten Steuerung, die geeignete Puffer umfasst, und eines nichtflüchtigen RAM besteht. Das Array ist in eine Prüfsummendomäne und eine Spiegeldomäne unterteilt, die sich jeweils gleichmäßig über die Anzahl von DASDs ausbreiten. Die benutzeradressierbaren Daten werden in die Prüfsummendomäne abgebildet, um jeglichen erwünschten RAID-Typ eines Prüfsummenarrays zu liefern. Der nicht-flüchtige RAM liefert eine Adressabbildung von Adressen, die von der Prüfsummendomäne zu der Spiegeldomäne umgeleitet werden. Der nicht-flüchtige RAM muss groß genug sein, um alle Adressen der gespiegelten Domäne abzubilden. Daten werden in die Spiegeldomäne und eine in dem nicht-flüchtigen RAM gespeicherte Adressabbildung geschrieben. Zu einem späteren Zeitpunkt werden die Daten in die Prüfsummendomäne eingespeichert. Lesevorgänge erfolgen je nach Bedarf aus der Spiegeldomäne oder der Prüfsummendomäne.

ZUSAMMENFASSUNG DER ERFINDUNG

[0008] Diese Erfindung, die in den beigefügten unabhängigen Ansprüchen ausführlich definiert ist, liefert ein redundantes hierarchisches Datenspeichersystem, das Daten gemäß verschiedener Redundanztechniken speichert, um Leistungsfähigkeit und Zuverlässigkeit zu optimieren. Die Erfindung bezieht sich vorwiegend auf Verfahren zum Vermeiden einer Überbeanspruchung eines dem Benutzer zur Verfügung stehenden virtuellen Speicherraums.

[0009] 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 einen virtuellen Speicherraum auf RAID-Ebene auf den physischen Speicherraum der Speicherplatten abzubilden. Der virtuelle Speicherraum auf RAID-Ebene präsentiert den physischen Speicherraum als mehrere RAID-Bereiche, wo sich einzelne RAID-Bereiche über mehrere Speicherplatten hinweg erstrecken. Die RAID-Bereiche umfassen Spiegel-RAID-Bereiche, die Spiegelzuweisungsblöcke zum Speichern von Daten gemäß einer Spiegelredundanz enthalten, und Paritäts-RAID-Bereiche, die Paritätszuweisungsblöcke zum Speichern von Daten gemäß einer Paritätsredundanz enthalten. Ein Speicherabbildungsspeicher ist wirksam mit dem RAID-Verwaltungssystem gekoppelt, um Abbildungsinformationen, die durch das RAID-Verwaltungssystem verwendet werden, um den virtuellen Speicherraum auf RAID-Ebene auf die Platten in dem Plattenarray zu speichern, dauerhaft zu speichern. Das RAID-Verwaltungssystem migriert bzw. bewegt Daten zwischen den Spiegel- und Paritäts-RAID-Bereichen auf eine dynamische Weise, die die Leistungsfähigkeit und Datenzuverlässigkeit optimiert. Für diese Migration ist ein gewisser Umfang an virtuellem Speicherraum auf RAID-Ebene reserviert.

[0010] Um eine Überbeanspruchung des virtuellen Speicherraums auf RAID-Ebene zu vermeiden, berechnet das RAID-Verwaltungssystem eine virtuelle Gesamtkapazität, die als Referenz für zukünftige Speicheranforderungen verwendet werden kann. Diese Berechnung beruht auf einer theoretischen Schlimmster-Fall-Anordnung des Spiegel- und des Paritätsspeicherbereichs. Das RAID-Verwaltungssystem sieht zuerst vor, dass ein minimaler Prozentsatz an Benutzerdaten in Spiegelspeicherbereichen behalten wird und dass der verbleibende Prozentsatz an Benutzerdaten in Paritätsspeicherbereichen behalten wird. Die virtuelle Gesamtkapazität des virtuellen Speicherraums auf Anwendungsebene, die einem Benutzer überlassen werden kann, wird auf der Basis eines Vorsehens dessen, dass der minimale Prozentsatz an Benutzerdaten in Spiegel-RAID-Bereichen gehalten wird und dass der verbleibende Prozentsatz an Benutzerdaten in Paritäts-RAID-Bereichen behalten wird, berechnet. Auf diese virtuelle Gesamtkapazität wird verwiesen, wenn zusätzliche Speicheranforderungen des Systems unternommen werden, um zu bestimmen, ob den Anforderungen stattgegeben werden kann, ohne den virtuellen Speicherraum überzubeanspruchen.

[0011] Ferner ist das RAID-Verwaltungssystem entworfen, um schnell auszuwerten, ob eine neue Speicheranforderung berücksichtigt werden kann, ohne auf Raumverwaltungstechniken wie z. B. ein Bewegen von Daten zwischen Spiegel- und Paritätsspeicher zurückgreifen zu müssen, um Raum für die neue Speicheranforderung zu schaffen. Diese Auswertung erfolgt, indem einzelne RAID-Bereiche untersucht werden, um etwaige ungenutzte Zuweisungsblöcke zu identifizieren, und indem die Anzahl ungenutzter Zuweisungsblöcke gezählt wird. Der Anforderung, einen neuen Zuweisungsblock zuzuweisen, wird stattgegeben, wenn die Anzahl ungenutzter Zuweisungsblöcke größer als der oder gleich dem reservierten Speicherraum ist, der für eine Migration reserviert wird.

Beschreibung der Zeichnungen

[0012] 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.

[0013] Fig. 1 ist ein schematisches Blockdiagramm eines Speichersystems gemäß dieser Erfindung;

[0014] Fig. 2 ist eine schematische Veranschaulichung, die eine Datenspeicherung auf RAID-Ebene 1 zeigt;

[0015] Fig. 3 ist eine schematische Veranschaulichung, die eine Datenspeicherung auf RAID-Ebene 5 zeigt;

[0016] Fig. 4 ist eine schematische Veranschaulichung einer Speicherabbildungsanordnung dieser Erfindung, bei der zwei virtuelle Speicherräume auf einen physischen Speicherraum abgebildet werden;

[0017] Fig. 5 ist ein Flussdiagramm eines allgemeinen Verfahrens zum Vermeiden einer Überbeanspruchung von virtuellem Speicherraum gemäß einem Aspekt dieser Erfindung;

[0018] Fig. 6 zeigt verschiedene Schritte, die anschließend an das bei Fig. 5 gezeigte allgemeine Verfahren unternommen werden können, um eine Überbeanspruchung zu vermeiden;

[0019] Fig. 7 ist eine schematische Veranschaulichung von Speicherraum auf mehreren Speicherplatten, und zeigt die Funktionsweise des Verfahrens der Fig. 5;

[0020] Fig. 8 ist ein Flussdiagramm eines ausführlicheren Verfahrens zum Vermeiden einer Überbeanspruchung von virtuellem Speicherraum gemäß einem weiteren Aspekt dieser Erfindung; und

[0021] Fig. 9 ist ein Flussdiagramm eines Verfahrens zum Betreiben eines redundanten hierarchischen Plattenarrays beim Handhaben von Speicheranforderungen gemäß einem weiteren Aspekt dieser Erfindung.

AUSFÜHRLICHE BESCHREIBUNG DER ERFINDUNG

[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 Hostcompu-

ter (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 Speicherbereiche 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 ers-

ten 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 Speicherraum 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 Speicherraum 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 Spei-

cherraum 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 adressier-

baren 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 werden.

[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 verwendet:

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] Wenn die virtuellen Ansichten abgebildet und zur Verwendung durch den Benutzer präsentiert werden, ist es wünschenswert, dass das Datenspeichersystem Speicherressourcen, die nicht zur Verfügung stehen, nicht überbeansprucht. Die Vermeidung einer Überbeanspruchung wird bei diesem redundanten hierarchischen Datenspeichersystem dadurch verkompliziert, dass Daten zwischen Spiegel- und Paritäts-RAID-Bereichen migriert bzw. bewegt werden können. Demgemäss überwacht das Datenspeichersystem dieser Erfindung seine Speicherressourcen, um eine Überbeanspruchung der virtuellen Speicherräume bezüglich des Benutzers zu verhindern.

[0059] Fig. 5 veranschaulicht ein allgemeines Verfahren zum Vermeiden einer Überbeanspruchung von virtuellem Speicherraum. Dieses Verfahren wird unter Bezugnahme auf Fig. 7 beschrieben, die ein Plattenarray **90** mit acht Speicherplatten 0 – 7 zeigt. Kurz gesagt leitet das Verfahren die virtuelle Gesamtkapazität des Speicherraums auf Anwendungsebene, der einem Benutzer überlassen werden kann, ab.

[0060] Schritte **60** und **62** stellen die virtuelle Ansicht **40** auf RAID-Ebene und den virtuellen Speicherraum **50** auf Anwendungsebene gemäß der obigen Beschreibung unter Bezugnahme auf Fig. 4 her. Bei den verbleibenden Schritten **64** – **68** berechnet das RAID-Verwaltungssystem die maximale Kapazität, die für den virtuellen Speicherraum auf Anwendungsebene verfügbar ist, wenn gegeben ist, dass ein minimaler Prozentsatz in dem Spiegelspeicher behalten werden muss. Bei Schritt **64** geht das RAID-Verwaltungssystem davon aus, dass ein Minimum der Benutzerdaten in dem Spiegelspeicher behalten werden soll. Ein beispielhafter minimaler Prozentsatz beträgt 10%. Das heißt, dass zumindest 10% der Anzahl von Blöcken in dem virtuellen Speicherraum auf Anwendungsebene als Blöcke einer Spiegelspeicherung zugewiesen werden. Man nimmt an, dass der verbleibende Prozentsatz an Benutzerdaten in Paritäts-RAID-Bereichen behalten wird (Schritt **66**). Diese theoretische Teilung von 10%/90% von Spiegel- und Paritäts-RAID-Bereichen wird verwendet, um die virtuelle Gesamtkapazität zu berechnen, da sie eine pessimistische Annahme liefert, dass sich das Plattenarray auf höchster Speicherkapazität befindet.

[0061] Eine weitere Annahme betrifft ein theoretisches Layout der Spiegel-RAID-Speicherung und der Paritäts-RAID-Speicherung. Im Einzelnen geht das RAID-Verwaltungssystem zum Zweck einer Berechnung einer virtuellen Gesamtkapazität davon aus, dass die Spiegel-RAID-Bereiche in physische Plattenbereiche abgebildet werden, die sich über alle Speicherplatten in dem Array erstrecken. Dies ist in Fig. 7 gezeigt, wo die Spiegel-RAID-Bereiche theoretisch an der Unterseite des Plattenarrays positioniert sind. Die Paritäts-RAID-Bereiche sind dann theoretisch in der verbleibenden virtuellen Ansicht auf RAID-Ebene platziert, die auf den oberen und möglicherweise nicht-einheitlichen Teil des Plattenarrays abbildet. Beispielhafte RAID-Bereiche RA_H, RA_{H+1} usw. sind als zusammenhängende Regionen R der Speicherplatten 0 – 7 abgebildet veranschaulicht.

