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

(10) **DE 695 32 030 T2** 2004.06.17

(12) # Übersetzung der europäischen Patentschrift

(97) **EP 0 726 521 B1**
(21) Deutsches Aktenzeichen: **695 32 030.0**
(96) Europäisches Aktenzeichen: **95 112 336.3**
(96) Europäischer Anmeldetag: **04.08.1995**
(97) Erstveröffentlichung durch das EPA: **14.08.1996**
(97) Veröffentlichungstag der Patenterteilung beim EPA: **29.10.2003**
(47) Veröffentlichungstag im Patentblatt: **17.06.2004**

(51) Int Cl.[7]: **G06F 3/06**
G06F 11/10, G06F 11/20, G11B 20/18

(30) Unionspriorität:
386574 **10.02.1995** **US**

(84) Benannte Vertragsstaaten:
DE, FR, GB

(73) Patentinhaber:
Hewlett-Packard Co. (n.d.Ges.d.Staates Delaware), Palo Alto, Calif., US

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

(72) Erfinder:
Nelson, Marvin D., Boise, Idaho 83704, US; Burkes, Theresa A., Meridian, Idaho 83642, US; Diamond, Bryan M., Boise, Idaho 83704, US; Jacobson, Michael B., Boise, Idaho 83714, US; Dolphin, Wade A., Boise, Idaho 83704, US; Voigt, Douglas L., Boise, Idaho 83702, US

(54) Bezeichnung: **Speicherplattenanordnung mit Ersatzbetriebsmitteln in Betrieb und Verfahren zur Verwendung von Ersatzbetriebsmitteln zum Speichern von Benutzerdaten**

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 Plattenarrays und insbesondere auf redundante hierarchische Plattenarraydatenspeichersysteme, die Heiße-Reserve-Ressourcen aufweisen, um einen ausreichenden Platz zur Rekonstruktion von Benutzerdaten und redundanten Informationen für den Fall, daß eine Speicherplatte ausfällt, zu liefern. Diese Erfindung bezieht sich ferner auf Verfahren zum Verwenden der Heiße-Reserve-Ressourcen, um Benutzerdaten zu speichern, bis derartige Ressourcen benötigt werden, um ein Wiederaufbauen zu erleichtern.

Hintergrund der Erfindung

[0002] 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.

[0003] 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.

[0004] 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.

[0005] Abgesehen von der Datenredundanz verbessern manche Plattenarraydatenspeichersysteme die Datenverfügbarkeit, indem sie eine zusätzliche physische Speicherplatte reservieren, die eine ausgefallene Speicherplatte ersetzen kann. Diese zusätzliche Speicherplatte wird als "Reserve" bezeichnet. Die Reserveplatte wird verwendet, um nach dem Plattenausfall Benutzerdaten zu rekonstruieren und eine Redundanz in dem Plattenarray wiederherzustellen, ein Vorgang, der als "Wiederaufbauen" bekannt ist. In manchen Fällen ist die zusätzliche Speicherplatte sogar an dem Plattenarray befestigt und in demselben vollständig betreibbar, bleibt jedoch außer Betrieb, bis eine Speicherplatte ausfällt. Diese unter Strom stehenden Speicherplatten werden als "heiße Reserven" bezeichnet.

[0006] Während die Heiße-Reserve-Platte im Fall eines Plattenausfalls nützlich ist, liefert sie dem Plattenarray während eines normalen Betriebs, wenn alle Platten ordnungsgemäß funktionieren, keine Dienste. Ein Reservieren und Vormerken einer gesamten physischen Speicherplatte für Heiße-Reserve-Zwecke erhöht somit die Kosten des Systems, ohne den Vorteil aufzuweisen, daß die zusätzlichen Ressourcen für Datenspeicherfunktionen vollständig genutzt werden.

[0007] Es wäre vorteilhaft, Heiße-Reserve-Ressourcen bereitzustellen und gleichzeitig alle Speicherplatten des Plattenarrays vollständig zu nutzen.

[0008] Aus der EP-A-0569313 ist ein Plattenspeichersystem einer RAID-Ebene 5 bekannt, bei dem eine der Mehrzahl von Speicherplatten als zweckgebundene Schreibunterstützungsspeichereinheit zur Verbesserung der Schreibleistungsfähigkeit verwendet wird, wodurch die Schreibunterstützungsspeichereinheit gleichzeitig auch als Heiße-Reserve-Einheit im Fall eines Plattenausfalls fungiert.

[0009] Aus der EP-A-0493984 ist ein RAID-Plattenspeichersystem bekannt, bei dem eine Rückkopier-Cache-Speichereinheit vorgesehen ist, an die Schreibvorgänge zu Spitzenlastzeiten des Systems geleitet werden. Daten, die in die Rückkopier-Cache-Speichereinheit geschrieben werden, werden während Ruheperioden des Systems zu den normalen RAID-Plätzen kopiert. Ferner wird gelehrt, daß eine zweckgebundene Heiße-Reserve-Speichereinheit als die Rückkopier-Cache-Speichereinheit fungieren kann.