[0062] Nachdem die Spiegel- und Paritäts-RAID-Bereiche theoretisch getrennt und in dem Plattenarray platziert wurden, kann das RAID-Verwaltungssystem die virtuelle Gesamtkapazität des virtuellen Speicherraums auf Anwendungsebene, der einem Benutzer überlassen werden kann, berechnen (Schritt **68**). Diese Berechnung beruht auf der potentiellen Quantität an virtuellem Speicherraum, der von dem minimalen Prozentsatz an Spiegel-RAID-Bereichen eingenommen werden soll, und auf der potentiellen Quantität an virtuellem Speicherraum, der von dem verbleibenden Prozentsatz an Paritäts-RAID-Bereichen eingenommen werden soll. Nachdem die virtuelle Gesamtkapazität ermittelt wurde, kann dieser Wert für zahlreiche andere, durch das Datenspeichersystem getroffene Bestimmungen verwendet werden, wie durch den Optionen-Schritt **70** dargestellt ist. Derartige Bestimmungen umfassen eine Berechnung einer nicht-überlassenen Kapazität, eine Auswertung einer Anforderung bezüglich einer neuen Quantität an virtuellem Speicherraum, eine Auswertung dessen, ob eine Speicherplatte entleert werden kann, ein Überwachen der Kapazitätserschöpfung und eine Auswertung dessen, ob eine Platte ohne schädliche Auswirkungen auf das System verloren gehen kann. Diese Bestimmungsprozesse werden nachfolgend unter Bezugnahme auf Fig. 6 ausführlicher beschrieben.

[0063] Bevor diese Teilbestimmungen adressiert werden, wird jedoch unter Bezugnahme auf Fig. 7 und Fig. 8 eine ausführlichere Erläuterung der primären Berechnung der virtuellen Gesamtkapazität erläutert. Das RAID-Verwaltungssystem definiert zuerst einen oder mehrere Sätze i in dem virtuellen Speicherraum auf RAID-Ebene (Schritt **100**). Jeder Satz enthält eine Gruppe von RAID-Bereichen, die sich über dieselbe Anzahl von Speicherplatten erstrecken. Bei einem „homogenen" Plattenarray, bei dem alle Speicherplatten dieselbe Größe aufweisen, gibt es nur einen Satz. Alle RAID-Bereiche weisen die gleiche Größe auf und erstrecken sich über alle Speicherplatten der gleichen Größe. Umgekehrt gibt es bei einem „heterogenen" Plattenarray, bei dem die Speicherplatten unterschiedliche Größen aufweisen, mehr als einen Satz.

[0064] Fig. 7 zeigt ein heterogenes Plattenarray, bei dem die Platten 0, 2, 3 und 7 dieselbe Größe aufweisen; die Platten 1, 5 und 6 dieselbe Größe aufweisen, jedoch größer sind als die Platten 0, 2, 3 und 7; und die Platte 4 die größte Platte ist. Bei diesem Plattenarray können drei Sätze definiert werden. Satz 0 umfasst die Gruppe von RAID-Bereichen, die auf alle acht Speicherplatten abbilden. Satz 0 deckt den Großteil des Speicherraums ab. Satz 1 umfasst die Gruppe von RAID-Bereichen, die auf vier Platten 1, 4, 5 und 6 abbilden. Satz 2 umfasst die Gruppe von RAID-Bereichen, die auf den obersten Abschnitt der Platte 4 abbilden.

[0065] Das RAID-Verwaltungssystem platziert die Spiegel-RAID-Bereiche theoretisch in einem oder mehreren Sätzen, die die Gruppe von RAID-Bereichen mit der größten Anzahl an Speicherplatten enthalten (Schritt **102**). Anders gesagt geht das RAID-Verwaltungssystem für die Zwecke des Berechnens der virtuellen Gesamtkapazität davon aus, dass die Spiegel-RAID-Bereiche in den unteren Abschnitt des Plattenarrays abgebildet werden, der alle Speicherplatten umspannt, wie in Fig. 7 gezeigt ist. Die Paritäts-RAID-Bereiche werden theoretisch in den verbleibenden virtuellen Speicherraum auf RAID-Ebene platziert.

[0066] Nachdem die Spiegel- und die Paritäts-RAID-Bereiche theoretisch getrennt wurden, enthält einer der definierten Sätze sowohl Spiegel- als auch Paritäts-RAID-Bereiche. Bei dem heterogenen Array der Fig. 7 ist Satz 0 der gemeinsame Satz, der sowohl Spiegel- als auch Paritäts-RAID-Bereiche enthält. Üblicherweise ist der Satz 0 der größte Satz und ist in der Regel der gemeinsame Satz. Bei einem homogenen Plattenarray enthält der einzige Satz selbstverständlich beide Arten von RAID-Bereichen. Bei ungeraden heterogenen Plattenarrays, die z. B. eine kleine Platte und mehrere sehr große Platten aufweisen, kann jedoch ein anderer Satz als der Satz 0 sowohl Spiegel- als auch Paritätsspeicherung enthalten.

[0067] Das RAID-Verwaltungssystem platziert den gemeinsamen Satz, mit G bezeichnet, für eine gegebene Plattenarraykonfiguration, indem es die folgende Beziehung [1] nach dem kleinsten G auflöst:

$$\sum_{i=0}^{G} \mathrm{VirCapacMirSet}(i) \geq \frac{C}{100 - C} \sum_{i=G+1}^{K} \mathrm{VirCapacParSet}(i)$$

wobei „VirCapacMirSet" die virtuelle Kapazität der Spiegel-RAID-Bereiche in dem Satz i darstellt, „VirCa-

pacParSet" die virtuelle Kapazität der Paritäts-RAID-Bereiche in dem Satz i darstellt, K die Gesamtanzahl von Sätzen darstellt und C der minimale Prozentsatz für Spiegel-RAID-Bereiche ist (als Ganzzahl ausgedrückt, z. B. 10 für 10%).

[0068] Nachdem der gemeinsame Satz G platziert wurde, kann das RAID-Verwaltungssystem die Kapazitäten der verschiedenen Sätze von RAID-Bereichen berechnen. Da manche Sätze entweder Spiegel- oder Paritätsbereiche, jedoch nicht beides, enthalten, ist die Berechnung bei diesen Sätzen einfach eine Funktion entweder der Quantität an virtuellem Speicherraum, der von Spiegel-RAID-Bereichen eingenommen wird (als MVT bezeichnet) oder der Quantität an virtuellem Speicherraum, der von Paritäts-RAID-Bereichen eingenommen wird (als PVT bezeichnet). Lediglich der gemeinsame Satz G, der sowohl Spiegel- als auch Paritäts-RAID-Bereiche enthält, stellt eine kompliziertere Berechnung auf der Basis einer Funktion sowohl von MVT als auch PVT dar. Nachdem die Kapazitäten jedes Satzes festgestellt wurden, ist die virtuelle Gesamtkapazität für den virtuellen Speicherraum auf Anwendungsebene einfach die Summe dieser Satz-Kapazitäten.

[0069] Bei Schritt **106** berechnet das RAID-Verwaltungssystem die Kapazität der Sätze, die lediglich Spiegel-RAID-Bereiche enthalten, falls vorhanden. Die Quantität an virtueller Kapazität, die von Spiegel-RAID-Bereichen eingenommen wird (MVT), ist durch die folgende Gleichung [2] gegeben:

$$MVT = \tfrac{1}{2}MR \cdot N$$

wobei MR die Anzahl von Spiegel-RAID-Bereichen ist und N die Anzahl von Speicherplatten in dem Satz ist. Bei der Konfiguration der Fig. 7 gibt es jedoch keine Sätze, die lediglich Spiegel-RAID-Bereiche enthalten.

[0070] Bei Schritt **108** berechnet das RAID-Verwaltungssystem die Kapazität von Sätzen, die lediglich Paritäts-RAID-Bereiche enthalten, falls vorhanden. Die Quantität an virtueller Kapazität, die von Paritäts-RAID-Bereichen eingenommen wird (PVT), ist durch die folgende Gleichung [31 gegeben:

$$PVT = PR(N - 1)$$

wobei PR die Anzahl von Paritäts-RAID-Bereichen ist. Bei Fig. 7 enthalten Sätze 1 und 2 lediglich Paritäts-RAID-Bereiche, und die Kapazitäten dieser Sätze werden durch Verwendung dieser Gleichung bestimmt.

[0071] Bei Schritt **110** berechnet das RAID-Verwaltungssystem die Kapazität des gemeinsamen Satzes G als Funktion von sowohl MVT als auch PVT. Diese Berechnung ist komplizierter, da zuerst die Anzahl von Spiegel-RAID-Bereichen MR und die Anzahl von Paritäts-RAID-Bereichen PR in dem Satz G ermittelt werden müssen. Die Anzahl von Spiegel-RAID-Bereichen MR ist durch die folgende Gleichung [4] gegeben:

$$MR = \frac{2C * V51 + 2C(N - 1)L - 2(100 - C)V11}{(100 - C)N + 2C(N - 1)}$$

wobei V51 die virtuelle RAID5-Kapazität in allen Sätzen außerhalb des Satzes G darstellt (in einer Speicherungseinheit angegeben, die äquivalent zu einer Region ist, die sich auf einer einzigen Platte innerhalb eines RAID-Bereichs befindet), V11 die virtuelle RAID1-Kapazität in allen Sätzen außerhalb des Satzes G darstellt (ebenfalls in einer Speicherungseinheit angegeben, die äquivalent zu einer Region ist, die sich auf einer einzigen Platte innerhalb eines RAID-Bereichs befindet), N die Anzahl von Platten in dem Satz G darstellt und L die Anzahl von RAID-Bereichen in dem Satz G darstellt. Nachdem die Anzahl von Spiegel-RAID-Bereichen MR berechnet wurde, wird die Anzahl von Paritäts-RAID-Bereichen PR in dem Satz G durch die Gleichung [5] berechnet:

$$PR = G - MR$$

[0072] Diese Werte für MR und PR werden in den obigen Gleichungen [2] und [3] verwendet. Die Quantität an virtueller Kapazität, die sowohl von Spiegel- als auch Paritäts-RAID-Bereichen eingenommen wird, ist somit durch die folgende Gleichung [6] gegeben:

virtuelle Kapazität des Satzes G =
$MVT + PVT = \tfrac{1}{2}MR \cdot N + PR(N - 1)$

[0073] Nachdem die Kapazitäten jedes Satzes berechnet wurden, kann die virtuelle Gesamtkapazität an virtuellem Speicherraum auf Anwendungsebene als Summe dieser Kapazitäten berechnet werden (Schritt **112**).