Zusammenfassung der Erfindung

[0010] Diese Erfindung, die in den beigefügten unabhängigen Ansprüchen ausführlich definiert ist,

schafft ein Plattenarraydatenspeichersystem, das alle an dem Array befestigten Speicherplatten vollständig nutzt und trotzdem Heiße-Reserve-Ressourcen liefert. Statt eine ganze physische Platte als heiße Reserve zu reservieren und vorzumerken, reservieren das System und die Verfahren dieser Erfindung die logische Kapazität einer Heiße-Reserve-Platte, die aus einem verfügbaren Speicherraum, der über mehrere Speicherplatten in dem Array verteilt ist, besteht. Auf diese Weise werden alle Speicherplatten verwendet, um Benutzerdaten zu speichern, während gewährleistet wird, daß eine entsprechende Quantität an Platz zum Rekonstruieren von Benutzerdaten und Wiederherstellen einer Redundanz im Fall eines Plattenausfalls zur Verfügung gestellt werden kann:

[0011] Gemäß einem Aspekt dieser Erfindung weist ein Plattenarraydatenspeichersystem eine Mehrzahl von Speicherplatten und eine Plattenarraysteuerung zum Koordinieren eines Transfers von Benutzerdaten zu und von den Speicherplatten auf. Ein RRID-Verwaltungssystem ist vorgesehen, um einen virtuellen Speicherplatz der RAID-Ebene auf den physischen Speicherplatz der Speicherplatten abzubilden. Der virtuelle Speicherplatz der RAID-Ebene präsentiert den physischen Speicherplatz als mehrere RAID-Bereiche, die Spiegel-RAID-Bereiche umfassen, die Daten gemäß einer Spiegelredundanz speichern, und Paritäts-RAID-Bereiche, die Daten gemäß einer Paritätsredundanz speichern, umfassen. Das RAID-Verwaltungssystem bildet ferner den virtuellen Speicherplatz der RAID-Ebene in einen virtuellen Speicherplatz einer Anwendungsebene ab, der dem Benutzer präsentiert wird.

[0012] Wenn die Heiße-Reserve-Option aktiviert ist, verwendet das RAID-Verwaltungssystem den Heiße-Reserve-Platz, um Benutzerdaten in einem Spiegelspeicher zu speichern, während es ferner garantiert, daß der Heiße-Reserve-Platz im Anschluß an einen Speicherplattenausfall zum Wiederaufbauen einer Redundanz zur Verfügung gestellt werden kann. Platz wird zur Verfügung gestellt, indem Daten von einem Spiegelspeicher zu einem Paritätsspeicher bewegt werden. Diese Bewegung bzw. Migration erhöht gleichzeitig den Umfang an Paritätsspeicher, während sie den Umfang an Spiegelspeicher verringert. Gleichzeitig mit diesem Migrationsvorgang werden die Streifengrößen für eine Datenspeicherung aufgrund des von der ausgefallenen Speicherplatte verlorenen Platzes verringert.

[0013] Gemäß diesem System werden alle Speicherplatten in dem Array verwendet, um Benutzerdaten zu speichern, und zusätzlich, um einen Speicherplatz beizubehalten, der im Fall eines Plattenausfalls ohne weiteres zu Heiße-Reserve-Zwecken zur Verfügung gestellt werden kann. Diese Verwendung eines Heiße-Reserve-Speicherplatzes verbessert die Leistungsfähigkeit beträchtlich, indem sie die Parallelität erhöht und mehr Kapazität für einen Spiegelspeicher liefert.

Beschreibung der Zeichnungen

[0014] 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.

[0015] **Fig.** 1 ist ein schematisches Blockdiagramm eines redundanten hierarchischen Plattenarraydatenspeichersystems gemäß dieser Erfindung.

[0016] **Fig.** 2 ist eine schematische Veranschaulichung eines Speicherplatzes auf mehreren Speicherplatten und zeigt eine Datenspeicherung gemäß einer RAID-Ebene 1.

[0017] **Fig.** 3 ist eine schematische Veranschaulichung eines Speicherplatzes auf mehreren Speicherplatten und zeigt eine Datenspeicherung gemäß einer RAID-Ebene 5.

[0018] **Fig.** 4 ist eine schematische Veranschaulichung einer Speicherabbildungsanordnung dieser Erfindung, bei der zwei virtuelle Speicherplätze auf einen physischen Speicherplatz abgebildet werden.

[0019] **Fig.** 5 ist eine schematische Veranschaulichung eines Speicherplatzes auf mehreren Speicherplatten und zeigt eine Verwendung von Heiße-Reserve-Ressourcen vor einem Ausfall einer Speicherplatte.

[0020] **Fig.** 6 ist eine schematische Veranschaulichung eines Speicherplatzes auf mehreren Speicherplatten und zeigt eine Verwendung von Heiße-Reserve-Ressourcen nach einem Ausfall einer Speicherplatte.

[0021] **Fig.** 7 ist eine schematische Veranschaulichung eines Speicherplatzes auf mehreren Speicherplatten und zeigt, wie eine Datenmigration von einer Speicherung der RAID-Ebene 1 zu einer Speicherung der RAID-Ebene 5 einen Platz für Heiße-Reserve-Zwecke zur Verfügung stellt.

[0022] **Fig.** 8 ist ein Flußdiagramm eines Verfahrens zum Betreiben eines Plattenarrays, um die Verfügbarkeit eines Heiße-Reserve-Platzes zu gewährleisten.

Ausführliche Beschreibung der Erfindung

[0023] **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**. Diese Erfindung wird im Kontext ihrer bevorzugten Implementierung als redundantes hierarchisches Plattenarraysystem beschrieben. Man beachte jedoch, daß Aspekte und Konzepte dieser Erfindung auch bei anderen Arten von Plattenarraydatenspeichersystemen,

beispielsweise nicht-hierarchischen Plattenarrays, verwendet werden können.

[0024] 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.

[0025] 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.

[0026] 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.

[0027] Die Plattenarraysteuerung **14** ist vorzugsweise als duale Steuerung implementiert, die aus der Plattenarraysteuerung A, die durch Bezugszeichen **14a** bezeichnet ist, und aus der Plattenarraysteuerung B, die durch Bezugszeichen **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.

[0028] Das hierarchische Plattenarray **11** kann als unterschiedliche Speicherplätze, einschließlich seines physischen Speicherplatzes und eines oder mehrerer virtueller Speicherplätze, charakterisiert werden. Diese verschiedenen Speicheransichten sind durch Abbildungstechniken verbunden. Beispielsweise kann der physische Speicherplatz des Plattenarrays in einen virtuellen Speicherplatz abgebildet werden, der Speicherbereiche gemäß den verschiedenen Datenzuverlässigkeitsebenen darstellt. Manche Bereiche in dem virtuellen Speicherplatz können einer ersten Zuverlässigkeitsspeicherebene zugewiesen sein, beispielsweise Spiegel- oder RAID-Ebene 1, und andere Bereiche können einer zweiten Zuverlässigkeitsspeicherebene zugewiesen sein, beispielsweise Paritätsoder RAID-Ebene 5. Die verschiedenen Abbildungstechniken und virtuellen Plätze, die RAID-Ebenen betreffen, werden nachfolgend ausführlicher beschrieben.

[0029] Das Datenspeichersystem **10** umfaßt einen Speicherabbildungsspeicher **21**, der eine durchgängige Speicherung der virtuellen Abbildungsinformationen liefert, die verwendet werden, um verschiedene Speicherplätze 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.

[0030] 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).

[0031] 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.

[0032] 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.

[0033] 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.

[0034] **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.

[0035] 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.

[0036] 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 den fünf Platten **30** gespeichert, und redundante "Paritäts"-Daten sind auf der sechsten Platte **32** gespeichert.

[0037] **Fig.** 3 zeigt einen Paritäts-RAID-Bereich-Entwurf etwas ausführlicher. Ähnlich dem Entwurf des Spiegel-RAID-Bereichs der **Fig.** 2 kann der physische Speicherplatz 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 "$\oplus$" dargestellt wird. Die Paritäten für die ersten vier Streifen (wobei die tiefgestellte Zahl den entsprechenden Streifen darstellt) lauten wie folgt:

P_0 = Segment 0 $\oplus$ Segment 1 $\oplus$ Segment 2
= Platte 0 $\oplus$ Platte 1 $\oplus$ Platte 2
P_1 = Segment 3 $\oplus$ Segment 4 $\oplus$ Segment 5
= Platte 0 $\oplus$ Platte 1 $\oplus$ 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

[0038] 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.

[0039] 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 Speicherplatz wird ein Abschnitt dieses Speicherplatzes 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 Speicherplatzes RAID-Bereiche zu bilden. Zusätzlich wird eine weitere (durch einen Host definierte) Ansicht eines Speicherplatzes 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.

[0040] 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.

[0041] 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.

[0042] 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.

[0043] **Fig.** 4 veranschaulicht ein Speicherabbilden des verfügbaren Speicherplatzes des Datenspeichersystems **10** als mehrere Lagen eines abgebildeten virtuellen Speicherplatzes. Jedes vertikal verlängerte Rechteck in dem Diagramm stellt eine Ansicht des physischen Speicherplatzes dar. Bei diesem Diagramm wird auf den physischen Speicherplatz **34** durch zwei virtuelle Speicheransichten **40** und **50** Bezug genommen. Der physische Speicherplatz **34** wird durch vier Platten (z. B.

[0044] 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 Speicherplatzes dar, bei der die Platten 1, 2 und 3 eine ungefähr gleiche Speicherkapazität aufweisen und die Platte 0 eine etwas gerin-

gere Speicherkapazität aufweist. Beispielhafte Speicherkapazitäten für derartige Platten betragen 1 bis 3 Gigabytes. Der Speicherplatz **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.

[0045] Der Speicherplatz der Platten wird in eine erste virtuelle Zwischenansicht **40** der RAID-Ebene des physischen Speicherplatzes **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 Speicherplatz darstellt. Die physische Konfiguration und RAID-Ansichten des Speicherplatzes entziehen sich der Anwendungsansicht.

[0046] Der RAID-Bereich-Speicherplatz **40** ist die Speicheransicht, die den Spiegel- und Paritätsspeicherplatz identifiziert. Beispielsweise kann ein RAID-Bereich **42** einen Spiegel-RAID-Bereich von M Zuweisungsblöcken **43** darstellen, während der RAID-Bereich **44** einen Parität-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 Speicherplatz **34**. Beispielsweise können sechzehn 64K-Zuweisungsblöcke **43** oder **45** an der virtuellen RAID-Ansicht auf eine einzige 1M-Region **35** abgebildet werden.

[0047] Die Spiegel- und Parität-RAID-Bereiche können den gesamten Speicherplatz **34** des Plattenarrays einnehmen, müssen aber nicht. Dementsprechend kann während bestimmter Anwendungen ein unbenutzter und unbezeichneter Speicherplatz vorliegen, der keinem bestimmten RAID-Bereich entspricht. Ein solcher Speicherplatz kann jedoch in einen Spiegel- oder Parität-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.