Im Einzelnen wird die virtuelle Gesamtkapazität TVC gemäß der folgenden Beziehung [7] abgeleitet:

$$TVC = \left[\tfrac{1}{2} MR * N_G + PR(N_G - 1)\right]_{innen} + \sum_{s<G}\left[\tfrac{1}{2} MR * N_S\right]_{aussen} +$$

$$\sum_{s>G}\left[PR(N_S - 1)\right]_{aussen}$$

wobei die Tiefstellung „innen" bedeutet, dass die Spiegel- und Paritäts-RAID-Bereiche innerhalb des gemeinsamen Satzes G liegen, N_G die Anzahl von Platten in dem Satz G darstellt, N_S die Anzahl von Platten in jenen Sätzen, die nicht der gemeinsame Satz G sind, darstellt, und die Tiefstellung „außen" bedeutet, dass die Spiegel- und Paritäts-RAID-Bereiche in anderen Sätzen als dem gemeinsamen Satz G angeordnet sind. Für Sätze außerhalb des gemeinsamen Satzes G werden zwei Summierungen verwendet, da man theoretisch davon ausgeht, dass diejenigen Sätze auf einer Seite des Satzes G, falls vorhanden, lediglich Spiegel-RAID-Bereiche enthalten, und dass diejenigen Sätze auf der anderen Seite des Satzes G, falls vorhanden, lediglich Paritäts-RAID-Bereiche enthalten.

[0074] Unter erneuter Bezugnahme auf Fig. 5 können viele andere Systembestimmungen getroffen werden, wie durch den Optionen-Block **70** dargestellt ist, nachdem die virtuelle Gesamtkapazität für den virtuellen Speicherraum auf Anwendungsebene bestimmt wurde. Diese Optionen sind in Fig. 6 gezeigt. Für manche der Optionen, nämlich Optionen **74** – **82**, bestimmt das Datenspeichersystem, ob eine „Heiße-Reserve-Option" durch den Benutzer aktiviert ist (Schritt **72**). Wenn die „Heiße-Reserve-Option" aktiviert ist, reserviert das Datenspeichersystem die Kapazität der größten Speicherplatte, die für den Fall, dass eine andere Speicherplatte ausfällt, verwendet werden soll. Herkömmliche bekannte Datenspeichersysteme benennen üblicherweise eine ganze Speicherplatte als sogenannte „Heiße-Reserve"-Platte. Die Heiße-Reserve-Platte ist an einer mechanischen Bucht und dem SCSI-Bus befestigt und ist in dem Plattenarray voll funktionstüchtig, bleibt jedoch außer Betrieb, bis eine Platte ausfällt. Auf einen Ausfall einer Platte hin wird die Heiße-Reserve-Platte verwendet, um zugängliche Benutzerdaten zu Zwecken der Rekonstruktion bzw. des Wiederaufbaus der Platte zu speichern.

[0075] Gemäß dem Datenspeichersystem dieser Erfindung wird ein Umfang an Speicherraum, der zu der Kapazität der größten Speicherplatte in dem Array äquivalent ist, über alle Speicherplatten in dem Array verteilt. Keine einzelne Speicherplatte wird als die „Heiße-Reserve"-Platte bezeichnet. Alle Platten in dem Array speichern zugängliche Benutzerdaten. Die erhöhte Verteilung verbessert die Leistungsfähigkeit beträchtlich, indem sie die Parallelität erhöht und indem sie mehr Kapazität für eine Spiegelspeicherung liefert.

[0076] Falls die „Heiße-Reserve-Option" aktiviert ist, ist die Kapazität des physischen Speicherraums zum Zweck des Berechnens der maximalen Kapazität des virtuellen Speicherraums auf Anwendungsebene, der dem Benutzer überlassen werden kann, um die Kapazität der größten Speicherplatte in dem Plattenarray verringert (Schritt **72**). Die größte Speicherplatte wird verwendet, da dies gewährleistet, dass der Rekonstruktionsprozess durchgeführt werden kann, falls irgendeine Speicherplatte, einschließlich der größten, in dem Array ausfällt. Bei einem „homogenen" Plattenarray, bei dem alle Speicherplatten dieselbe Größe aufweisen, kann die Kapazität einer beliebigen Speicherplatte verwendet werden. Unter Bezugnahme auf Fig. 7 steht die Kapazität der Speicherplatte 4 (der größten Platte in dem Array **90**) nicht für die maximale Kapazität des virtuellen Speicherraums auf Anwendungsebene zur Verfügung, der in den virtuellen Speicherraum auf RAID-Ebene abgebildet werden kann. Wenn dagegen die „Heiße-Reserve-Option" nicht aktiviert ist, wird der maximale virtuelle Speicherraum nicht durch die Reservierung der physischen Kapazität der größten Speicherplatte verringert.

Option 1: Nicht-überlassene Kapazität

[0077] Bei Schritt **74** berechnet das Datenspeichersystem „die nicht-überlassene Kapazität", die verbleibende Kapazität an virtuellem Speicherraum, der nicht einer Benutzung seitens des Benutzers überlassen wird. Um die nicht-überlassene Kapazität zu berechnen, subtrahiert das RAID-Verwaltungssystem die überlassene Kapazität, die einer Benutzung seitens des Benutzers überlassen wurde, von der virtuellen Gesamtkapazität. Die nicht-überlassene Kapazität kann somit durch die folgende Gleichung [8] dargestellt werden:
Nicht-überlassene Kapazität =
bei einer Reservierung eines Heiße-Reserve-Raums,
TVC ohne die größte Plattenvorrichtung – überlassene Kapazität
andernfalls
TVC – überlassene Kapazität.

[0078] Man beachte, dass, wenn die Heiße-Reserve-Option aktiviert ist, die virtuelle Gesamtkapazität (TVC) unter der Annahme abgeleitet wird, dass die Kapazität der größten Speicherplatte aus dem Plattenarray entfernt wurde.

Option 2: Anforderung eines neuen virtuellen Speicherraums

[0079] Ein Benutzer kann anfordern, einen neuen virtuellen Speicherraum einer gegebenen Größe zu definieren (Schritt **76**). Das Datenspeichersystem billigt eine Anforderung bezüglich eines neuen virtuellen Speicherraums, wenn der neue Raum geringer als die oder gleich der Differenz zwischen der virtuellen Gesamtkapazität und der überlassenen Kapazität ist. Genauso gut könnte man sagen, dass die Anforderung gebilligt wird, wenn der neue virtuelle Speicherraum geringer als die oder gleich der nicht-überlassenen Kapazität ist. Dies wird durch die folgende Beziehung [9] wiedergegeben:
Billige Anforderung bezüglich neuem virtuellen Raum, wenn:
falls ein Heiße-Reserve-Raum reserviert wird,
Neuer Raum ≤ TVC ohne größte Speicherplatte – überlassene Kapazität
andernfalls
Neuer Raum ≤ TVC – überlassene Kapazität.

Option 3: Anforderung, Speicherplatte zu entleeren

[0080] Es kann eine Anforderung gemacht werden, Daten aus einer Speicherplatte zu entleeren (Schritt **78**). Der „Entleerungs"-Vorgang entfernt alle Daten von der Speicherplatte und speichert sie an anderen Stellen. Für die neue Speicherkonfiguration auf den anderen Platten wird eine Redundanz regeneriert. Die entleerte Speicherplatte kann dann aus dem Array entfernt werden. Die Anforderung eines Entleerens einer Speicherplatte wird gebilligt, wenn die virtuelle Gesamtkapazität, ohne die Kapazität der zu entleerenden Speicherplatte, und minus die überlassene Kapazität größer oder gleich null ist, wie durch die Beziehung [10] dargestellt ist:
Anforderung, eine Speicherplatte zu entleeren, wird gebilligt, wenn:
falls ein Heiße-Reserve-Raum reserviert wird,
TVC ohne Kapazität der zu entleerenden Speicherplatte und ohne größte Speicherplatte – überlassene Kapazität ≥ 0
andernfalls
TVC ohne Kapazität der zu entleerenden Speicherplatte – überlassene Kapazität ≥ 0.

Option 4: Warnung eines Benutzers vor einer Kapazitätserschöpfung

[0081] Das Datenspeichersystem verwendet die bei Gleichung [7] berechnete virtuelle Gesamtkapazität bei einem Prozess eines Warnens eines Benutzers vor einer Kapazitätserschöpfung (Schritt **80**). Anfänglich wird durch den Benutzer ein gewünschtes Schwellenverhältnis an ungenutztem virtuellem Speicherraum zu der virtuellen Gesamtkapazität (TVC) eingestellt. Während des Betriebs wird ein tatsächliches Verhältnis des tatsächlichen Umfangs an ungenutztem virtuellem Speicherraum zu der virtuellen Gesamtkapazität berechnet und mit dem Schwellenverhältnis verglichen. Falls das tatsächliche Verhältnis das Schwellenverhältnis übersteigt, informiert das Datenspeichersystem den Benutzer, dass die Kapazität des Systems erschöpft wurde. Wenn das tatsächliche Verhältnis das Schwellenverhältnis nicht übersteigt, wird keine Warnung gegeben. Diese Option wird durch die folgende Gleichung [11] dargestellt:
Warnung „Kapazität erschöpft" =
falls ein Heiße-Reserve-Raum reserviert wird,
Verwendung von TVC = TVC ohne größte Speicherplatte
andernfalls,
Verwendung von TVC = TVC:
(Schwellenverhältnis – (ungenutzter virtueller Speicherraum/TVC)) ≥ 0.

Option 5: Auswertung, ob überlassene Kapazität von fehlender Speicherplatte abhängt

[0082] Bei Schritt **82** besteht eine andere Option, die auf eine Bestimmung der bei Gleichung [7] berechneten virtuellen Gesamtkapazität folgt, darin, zu bestimmen, ob die dem Benutzer überlassene Kapazität von einer fehlenden Speicherplatte abhängt. Eine derartige Situation entsteht, wenn ein Benutzer eine neue Speicherplatte hinzugefügt hat, das System jedoch noch keinen virtuellen Speicherraum auf RAID-Ebene, der den neuen Speicherplattenraum umfasst, zugewiesen hat. Die Speicherplatte könnte entfernt werden, ohne die überlassene Kapazität, die dem Benutzer bereits zugewiesen wurde, zu beeinflussen.

[0083] Anhand der folgenden Beziehung [12] wertet das System aus, ob eine Platte entfernt werden kann, ohne die vorhandene überlassene Kapazität zu beeinflussen:
Überlassene Kapazität hängt von Speicherplatte ab, wenn:
falls ein Heiße-Reserve-Raum reserviert wird,
TVC ohne Kapazität der fehlenden Speicherplatte und ohne größte Speicherplatte – überlassene Kapazität < 0
andernfalls
TVC ohne Kapazität der fehlenden Speicherplatte – überlassene Kapazität < 0.