[0048] Der in den RRID-Bereichen verfügbare Speicherplatz 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 Speicherplatz auf den Speicherplatten **12** angibt. Der virtuelle Speicherplatz **50** präsentiert eine Ansicht eines linearen Satzes von gleich großen virtuellen Speicherblökken **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 Blockspeicherplatz **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.

[0049] Das RAID-Verwaltungssystem **16** kann die Konfiguration der RAID-Bereiche über den physischen Speicherplatz 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 Plattenspeicherplatzes 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 Ansicht **50** der Anwendungsebene auf die RAID-Ansicht **40** allgemein in einem Veränderungszustand. Der Speicherabbildungsspeicher in den NV-RAMs **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.

[0050] Die Migrationsoperation des Speichersystems **10** wird nun unter Bezugnahme auf **Fig.** 1 und 4 beschrieben.

[0051] Für die Zwecke einer fortführenden Erläuterung nehmen virtuelle Blöcke **53** des virtuellen Speicherplatzes **50** der Anwendungsebene Bezug auf zugeordnete Zuweisungsblöcke **45** in dem Parität-RAID-Bereich **44**, die in dem Bereich A1 des physischen Speicherplatzes **34** gespeichert sind. Derartige virtuelle Blöcke **53** können als "virtuelle Paritätsblöcke" bezeichnet werden, während die zugeordneten Zuweisungsblökke **45** als "Paritätszuweisungsblöcke" bezeichnet werden.

[0052] Desgleichen nehmen virtuelle Blöcke **52** Bezug auf zugeordnete Zuweisungsblöcke **43** in dem Spiegel-RAID-Bereich **42**, die in dem Bereich A0 des physischen Speicherplatzes **34** gespeichert sind. Derartige virtuelle Blöcke **52** können hierin als "virtuelle Spiegelblöcke" bezeichnet werden, während die zugeordneten Zuweisungsblöcke **43** als "Spiegelzuweisungsblöcke" bezeichnet werden.

[0053] Um Daten von einem RAID-Bereich zu einem anderen zu bewe gen, 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 Speicherplatzes **50** der Anwendungsebene auf den virtuellen Speicherplatz **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.

[0054] 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.

[0055] 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 Speicherplatz bewegten sich Daten von dem Bereich A1 zum Bereich A0.

[0056] 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 Speicherplatz belegen als Paritätszuweisungsblöcke, führt diese letzte Technik zu einer Nettoerhöhung des Umfangs an unbenutztem Speicher der RAID-Ebene.

[0057] 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 Spei-

cherplatzes 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.

[0058] 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 Blockspeicherplatz 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.

[0059] 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.

[0060] Die obigen zwei aufgezählten Sequenzen liefern Beispiele dessen, wie das Speichersystem dieser Erfindung arbeiten kann, um Daten zwischen zwei verschiedenen Redundanzebenen zu bewegen.

[0061] Abgesehen von der Datenredundanz bietet das Plattenarraydatenspeichersystem dieser Erfindung "Heiße-Reserve"-Fähigkeiten, um eine Datenverfügbarkeit zu verbessern. Wie in dem Abschnitt "Hintergrund der Erfindung" erörtert wurde, reservieren bekannte Plattenarrays eine zusätzliche physische Speicherplatte bzw. eine "Reserve", die in dem Fall, daß eine Speicherplatte ausfällt, verwendet wird. Die zusätzliche Speicherplatte ist an dem Plattenarray befestigt und in demselben vollständig betreibbar, bleibt jedoch im Ruhezustand, bis der Speicherplattenausfall auftritt. Die unter Strom stehende Speicherplatte wird als "heiße Reserve" bezeichnet. Die Heiße-Reserve-Platte wird verwendet, um die Daten nach dem Plattenausfall zu rekonstruieren und die Redundanz in der Platte wiederherzustellen, ein Vorgang, der als "Wiederaufbauen" bezeichnet wird. Ungünstigerweise liefern die zweckgebundenen Heiße-Reserve-Platten während eines normalen Betriebs, wenn alle Platten ordnungsgemäß funktionieren, keinen brauchbaren Dienst, sondern sind statt dessen nur im Fall eines Plattenausfalls von Nutzen.

[0062] Gemäß dem Plattenarray dieser Erfindung wird eine Speicherplatzmenge, die ausreichend ist, um ein Wiederaufbauen von Benutzerdaten im Fall eines Plattenausfalls zu ermöglichen, in einer verteilten Weise über viele, vorzugsweise alle, Speicherplatten in dem Array beibehalten und verwendet, um eine RAID-1-Speicherung zu erhöhen. Keine einzelne Speicherplatte ist als die "Heiße-Reserve"-Platte bezeichnet. Alle Platten in dem Array speichern zugängliche Benutzerdaten derart, daß ein Speicherplatz zum Wiederaufbauen zur Verfügung gestellt werden kann. Die Verwendung eines Heiße-Reserve-Platzes verbessert die Leistungsfähigkeit beträchtlich, indem die Parallelität erhöht wird und indem mehr Kapazität für eine Spiegelspeicherung bereitgestellt wird. Im wesentlichen ist die zusätzliche Speicherplatte, die in der Vergangenheit ausschließlich für Heiße-Reserve-Ressourcen reserviert war, nun zum Zweck eines normalen Betriebs des Speicherns von Benutzerdaten in das Array integriert.

[0063] **Fig.** 5 zeigt ein beispielhaftes homogenes Plattenarray aus sechs Platten, das Speicherplatten 0 bis 5 aufweist, bevor ein Plattenausfall auftritt. Der gesamte Speicherplatz wird verwendet, um Benutzer- und redundante Daten gemäß Spiegel- oder Paritätstechniken zu speichern. Kein einziger Bereich und keine einzige Platte ist lediglich für Heiße-Reserve-Ressourcen reserviert. Hier sind 55% der Benutzerdaten in RAID-5-Paritätsbereichen gespeichert, und 45% sind in RAID-1-Spiegelbereichen gespeichert. Obwohl der gesamte Speicherplatz für eine normale Datenspeicherung verwendet wird, kann eine Quantität des Platzes im Fall eines Plattenausfalls zum Rekonstruieren von Benutzerdaten und zum Wiederherstellen einer Redundanz zur Verfügung gestellt werden. In dieser Figur wird der Heiße-Reserve-Platz, der zur Verfügung gestellt werden kann, derzeit als RAID-1-Speicherplatz verwendet, der über die Speicherplatten in dem Array verteilt ist.

[0064] **Fig.** 6 zeigt dasselbe Plattenarray wie **Fig.** 5, aber im Anschluß an einen Ausfall der Speicherplatte 5. Die auf der Platte 5 gespeicherten Benutzer- und/oder redundanten Daten sind verloren. Auf einen Ausfall hin beginnt das Plattenarray einen Wiederaufbauvorgang, um eine Redundanz in dem System wiederaufzubauen. Durch Bewegen von Daten von dem Spiegelspeicher zu dem Paritätsspeicher wird Platz zur Verfügung gestellt. Diese Migration erhöht gleichzeitig den Umfang an Paritätsspeicher, während sie den Umfang an Spiegelspeicher verringert. Gleichzeitig mit diesem Migrationsvorgang wird die Größe aller Streifen in dem System um ein Segment, das aufgrund der ausgefallenen Speicherplatte verloren ist, verringert. Man beachte, daß sich die Streifen nun lediglich über die Platten 0 bis 4 und nicht über die Platte 5 erstrecken. Infolge dieser Datenmigration und Verringerung der Streifengröße sind nun 90% der Benutzerdaten in RAID-5-Bereichen gespeichert, und 10% sind in ARID-1-Bereichen gespeichert. Im Anschluß an diesen Wiederaufbauvorgang wurde die Redundanz in dem Plattenarray wiederhergestellt.

[0065] **Fig.** 7 veranschaulicht, wie eine Datenmigration im Anschluß an einen Ausfall der Platte 5 zusätzlichen Platz freimacht. Diese Figur zeigt Abschnitte der Speicherplatten 0 bis 5, die zwei Sätze von drei Streifen enthalten, die allgemein mit den Zahlen **90** und **92** bezeichnet sind. Daten in den oberen drei Streifen **90** sind gemäß Spiegelspeichertechniken gespeichert, bei denen Benutzerdaten in den Segmenten 1 bis 9 gespeichert sind und redundante Daten in den Segmenten 1' bis 9' gespeichert sind. Man beachte, daß die redundanten Daten 3', 6' und 9' auf der Platte 5 aufgrund des Ausfalls dieser Platte verlorengingen.

[0066] Bei dem Beispiel der **Fig.** 7 sei angenommen, daß die Benutzerdaten in den Streifen **90** zu dem unteren Satz von Streifen **92** bewegt werden. Aufgrund der Differenz bezüglich der redundanten Speichertechniken können die Benutzerdaten 1 bis 9 hier in weniger Segmenten gespeichert werden. Trotz des Verlustes der Platte 5 kann die Datenmigration, die an drei zusätzlichen Segmenten (als "unbenutzt" bezeichnet) zur Verfügung gestellt ist, nun verwendet werden, um Benutzerdaten zu speichern.

[0067] Die Fähigkeit, Heiße-Reserve-Platz als RAID-1-Speicher zu verwenden, während ein Wiederaufbauen im Anschluß, an einen Plattenausfall garantiert wird, wird durch einen Speicherverwalter verwaltet. Bei der bevorzugten Implementierung ist der Speicherverwalter als das RAID-Verwaltungssystem **16** (**Fig.** 1) verkörpert, das die Benutzerdaten, die redundanten Daten und den unbenutzten Platz auf dem Plattenarray nachverfolgt.

[0068] Das RAID-Verwaltungssystem gewährleistet, daß ein unbenutzter Speicherplatz zu Heiße-Reserve-Zwecken zur Verfügung gestellt werden kann, indem dem Umfang an virtuellem Speicherplatz der Anwendungsebene (bei **Fig.** 4 Platz **50**), der einem Benutzer überlassen werden kann, eine künstliche Obergrenze auferlegt wird ("künstlich" insofern, als die Menge an virtuellem Speicherplatz der Anwendungsebene in bezug auf das, was zur Verfügung gestellt würde, wenn die Heiße-Reserve-Option nicht bestünde, verringert ist). Um diese Obergrenze zu liefern, berechnet das RAID-Verwaltungssystem routinemäßig eine Kapazität des virtuellen Speicherplatzes der Anwendungsebene, der einem Benutzer überlassen werden kann, während gewährleistet wird, daß das Plattenarray trotzdem noch seine erforderlichen Funktionen (z. B. Datenmigration) erfüllen kann. Wenn der Benutzer die Heiße-Reserve-Option wählt, berechnet das RAID-Verwaltungssystem die Anwendungsebenenkapazität so, als ob sie nicht die Kapazität der Speicherplatte der größten Größe in dem Array umfassen würde. Die Kapazität der größten Speicherplatte wird verwendet, um zu garantieren, daß der Wiederaufbauvorgang 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 jeglicher Speicherplatte verwendet werden.