Option 6: Rekonstruktion der Platte

[0084] Eine weitere Option besteht darin, zu bestimmen, ob in dem Fall, dass eine Speicherplatte ausfällt, eine ausreichende Kapazität vorliegt, um eine Speicherplatte zu rekonstruieren bzw. wiederaufzubauen (Schritt **84**). Bezüglich einer Rekonstruktion einer Platte liegt eine ausreichende Kapazität vor, eine Rekonstruktion einer Speicherplatte zu ermöglichen, falls die um die überlassene Kapazität verringerte virtuelle Gesamtkapazität größer als oder gleich null ist. An diesem Punkt ist die Speicherplatte ausgefallen, und somit ist die Kapazität der ausgefallenen Speicherplatte kein Bestandteil der physischen Kapazität oder der virtuellen Gesamtkapazität (TVC). Dies kann durch die folgende Beziehung [13] dargestellt werden:
Erlaube Rekonstruktion einer Platte, wenn:
TVC – überlassene Kapazität ≥ 0.

Option 7: Aktivierung einer heißen Reserve

[0085] Eine Anforderung, die Heiße-Reserve-Option zu aktivieren, kann auftreten, nachdem der Benutzer bereits Daten in dem System gespeichert hat. Es existiert ausreichend Raum, um die Heiße-Reserve-Option zu berücksichtigen, wenn die virtuelle Gesamtkapazität (die um die Kapazität der größten Speicherplatte verringert ist), minus der überlassenen Kapazität, größer oder gleich null ist. Diese Formel ist dieselbe wie beim Prüfen, ob die nicht-überlassene Kapazität, die bei Gleichung [8] unter der Annahme abgeleitet wurde, dass die Heiße-Reserve-Option aktiviert ist (d. h. um die virtuelle Gesamtkapazität TVC um die Größe der größten Speicherplatte zu verringern), größer als null ist.

[0086] Die Heiße-Reservebildung-Anforderung kann durch die folgende Beziehung [14] dargestellt werden:
Erlaube Anforderung bezüglich heißer Reserve, wenn:
TVC ohne Kapazität der größten Speicherplatte – überlassene Kapazität ≥ 0.

[0087] Die oben beschriebenen Optionen stellen mögliche Verwendungen des in den Gleichungen [1] – [7] abgeleiteten Werts der virtuellen Gesamtkapazität dar. Auch andere Verwendungen sind möglich. Indem es die virtuelle Gesamtkapazität ständig überwacht, vermeidet das Datenverwaltungssystem dieser Erfindung effektiv eine Überbeanspruchung von Ressourcen, da jede Anforderung auf der Basis der vorliegenden virtuellen Gesamtkapazität und bezüglich dessen, wie sie weiterhin nicht-überlassen bleibt, analysiert werden kann.

[0088] Fig. 9 zeigt ein Verfahren zum Betreiben eines redundanten hierarchischen Datenspeichersystems gemäß einem weiteren Aspekt dieser Erfindung. Dieses Verfahren untersucht, ob angesichts der vorliegenden Anforderungen an die Speicherkonfiguration und Arbeitslast mehr Spiegelzuweisungsblöcke zugewiesen werden können. Schritt **120** stellt den virtuellen Speicherraum auf RAID-Ebene fest, was oben unter Bezugnahme auf Fig. 4 ausführlich beschrieben wurde. Bei Schritt **122** reserviert das RAID-Verwaltungssystem einen Speicherraum auf nicht-zusammenhängende Weise in dem gesamten virtuellen Speicherraum auf RAID-Ebene. Das heißt, dass das RAID-Verwaltungssystem gewährleistet, dass ein Umfang an ungenutztem Raum, der über den gesamten virtuellen Speicherraum (sowohl in Spiegel- als auch Paritätsspeicherbereichen) verteilt ist, reserviert ist. Auf einen Empfang einer Anforderung, einen neuen Spiegelzuweisungs- oder Paritätszuweisungsblock zuzuweisen (Schritt **124**), untersucht das RAID-Verwaltungssystem die RAID-Bereiche, um etwaige ungenutzte Zuweisungsblöcke zu identifizieren (Schritt **126**). Ungenutzte Zuweisungsblöcke sind Speicherblöcke in dem virtuellen Speicherraum auf RAID-Ebene, die noch nicht zugewiesen wurden.

[0089] Bei Schritt **128** werden die ungenutzten Zuweisungsblöcke gezählt. Man erinnere sich aus den Gleichungen [2] und [3], dass im Vergleich zu Spiegel-RAID-Bereichen mehr Daten in Paritäts-RAID-Bereichen gespeichert werden können, wenn derselbe Umfang an virtuellem Speicherraum gegeben ist. Dementsprechend beginnt das RAID-Verwaltungssystem eine Schleifenbildung durch die RAID-Bereiche (Spiegel-, Paritäts- und nicht Benannte) und zählt die Zuweisungsblöcke gemäß ihrem derzeit zugewiesenen Spiegel- und Paritätszustand. Bei nicht benannten RAID-Bereichen geht man für die Zwecke der Berechnung davon aus, dass alle Zuweisungsblöcke in derartigen nicht benannten RAID-Bereichen ungenutzte Paritätszuweisungsblöcke sind.

[0090] Wenn der Zählwert ungenutzter Zuweisungsblöcke während des Zählvorgangs größer als der oder gleich dem Umfang an reserviertem Speicherraum wird (d. h. der „Ja"-Zweig von Schritt **130**), erlaubt das RAID-Verwaltungssystem unmittelbar die Anforderung, den neuen Zuweisungsblock zuzuweisen (Schritt **132**), und hört auf zu zählen. Wenn der Zählwert ungenutzter Zuweisungsblöcke im Anschluss an einen gesamten Durchlauf durch den virtuellen Speicherraum auf RAID-Ebene dagegen geringer ist als der Umfang an reserviertem Speicherraum (d. h. der „Nein"-Zweig von Schritt **130**), lehnt das RAID-Verwaltungssystem die Anforderung, den neuen Zuweisungsblock zuzuweisen, ab (Schritt **134**).

[0091] In manchen Umständen kann das RAID-Verwaltungssystem eine Anforderung empfangen, die zu einer Zuweisung eines ganzen neuen Spiegel-RAID-Bereichs einer ausgewählten Größe führt. Dies erfolgt, wenn keine ungenutzten Spiegelzuweisungsblöcke gefunden werden konnten. Auf eine derartige Anforderung hin leitet das RAID-Verwaltungssystem eine maximal zulässige Größe eines Spiegel-RAID-Bereichs als Funktion der Anzahl ungenutzter Zuweisungsblöcke (Schritt **128**), des reservierten Speicherraums (in Zuweisungsblöcken ausgedrückt) und der Anzahl von Zuweisungsblöcken, die pro Region eines physischen Speicherraums vorgesehen sind, ab. Ein beispielhafter Wert dieser letzteren Variable beträgt sechzehn, da in eine 1M-Region sechzehn 64K-Zuweisungsblöcke abgebildet werden können. Genauer gesagt ist die maximal zulässige Größe eines Spiegel-RAID-Bereichs (MRA_max) durch die folgende Gleichung [15] gegeben:

$$MRA_max = (2 \cdot diff/AB\ pro\ Region) + 2$$

wobei „diff" den Unterschied zwischen der Anzahl ungenutzter Zuweisungsblöcke und dem reservierten Speicherraum darstellt und „AB pro Region" die Anzahl von Zuweisungsblöcken pro Region eines physischen Speicherraums darstellt.

[0092] Die Anforderung wird erlaubt, wenn die Größe des neuen Spiegel-RAID-Bereichs geringer als die oder gleich der maximal zulässigen Größe ist.

[0093] Diese Erfindung weist viele Vorteile auf. Erstens liefert sie einen unkomplizierten Lösungsansatz zur Vermeidung einer Überbeanspruchung eines virtuellen Speicherraums bezüglich eines Benutzers. Dies erfolgt, indem die überlassene Kapazität unter die dem Benutzer zur Verfügung stehende virtuelle Gesamtkapazität um den folgenden kombinierten Umfang begrenzt wird: (1) um zumindest einen für die Migration reservierten RAID-Bereich, (2) um den Raum, der erforderlich ist, um zu gewährleisten, dass zumindest ein minimaler Umfang an Benutzerdaten in einer Spiegelspeicherung behalten werden kann, und (3) um die Kapazität der größten Speicherplatte für den Fall, dass die „Heiße-Reserve"-Option aktiviert wird. Zweitens können, nachdem die virtuelle Gesamtkapazität bekannt ist, viele andere Anforderungen bezüglich einer umfassenderen Speicherung überprüft werden, um zu bestimmen, ob Raum zugewiesen werden kann, um die Anforderung zu berücksichtigen. Diese vorab erfolgende Auswertung verhindert eine Überbeanspruchung. Drittens wertet diese Erfindung Anforderungen bezüglich mehr Spiegel-RAID-Bereichen aus und lehnt derartige Anforderungen ab, wenn das System nicht in der Lage ist, mehr Spiegel-RAID-Bereiche zu berücksichtigen.

[0094] 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 Vermeiden einer Überbeanspruchung von virtuellem Speicherraum in einem hierarchischen Datenspeichersystem; wobei das Datenspeichersystem (**10**) ein Plattenarray (**11**) von mehreren Speicherplatten (**12**) ausgewählter Kapazitäten aufweist, die einen physischen Speicherraum (**34**) einer physischen Gesamtkapazität definieren; wobei das Verfahren die folgenden Schritte umfasst:
Abbilden des physischen Speicherraumes (**34**) in einen virtuellen Speicherraum (**40**) auf RAID-Ebene, der den physischen Speicherraum als mehrere RAID-Bereiche präsentiert, wobei die RAID-Bereiche Spiegel-RAID-Bereiche, die Daten gemäß einer Spiegelredundanz speichern, und Paritäts-RAID-Bereiche, die Daten gemäß einer Paritätsredundanz speichern, umfassen, wobei die Spiegel-RAID-Bereiche im Vergleich zu einer Quantität an virtuellem Speicherraum auf RAID-Ebene, der durch Paritäts-RAID-Bereiche eingenommen wird, um dieselbe Datenmenge zu speichern, eine gleiche oder größere Quantität an virtuellem Speicherraum auf RAID-Ebene einnehmen, um eine Datenmenge zu speichern;
Abbilden des virtuellen Speicherraums (**40**) auf RAID-Ebene in einen Speicherraum (**50**) auf Anwendungsebene;

Vorsehen, dass ein minimaler Prozentsatz an Benutzerdaten in den Spiegel-RAID-Bereich behalten wird;
Vorsehen, dass ein verbleibender Prozentsatz von Benutzerdaten in Paritäts-RAID-Bereichen behalten wird; und
Berechnen einer virtuellen Gesamtkapazität des virtuellen Speicherraums (**50**) auf Anwendungsebene, die einem Benutzer überlassen werden kann, auf der Basis des Vorsehens, dass der minimale Prozentsatz von Benutzerdaten in Spiegel-RAID-Bereichen behalten wird und dass der verbleibende Prozentsatz an Benutzerdaten in Paritäts-RAID-Bereichen behalten wird.

2. Ein Verfahren gemäß Anspruch 1, das ferner die folgenden zusätzlichen Schritte umfasst:
Bereitstellen einer Heisse-Reserve-Option, bei der eine Quantität des virtuellen Speicherraumes (**40**) auf RAID-Ebene für einen Plattenersatzraum reserviert ist;
Erfassen, ob der Benutzer die Heisse-Reserve-Option aktiviert hat; und
für den Fall, dass die Heisse-Reserve-Option aktiviert ist, Berechnen der virtuellen Gesamtkapazität des virtuellen Speicherraums (**50**) auf Anwendungsebene, ohne eine Kapazität einer größten Speicherplatte zu enthalten.

3. Ein Verfahren gemäß Anspruch 1, das ferner die folgenden zusätzlichen Schritte umfasst:
Liefern einer überlassenen Kapazität des virtuellen Speicherraums (**50**) auf Anwendungsebene, die einer Verwendung durch den Benutzer überlassen wurde; und
Berechnen einer nicht überlassenen Kapazität, die nicht einer Verwendung durch den Benutzer überlassen wurde, indem die überlassene Kapazität von der virtuellen Gesamtkapazität subtrahiert wird.

4. Ein Verfahren gemäß Anspruch 1, das ferner die folgenden zusätzlichen Schritte umfasst:
Liefern einer überlassenen Kapazität des virtuellen Speicherraums (**50**) auf Anwendungsebene, die einer Verwendung durch den Benutzer überlassen wurde; und
für den Fall, dass eine Speicherplatte ausfällt, umfasst der Rechenschritt ein Berechnen der virtuellen Gesamtkapazität, ohne eine Kapazität der ausgefallenen Speicherplatte zu enthalten; und
Bestimmen, ob eine ausreichende Kapazität vorliegt, um einen Neuaufbau der ausgefallenen Speicherplatte zu ermöglichen, indem ausgewertet wird, ob die virtuelle Gesamtkapazität, die berechnet wurde, ohne die Kapazität der ausgefallenen Speicherplatte zu enthalten, größer als die oder gleich der überlassenen Kapazität ist.

5. Ein Verfahren gemäß Anspruch 1, das ferner die folgenden zusätzlichen Schritte umfasst:
Liefern einer überlassenen Kapazität des virtuellen Speicherraums (**50**) auf Anwendungsebene, die einer Verwendung durch den Benutzer überlassen wurde; und
Empfangen einer Anforderung von dem Benutzer, eine neue Quantität an virtuellem Speicherraum (**54**) auf Anwendungsebene hinzuzufügen;
Auswerten, ob die neue Quantität an virtuellem Speicherraum (**50**) auf Anwendungsebene dem Benutzer überlassen werden kann, auf der Basis der virtuellen Gesamtkapazität, der überlassenen Kapazität und der angeforderten neuen Quantität an virtuellem Speicherraum auf Anwendungsebene; und
Bewilligen eines Dem-Benutzer-Überlassens der angeforderten neuen Quantität an virtuellem Speicherraum auf Anwendungsebene, wenn die angeforderte neue Quantität geringer ist als eine Differenz zwischen der virtuellen Gesamtkapazität und der überlassenen Kapazität.

6. Ein Verfahren gemäß Anspruch 1, das ferner die folgenden zusätzlichen Schritte umfasst:
Liefern einer überlassenen Kapazität des virtuellen Speicherraums (**50**) auf Anwendungsebene, die einer Verwendung durch den Benutzer überlassen wurde; und
Auswerten, ob die Speicherkapazität einer Speicherplatte verloren gehen kann, ohne sich auf die dem Benutzer überlassene Kapazität auszuwirken, wobei die Auswertung auf der virtuellen Gesamtkapazität, der überlassenen Kapazität und der Kapazität der einen Speicherplatte beruht.

7. Ein Verfahren gemäß Anspruch 1, das ferner die folgenden zusätzlichen Schritte umfasst:
Einstellen eines gewünschten Schwellenverhältnisses von ungenutztem virtuellem Speicherraum (**50**) auf Anwendungsebene zu der virtuellen Gesamtkapazität;
Bestimmen eines tatsächlichen Verhältnisses einer tatsächlichen Menge an ungenutztem virtuellem Speicherraum (**50**) auf Anwendungsebene zu der virtuellen Gesamtkapazität; und
Informieren des Benutzers, dass die Kapazität des virtuellen Speicherraums auf Anwendungsebene erschöpft ist, wenn das tatsächliche Verhältnis das Schwellenverhältnis übersteigt.

8. Ein Verfahren gemäß Anspruch 1, bei dem sich die einzelnen RAID-Bereiche über mehrere Speicher-

platten erstrecken und einer Region von einzelnen Speicherplatten zugeordnet sind, wobei der virtuelle Speicherraum auf RAID-Ebene auch ungenutzte Zuweisungsblöcke aufweist; wobei das Verfahren ferner die folgenden Schritte umfasst:
Liefern eines reservierten Speicherraums auf nicht zusammenhängende Weise in dem gesamten virtuellen Speicherraum (**40**) auf RAID-Ebene;
Empfangen einer Anforderung, einen neuen Zuweisungsblock entweder als einen Spiegelzuweisungsblock oder einen Paritätszuweisungsblock zuzuweisen;
Untersuchen einzelner RAID-Bereiche, um etwaige ungenutzte Zuweisungsblöcke zu identifizieren;
Zählen einer Anzahl ungenutzter Zuweisungsblöcke; und
Ermöglichen, dass die Anforderung einen neuen Zuweisungsblock zuweist, wenn die Anzahl ungenutzter Zuweisungsblöcke größer als der oder gleich dem reservierten Speicherraum ist.

9. Ein Verfahren zum Vermeiden einer Überbeanspruchung von virtuellem Speicherraum in einem hierarchischen Datenspeichersystem; wobei das Datenspeichersystem (**10**) ein Plattenarray (**11**) von N Speicherplatten) ausgewählter Kapazitäten aufweist, die einen physischen Speicherraum einer physischen Gesamtkapazität definieren; wobei das Verfahren die folgenden Schritte umfasst:
(a) Abbilden des physischen Speicherraumes in einen virtuellen Speicherraum (**40**) auf RAID-Ebene, der den physischen Speicherraum als mehrere RAID-Bereiche präsentiert, wobei sich einzelne RAID-Bereiche über mehrere Speicherplatten erstrecken, wobei die RAID-Bereiche eine Anzahl von Spiegel-RAID-Bereichen (MR), die Daten gemäß einer Spiegelredundanz speichern, und eine Anzahl von Paritäts-RAID-Bereichen (PR), die Daten gemäß einer Paritätsredundanz speichern, umfassen, wobei die Spiegel-RAID-Bereiche eine Quantität an virtuellem Speicherraum auf RAID-Ebene (MVT) einnehmen, die gleich ½MR · N ist und die Paritäts-RAID-Bereiche eine Quantität an virtuellem Speicherraum auf RAID-Ebene (PVT) einnehmen, die gleich ½ PR(N – 1) ist; und
Abbilden des virtuellen Speicherraums (**40**) auf RAID-Ebene in einen Speicherraum (**50**) auf Anwendungsebene;
(b) Vorsehen, dass ein minimaler Prozentsatz an Benutzerdaten in den Spiegel-RAID-Bereichen (MR) behalten wird;
(c) Definieren eines oder mehrerer Sätze von virtuellem Speicherraum auf RAID-Ebene, wobei einzelne Sätze eine Gruppe von RAID-Bereichen umfassen, die sich über dieselbe Anzahl von Speicherplatten erstrecken;
(d) Annehmen, dass sich die Spiegel-RAID-Bereiche (MR) in einem oder mehreren Sitzen befinden, die die Gruppe von RAID-Bereichen mit der höchsten Anzahl von Speicherplatten enthalten;
(e) Annehmen, dass sich die Paritäts-RAID-Bereiche (PR) in dem verbleibenden virtuellen Speicherraum auf RAID-Ebene befinden;
(f) Lokalisieren eines Satzes, der sowohl die Spiegel-RAID-Bereiche als auch die Paritäts-RAID-Bereiche enthält;
(g) Berechnen einer Kapazität des virtuellen Speicherraums auf RAID-Ebene des bei Schritt (f) lokalisierten Satzes als Funktion entweder der Quantität an virtuellem Speicherraum auf RAID-Ebene, der durch Spiegel-RAID-Bereiche (MVT) eingenommen wird, oder der Quantität an virtuellem Speicherraum auf RAID-Ebene, der durch die Paritätsredundanzbereiche (PVT) eingenommen wird; und
(h) Berechnen einer Kapazität des virtuellen Speicherraums auf RAID-Ebene eines beliebigen anderen Satzes als des bei Schritt (f) lokalisierten Satzes als Funktion entweder der Quantität an virtuellem Speicherraum auf RAID-Ebene, der durch Spiegel-RAID-Bereiche (MVT) eingenommen wird, oder der Quantität an virtuellem Speicherraum auf RAID-Ebene, der durch die Paritätsredundanzbereiche (PVT) eingenommen wird; und
(i) Summieren der Kapazitäten der virtuellen Speicherräume auf RAID-Ebene für jeden in den Schritten (g) und (h) berechneten Satz, um eine virtuelle Gesamtkapazität des virtuellen Speicherraums (**50**) auf Anwendungsebene, die einem Benutzer überlassen werden kann, zu ergeben.

10. Ein Verfahren gemäß Anspruch 9, bei dem sich die einzelnen RAID-Bereiche über mehrere Speicherplatten erstrecken und einer Region von einzelnen Speicherplatten zugeordnet sind, wobei der virtuelle Speicherraum auf RAID-Ebene auch ungenutzte Zuweisungsblöcke aufweist; wobei das Verfahren ferner die folgenden Schritte umfasst:
Liefern eines reservierten Speicherraums auf nicht zusammenhängende Weise in dem gesamten virtuellen Speicherraum (**40**) auf RAID-Ebene;
Empfangen einer Anforderung, einen neuen Zuweisungsblock entweder als einen Spiegelzuweisungsblock oder einen Paritätszuweisungsblock zuzuweisen;
Untersuchen einzelner RAID-Bereiche, um etwaige ungenutzte Zuweisungsblöcke zu identifizieren;
Zählen einer Anzahl ungenutzter Zuweisungsblöcke; und
Ermöglichen, dass die Anforderung einen neuen Zuweisungsblock zuweist, wenn die Anzahl ungenutzter Zuweisungsblöcke größer als der oder gleich dem reservierten Speicherraum ist.