[0069] Nachdem die künstliche Obergrenze berechnet wurde, setzt das RAID-Verwaltungssystem diese Grenze durch, indem es ein Überlassen von zusätzlichem virtuellem Speicherplatz der Anwendungsebene über die berechnete Kapazität hinaus verhindert. Diese kombinierte Kapazitätsanalyse und -durchsetzung verringert künstlich die Größe des virtuellen Speicherplatzes der Anwendungsebene, um dadurch zu gewährleisten, daß eine ausreichende Quantität des virtuellen Speicherplatzes der RAID-Ebene zum Wiederaufbauen im Falle des Ausfalls einer Speicherplatte zur Verfügung gestellt werden kann.

[0070] Falls der gegenwärtige Umfang an virtuellem Speicherplatz der Anwendungsebene nicht gefüllt werden kann, während gleichzeitig eine heiße Reserve garantiert wird, erzeugt das RAID-Verwaltungssystem eine Warnung, um den Benutzer zu informieren, daß kein Heiße-Reserve-Platz verfügbar ist. Diese Situation kann entstehen, wenn das Plattenarray bereits eine Speicherplatte verloren hat. Beispielsweise könnte das Plattenarray der **Fig.** 6, das bereits eine ausgefallene Platte 5 aufweist, keinen weiteren Plattenverlust hinnehmen und trotzdem die Redundanz der Benutzerdaten aufrechterhalten. Eine andere Situation könnte entstehen, wenn der Benutzer die Heiße-Reserve-Option anfänglich nicht freigibt, jedoch später dieses Merkmal anfordert, nachdem die Größe des virtuellen Speicherplatzes der Anwendungsebene bereits eingestellt wurde. Das Plattenarray ist an diesem Punkt eventuell nicht in der Lage, Heiße-Reserve-Ressourcen zu garantieren.

[0071] **Fig.** 8 zeigt ein bevorzugtes Verfahren zum Betreiben des Plattenarrays während normaler Bedingungen und im Falle eines Plattenausfalls. Bei Schritt **100** bildet das RAID-Verwaltungssystem **16** den durch das Plattenarray **12** bereitgestellten physischen Speicherplatz in einen ersten virtuellen Spei-

cherplatz **40** bzw. einen virtuellen Speicherplatz **40** der RAID-Ebene ab (siehe **Fig.** 4). Wie oben beschrieben wurde, präsentiert der virtuelle Speicherplatz der RAID-Ebene den physischen Speicherplatz als Spiegel- und Paritäts-RAID-Bereiche. Bei Schritt **102** gewährleistet das RAID-Verwaltungssystem, daß über das Plattenarray genügend Speicherplatz zur Verwendung im Fall eines Plattenausfalls zur Verfügung gestellt werden kann.

[0072] Anfänglich, wenn Daten zuerst in dem System plaziert werden, werden die Daten gänzlich in dem Spiegelspeicher gespeichert (Schritt **104**). Die Benutzerdaten und ihre redundanten Gegenstücke werden vorzugsweise gleichmäßig über die Platten in dem Array verteilt. Eine Spiegelspeicherung liefert eine bessere Leistungsfähigkeit und Zuverlässigkeit im Vergleich zu einer Paritätsspeicherung und ist somit der bevorzugte Speichertyp. Wenn die Menge an Benutzerdaten ungefähr die Hälfte der physischen Kapazität der Platten in dem System erreicht, ist das Plattenarray effektiv voll, da die Benutzerdaten ungefähr die Hälfte der Speicherkapazität einnehmen und die redundanten Daten ungefähr die andere Hälfte einnehmen. Wenn die Menge an Benutzerdaten die 50%-Schwelle überschreitet, beginnt das RAID-Verwaltungssystem, Daten unter Verwendung einer Datenmigration zwischen den Typen sowohl in Spiegel- als auch Paritäts-RRID-Bereichen zu speichern (Schritt **106**). Im Gegensatz zu einer Spiegelspeicherung können unter Verwendung einer Paritätsspeicherung mehr Benutzerdaten in demselben Platzumfang gespeichert werden (wie oben bei **Fig.** 7 veranschaulicht ist). Während sich das Datenspeichersystem weiterhin füllt, werden proportional mehr Daten zu dem Paritätsspeicher bewegt, um das zunehmende Datenvolumen zu berücksichtigen. Die Menge an Benutzerdaten in dem Spiegelspeicher nimmt ab.

[0073] Für den Fall, daß eine Speicherplatte zu einem beliebigen Zeitpunkt während des normalen Betriebs ausfällt (wie durch die gestrichelten Linien zum Schritt **108** veranschaulicht ist), initiiert das RAID-Verwaltungssystem den Wiederaufbauvorgang. Bei Schritt **109** wertet das RAID-Verwaltungssystem aus, ob ein entsprechender Umfang an Heiße-Reserve-Platz vorliegt, der im Fall eines Plattenausfalls zur Verfügung gestellt werden kann. Mit anderen Worten, kann das System ohne die ausgefallene Speicherplatte in der Lage sein, die gesamte überlassene Kapazität mit einer Redundanz zu speichern? Falls kein ausreichender Platz vorliegt, um im Fall eines Plattenausfalls die gesamte überlassene Kapazität mit einer Redundanz zu speichern (d. h. der "Nein"-Zweig aus Schritt **109**), wird das Plattenarrayspeichersystem in einem verschlechterten Zustand ohne Heiße-Reserve-Ressourcen oder die Fähigkeit, eine Redundanz wiederherzustellen, betrieben (Schritt **111**).