11. Ein hierarchisches Datenspeichersystem, das folgende Merkmale aufweist: ein Plattenarray (**11**), das eine Anzahl von Speicherplatten (**12**) aufweist, die einen physischen Speicherraum definieren; eine Plattenarraysteuerung (**14**), die mit dem Plattenarray (**11**) gekoppelt ist, um einen Datentransfer zu und von den Platten (**12**) zu koordinieren; ein RAID-Verwaltungssystem (**16**), das wirksam mit der Plattenarraysteuerung gekoppelt ist, um einen virtuellen Speicherraum (**40**) auf RAID-Ebene auf den physischen Speicherraum (**34**) der Speicherplatten abzubilden; wobei der virtuelle Speicherraum auf RAID-Ebene den physischen Speicherraum als mehrere RAID-Bereiche präsentiert, wobei sich einzelne RAID-Bereiche über mehrere Speicherplatten erstrecken, wobei die RAID-Bereiche Spiegel-RAID-Bereiche, die Daten gemäß einer Spiegelredundanz speichern, und Paritäts-RAID-Bereiche, die Daten gemäß einer Paritätsredundanz speichern, umfassen; wobei das RAID-Verwaltungssystem (**16**) vorsieht, dass ein minimaler Prozentsatz von Benutzerdaten in den Spiegel-RAID-Bereichen behalten wird und dass ein verbleibender Prozentsatz an Benutzerdaten in Paritäts-RAID-Bereichen behalten wird;
wobei das RAID-Verwaltungssystem (**16**) den virtuellen Speicherraum (**40**) auf RAID-Ebene in einen Speicherraum (**50**) auf Anwendungsebene, der einem Benutzer präsentiert wird, abbildet; und
wobei das RAID-Verwaltungssystem auf der Basis des Vorsehens, dass der minimale Prozentsatz von Benutzerdaten in den Spiegel-RAID-Bereichen behalten wird und dass der verbleibende Prozentsatz an Benutzerdaten in den Paritäts-RAID-Bereichen behalten wird, eine virtuelle Gesamtkapazität des virtuellen Speicherraums auf Anwendungsebene berechnet, die einem Benutzer überlassen werden kann.

Es folgen 8 Blatt Zeichnungen

Anhängende Zeichnungen

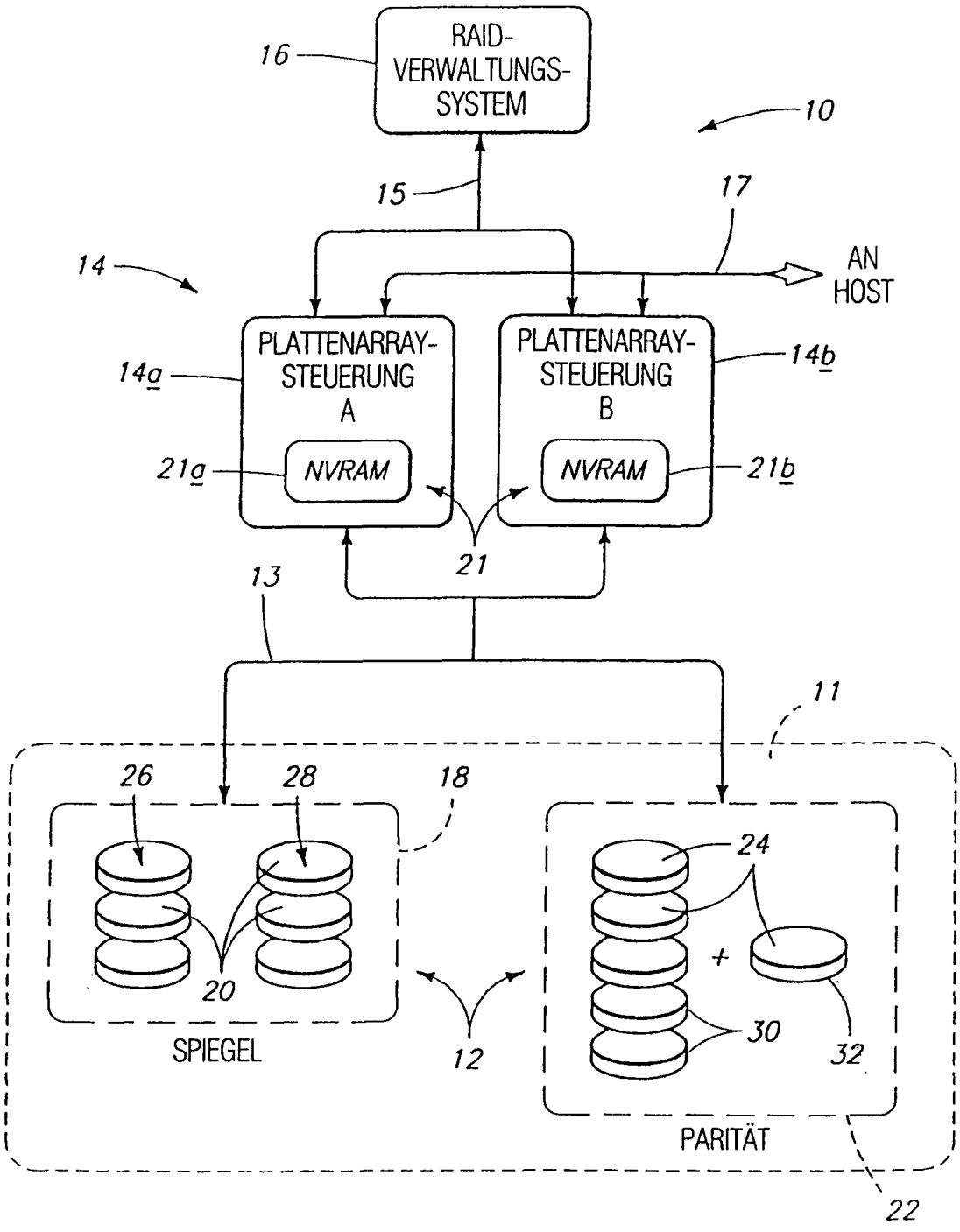


Fig. 1

PLATTEN

0	1	2	3	STREIFEN-NUMMER
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

0	1	2	3	STREIFEN-NUMMER
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

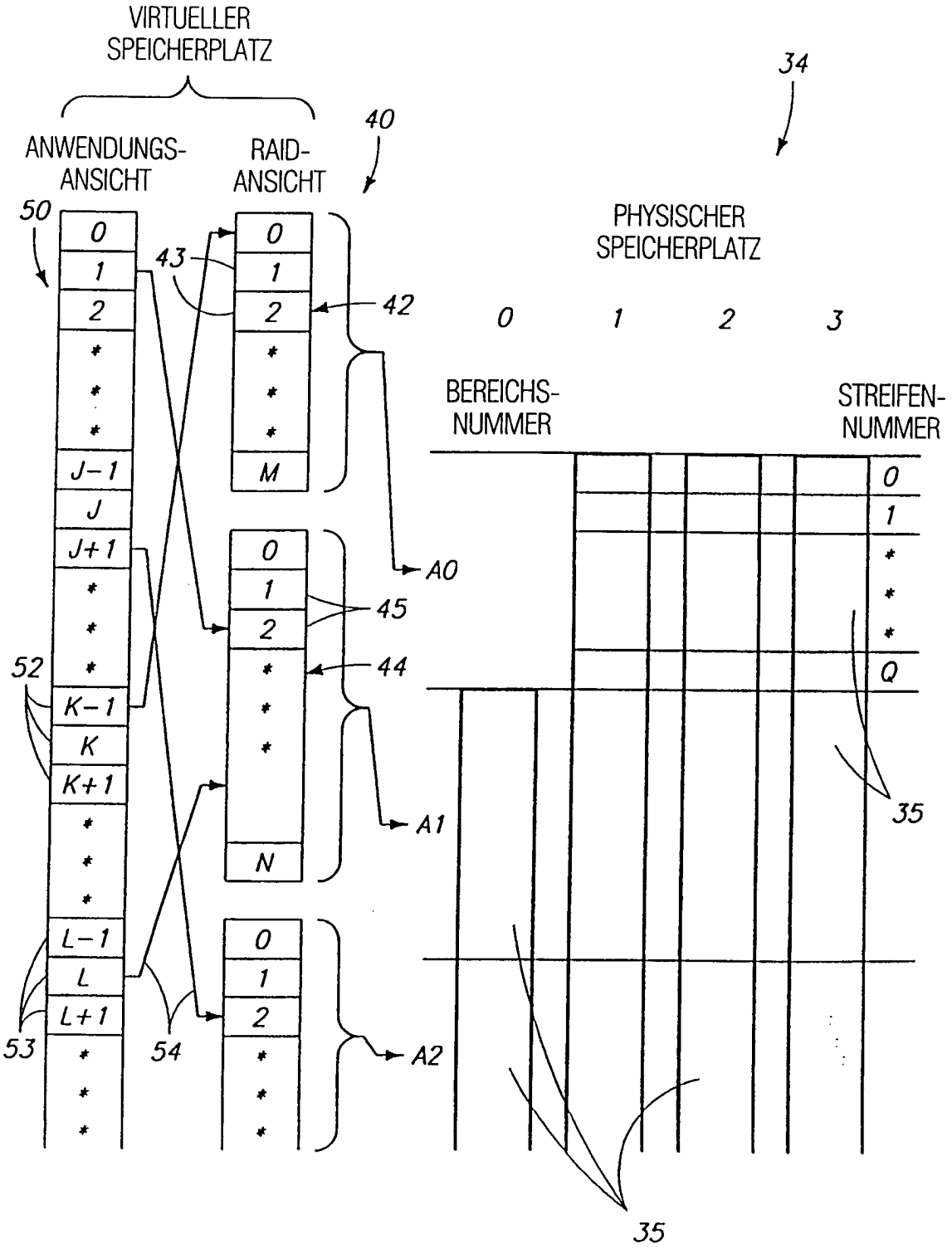
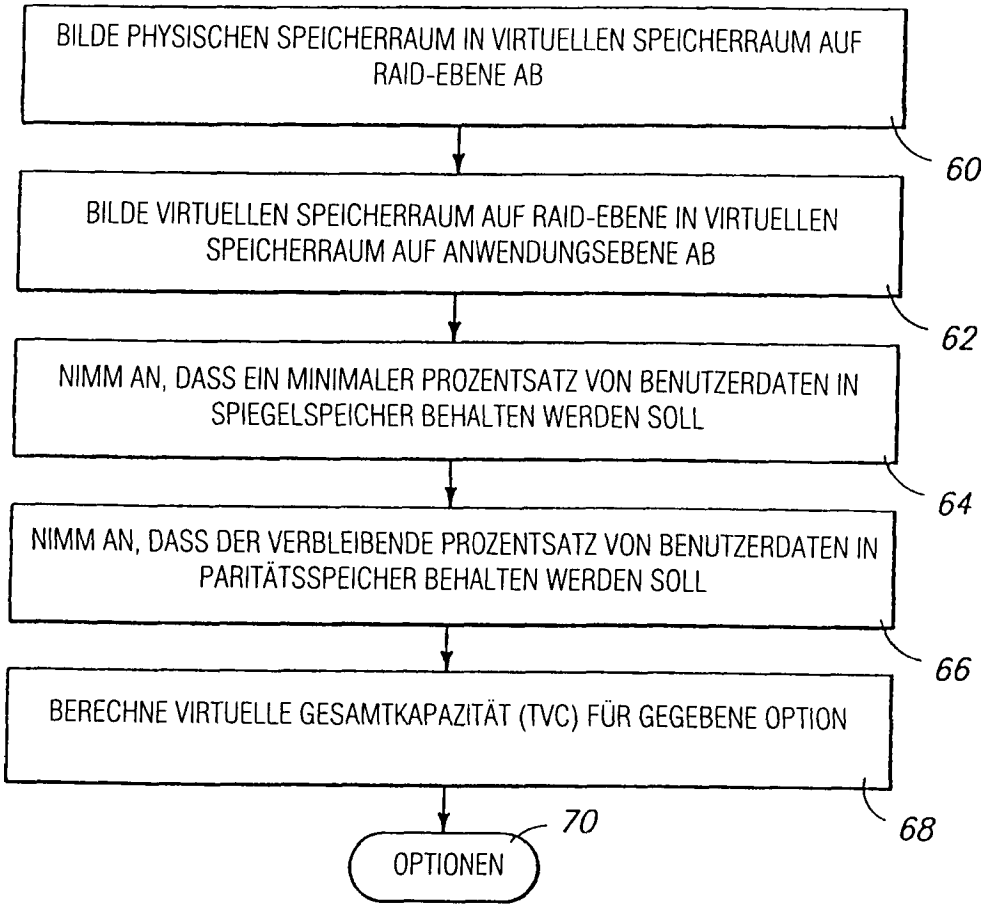
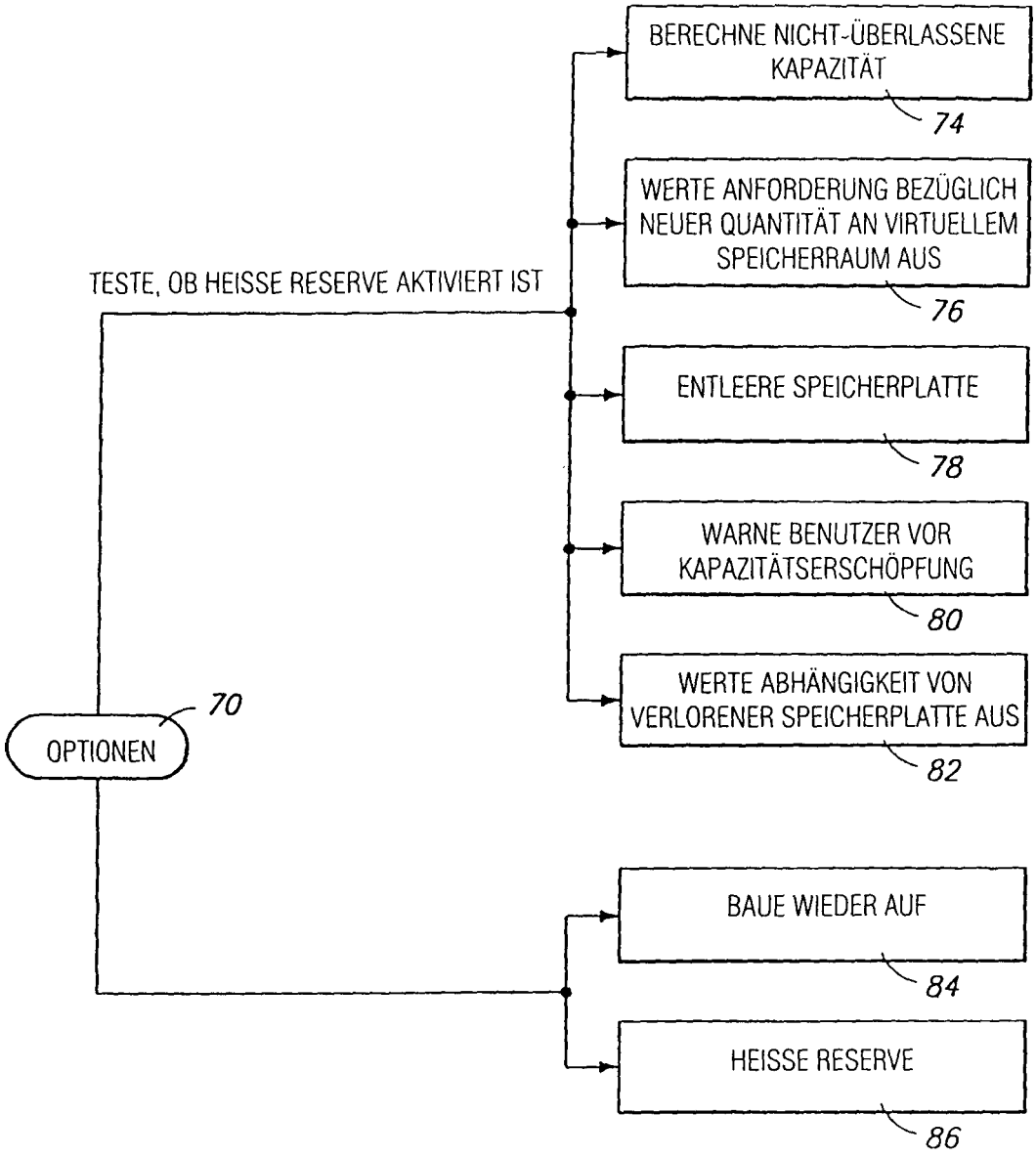

VIRTUELLER SPEICHERPLATZ
ANWENDUNGS-ANSICHT
RAID-ANSICHT
40
34
50
PHYSISCHER SPEICHERPLATZ
0
1
2
3
BEREICHS-NUMMER
STREIFEN-NUMMER
43
42
45
44
52
53
54
A0
A1
A2
35
35

Fig 4

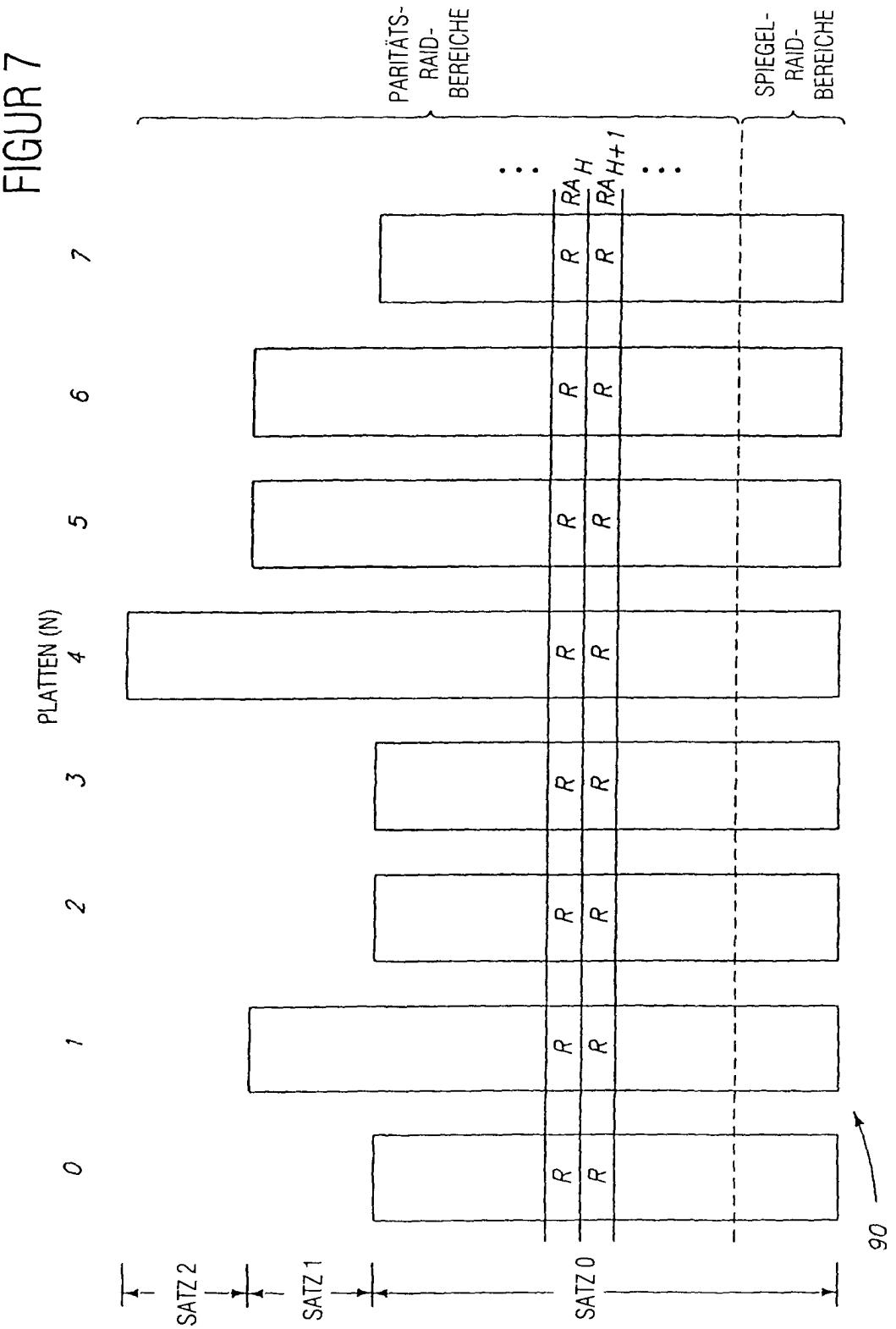

BILDE PHYSISCHEN SPEICHERRAUM IN VIRTUELLEN SPEICHERRAUM AUF RAID-EBENE AB
60
BILDE VIRTUELLEN SPEICHERRAUM AUF RAID-EBENE IN VIRTUELLEN SPEICHERRAUM AUF ANWENDUNGSEBENE AB
62
NIMM AN, DASS EIN MINIMALER PROZENTSATZ VON BENUTZERDATEN IN SPIEGELSPEICHER BEHALTEN WERDEN SOLL
64
NIMM AN, DASS DER VERBLEIBENDE PROZENTSATZ VON BENUTZERDATEN IN PARITÄTSSPEICHER BEHALTEN WERDEN SOLL
66
BERECHNE VIRTUELLE GESAMTKAPAZITÄT (TVC) FÜR GEGEBENE OPTION
68
70
OPTIONEN