[0074] Wenn dagegen ein ausreichender Platz für eine heiße Reserve vorliegt (d. h. der "Ja"-Zweig aus Schritt **109**), bestimmt das RAID-Verwaltungssystem, ob ausreichend Platz für die Zwecke des Rekonstruierens von Benutzerdaten und des Wiederherstellens einer Redundanz zur Verfügung gestellt wurde (Schritt **110**). Falls dies der Fall ist (d. h. der "Ja"-Zweig aus Schritt **110**), verwendet das Plattenarray den verfügbaren freien Platz zum Rekonstruieren der Benutzerdaten von der ausgefallenen Speicherplatte und zum Wiederherstellen der Redundanz (Schritt **112**). Alternativ dazu, falls nicht genügend Platz zur Verfügung gestellt wurde (d. h. der "Nein"-Zweig von Schritt **110**), bewegt das RAID-Verwaltungssystem Benutzerdaten von dem Spiegelspeicher zu dem Paritätsspeicher, um einen verfügbaren Speicherplatz freizumachen, um den verlorenen Platz auf der ausgefallenen Speicherplatte zu kompensieren. Das RAID-Verwaltungssystem bewegt so lange Daten von dem Spiegel- zum Paritätsspeicher (wie durch die Schleife der Schritte **110** und **114** dargestellt ist), bis ein ausreichender Umfang an verfügbarem Speicherplatz freigemacht ist, um den Bedarf an dem ausgefallenen Plattenplatz zu eliminieren.

[0075] Das Plattenarraysystem dieser Erfindung weist insofern Vorteile gegenüber bekannten Plattenarrays auf, als alle Speicherplatten gleichermaßen verwendet werden, um Daten in einem Spiegelspeicher zu speichern, während ein Wiederaufbau im Anschluß an einen Plattenausfall garantiert wird. Es gibt keine zweckgebundene Speicherplatte, die ungenutzt bleibt, bis ein Ausfall auftritt. Die Verwendung eines Heiße-Reserve-Speichers als Spiegelspeicher verbessert die Leistungsfähigkeit des gesamten Plattenarrays beträchtlich, indem die Parallelität erhöht wird und die Komplexität des Verwaltens der Schreiboperationen in redundante Speicherplätze verringert wird.

[0076] 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. Deshalb ist die Erfindung in jeglicher ihrer Formen oder Modifikationen in dem ordnungsgemäßen Schutzumfang der beigefügten Patentansprüche beansprucht, die gemäß der Äquivalente-Doktrin entsprechend interpretiert werden.

Patentansprüche

1. Ein Verfahren zum Bereitstellen von Speicherplatz in einem hierarchischen Speicherplattenarray zur Verwendung beim Wiederaufbauen in dem Fall, daß eine Speicherplatte ausfällt, wobei das Plattenarray (**11**) eine Mehrzahl von Speicherplatten (**12**) aufweist, die einen physischen Speicherplatz definieren, wobei der physische Speicherplatz in einen virtuellen Speicherplatz einer RAID-Ebene abgebildet

ist, der den physischen Speicherplatz als mehrere RAID-Bereiche präsentiert, die Spiegel- und Paritäts-RAID-Bereiche umfassen, wobei das Verfahren folgende Schritte aufweist:
vor einem Plattenausfall, Speichern von Benutzerdaten sowohl in den Spiegel- als auch den Paritäts-RAID-Bereichen, derart, daß die Benutzerdaten über alle Speicherplatten (**12**) verteilt sind;
Speichern von redundanten Daten auf zumindest einer der Speicherplatten (**12**);
Gewährleisten, daß eine ausreichende Quantität an Speicherplatz für ein Rekonstruieren von Benutzerdaten und ein Wiederherstellen einer Redundanz in dem Fall, daß eine der Speicherplatten ausfällt, zur Verfügung gestellt werden kann, so daß die Benutzerdaten vor und nach dem Speicherplattenausfall sowohl in den Spiegelals auch den Paritäts-RAID-Bereichen auf allen Speicherplatten (**12**) gespeichert sind; und
Verwenden der Quantität an Speicherplatz, um sowohl Benutzer- als auch Redundanzdaten zu speichern, bis der Speicherplattenausfall auftritt.

2. Ein Verfahren gemäß Anspruch 1, das ferner den Schritt des Umwandelns des zum Speichern von Benutzer- und redundanten Daten verwendeten Speicherplatzes in einen Heiße-Reserve-Platz zum Rekonstruieren der Benutzerdaten und zum Wiederherstellen der Redundanz aufweist.

3. Ein Verfahren gemäß Anspruch 1, das ferner den Schritt des Erzeugens einer Warnung aufweist, um einen Benutzer zu informieren, daß Heiße-Reserve-Ressourcen nicht garantiert werden können, wenn bestimmt wird, daß im Anschluß an einen Plattenausfall keine ausreichende Quantität an Speicherplatz zum Rekonstruieren von Benutzerdaten und zum Wiederherstellen der Redundanz in dem Plattenarray zur Verfügung gestellt werden kann.