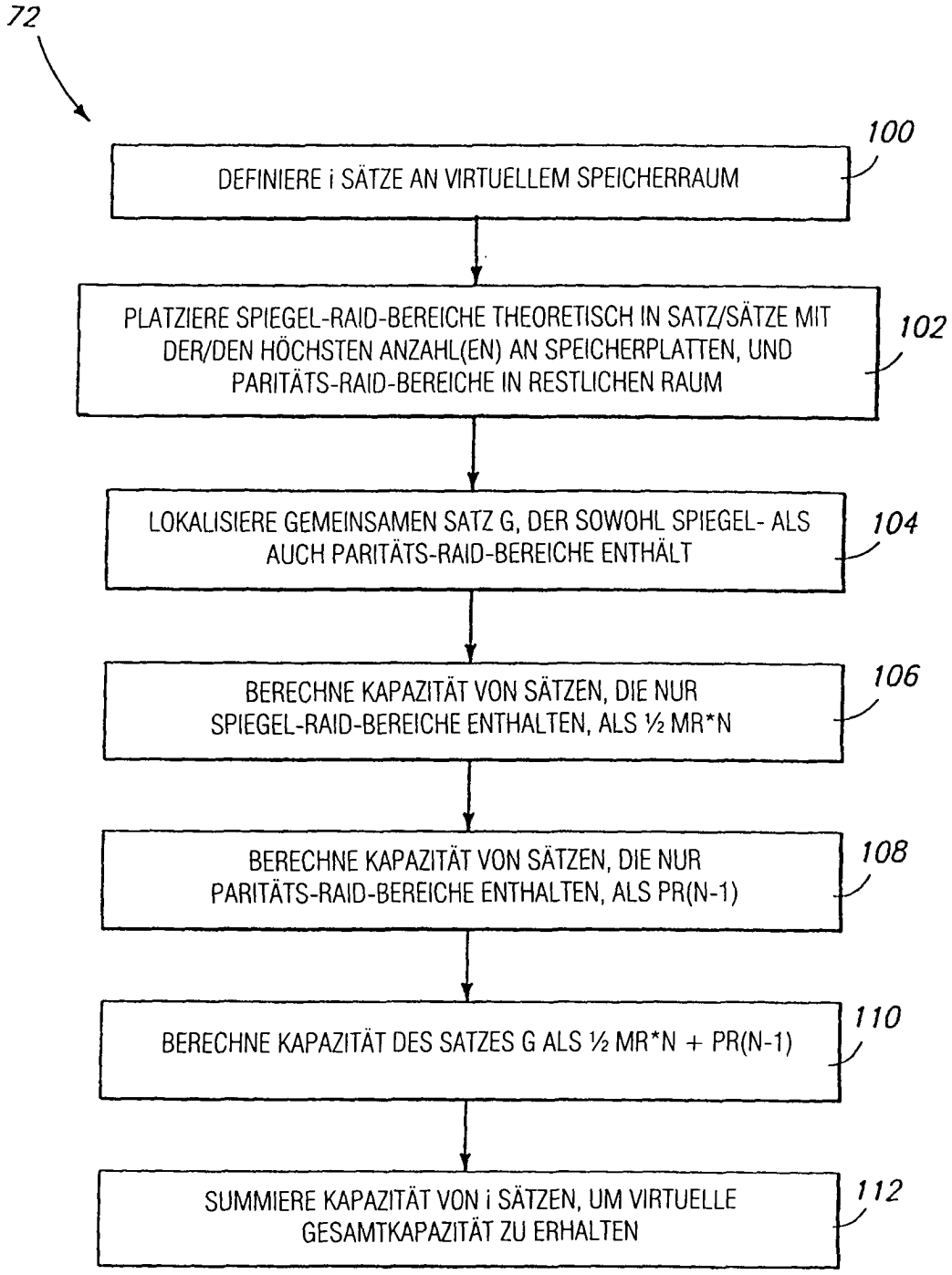
FIGUR 5


BERECHNE NICHT-ÜBERLASSENE KAPAZITÄT
74
WERTE ANFORDERUNG BEZÜGLICH NEUER QUANTITÄT AN VIRTUELLEM SPEICHERRAUM AUS
76
TESTE, OB HEISSE RESERVE AKTIVIERT IST
ENTLEERE SPEICHERPLATTE
78
WARNE BENUTZER VOR KAPAZITÄTSERSCHÖPFUNG
80
WERTE ABHÄNGIGKEIT VON VERLORENER SPEICHERPLATTE AUS
82
70
OPTIONEN
BAUE WIEDER AUF
84
HEISSE RESERVE
86

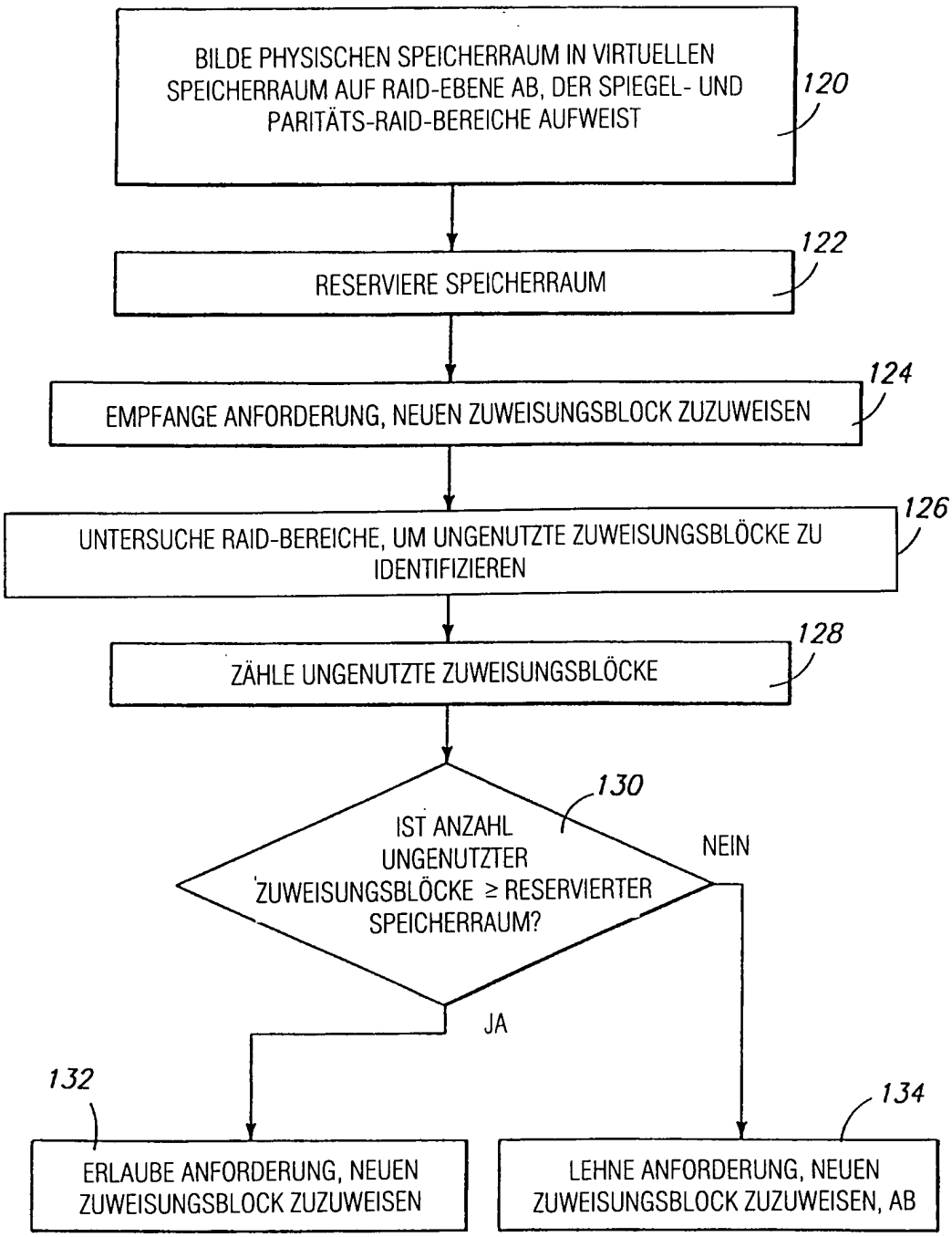
FIGUR 6

FIGUR 7


PARITÄTS-
RAID-
BEREICHE
SPIEGEL-
RAID-
BEREICHE
PLATTEN (N)
0
1
2
3
4
5
6
7
RA H
RA H+1
R
SATZ 2
SATZ 1
SATZ 0
90


72
100
DEFINIERE i SÄTZE AN VIRTUELLEM SPEICHERRAUM
PLATZIERE SPIEGEL-RAID-BEREICHE THEORETISCH IN SATZ/SÄTZE MIT DER/DEN HÖCHSTEN ANZAHL(EN) AN SPEICHERPLATTEN, UND PARITÄTS-RAID-BEREICHE IN RESTLICHEN RAUM
102
LOKALISIERE GEMEINSAMEN SATZ G, DER SOWOHL SPIEGEL- ALS AUCH PARITÄTS-RAID-BEREICHE ENTHÄLT
104
BERECHNE KAPAZITÄT VON SÄTZEN, DIE NUR SPIEGEL-RAID-BEREICHE ENTHALTEN, ALS ½ MR*N
106
BERECHNE KAPAZITÄT VON SÄTZEN, DIE NUR PARITÄTS-RAID-BEREICHE ENTHALTEN, ALS PR(N-1)
108
BERECHNE KAPAZITÄT DES SATZES G ALS ½ MR*N + PR(N-1)
110
SUMMIERE KAPAZITÄT VON i SÄTZEN, UM VIRTUELLE GESAMTKAPAZITÄT ZU ERHALTEN
112

FIGUR 8


BILDE PHYSISCHEN SPEICHERRAUM IN VIRTUELLEN SPEICHERRAUM AUF RAID-EBENE AB, DER SPIEGEL- UND PARITÄTS-RAID-BEREICHE AUFWEIST
120
RESERVIERE SPEICHERRAUM
122
EMPFANGE ANFORDERUNG, NEUEN ZUWEISUNGSBLOCK ZUZUWEISEN
124
UNTERSUCHE RAID-BEREICHE, UM UNGENUTZTE ZUWEISUNGSBLÖCKE ZU IDENTIFIZIEREN
126
ZÄHLE UNGENUTZTE ZUWEISUNGSBLÖCKE
128
IST ANZAHL UNGENUTZTER ZUWEISUNGSBLÖCKE ≥ RESERVIERTER SPEICHERRAUM?
130
NEIN
JA
132
ERLAUBE ANFORDERUNG, NEUEN ZUWEISUNGSBLOCK ZUZUWEISEN
134
LEHNE ANFORDERUNG, NEUEN ZUWEISUNGSBLOCK ZUZUWEISEN, AB

FIGUR 9