4. Ein Verfahren zum Bereitstellen eines Speicherplatzes zur Verwendung beim Wiederaufbauen in einem hierarchischeu Datenspeichersystem, wobei das Datenspeichersystem (**10**) ein Plattenarray (**11**) von mehr als zwei Speicherplatten (**12**) ausgewählter Kapazitäten aufweist, die einen physischen Speicherplatz (**34**) definieren, wobei das Verfahren folgende Schritte aufweist:
Abbilden des physischen Speicherplatzes (**34**) in einen virtuellen Speicherplatz (**40**) einer RAID-Ebene, der den physischen Speicherplatz als mehrere RAID-Bereiche präsentiert, wobei die RAID-Bereiche Spiegel- RAID-Bereiche, die Daten gemäß einer Spiegelredundanz speichern, und Paritäts-RAID-Bereiche umfassen, die Daten gemäß einer Paritätsredundanz speichern, wobei sich die Spiegel- und Paritäts-RAID-Bereiche über alle Speicherplatten in dem Plattenarray erstrecken;
Speichern von Benutzerdaten sowohl in den Spiegelals auch den Paritäts-RAID-Bereichen, derart, daß die Benutzerdaten über alle Speicherplatten (**12**) in dem Plattenarray (**11**) verteilt sind;
Gewährleisten, daß eine ausreichende Quantität des virtuellen Speicherplatzes (**40**) der RAID-Ebene zum Rekonstruieren von Benutzerdaten und Wiederherstellen einer Redundanz im Fall eines Speicherplattenausfalls zur Verfügung gestellt werden kann;
Verwenden der Quantität des virtuellen Speicherplatzes der RAID-Ebene, um Daten zu speichern, bis der Speicherplattenausfall auftritt; und
im Fall eines Speicherplattenausfalls, Durchführen eines der folgenden zwei Schritte:
(a) Verwenden eines unbenutzten Speicherplatzes, der nicht als Spiegel- oder Paritäts-RAID-Bereiche zugewiesen ist, als den Speicherplatz, um Benutzerdaten zu rekonstruieren und die Redundanz von der ausgefallenen Speicherplatte wiederherzustellen, oder
(b) Bewegen von Benutzerdaten von Spiegel-RAID-Bereichen zu Päritäts-RAID-Bereichen und Verdichten der Spiegel-RAID-Bereiche, um Speicherplatz freizumachen, um den von der ausgefallenen Speicherplatte verlorenen Platz zu kompensieren;
wobei jede Speicherplatte vor und nach einem Speicherplattenausfall Daten gemäß sowohl einer Spiegel- als auch einer Paritätsredundanz speichert.

5. Ein Verfahren gemäß Anspruch 4, das ferner folgende Schritte aufweist:
Abbilden des virtuellen Speicherplatzes (**40**) der RAID-Ebene in einen virtuellen Speicherplatz (**50**) einer Anwendungsebene;
dynamisches Bewegen der Benutzerdaten zwischen den Spiegel- und den Paritäts-RAID-Bereichen; und
Berechnen einer Kapazität des virtuellen Speicherplatzes (**50**) der Anwendungsebene, die einem Benutzer überlassen werden kann, ohne eine Kapazität einer Speicherplatte einer größten Größe zu umfassen.

6. Ein Verfahren gemäß Anspruch 5, das ferner den zusätzlichen Schritt des Verhinderns eines Überlassens von zusätzlichem virtuellem Speicherplatz (**50**) der Anwendungsebene über die berechnete Kapazität hinaus aufweist, um zu gewährleisten, daß die ausreichende Quantität des virtuellen Speicherplatzes (**40**) der RRID-Ebene zum Wiederaufbauen zur Verfügung gestellt werden kann.

7. Ein Datenspeichersystem, das folgende Merkmale aufweist:
ein hierarchisches Plattenarray (**11**), das eine Mehrzahl von Speicherplatten (**12**) aufweist, die einen physischen Speicherplatz definieren;
eine Plattenarraysteuerung (**14**), die mit dem Plattenarray (**11**) gekoppelt ist, zum Koordinieren eines Transfers von Benutzerdaten zu und von den Speicherplatten (**12**);
einen Speicherverwalter, der wirksam mit der Plat-

tenspeichersteuerung gekoppelt ist, um eine Abbildung des physischen Speicherplatzes in einen virtuellen Speicherplatz der RAID-Ebene, die den physischen Speicherplatz als mehrere RAID-Bereiche präsentiert, die Spiegel-RAID-Bereiche, die Daten gemäß einer Spiegelredundanz speichern, und Paritäts-RAID-Bereiche, die Daten gemäß einer Paritätsredundanz speichern, einschließen, zu verwalten, wobei der Speicherverwalter Benutzerdaten sowohl in dem Spiegel- als auch dem Paritäts-RAID-Bereich speichert, derart, daß die Benutzerdaten vor einem etwaigen Plattenausfall über alle Speicherplatten in dem Plattenarray verteilt sind; und
wobei der Speicherverwalter (**16**) eine ausreichende Quantität an Speicherplatz beibehält, der zum Rekonstruieren der Benutzerdaten und Wiederherstellen einer Redundanz zur Verfügung gestellt werden kann, falls eine der Speicherplatten ausfällt, wobei der Speicherverwalter (**16**) die Quantität an Speicherplatz verwendet, um sowohl Benutzer- als auch redundante Daten zu speichern, während er garantiert, daß der Speicherplatz im Anschluß an einen Speicherplattenausfall zum Wiederaufbauen zur Verfügung gestellt werden kann, so daß jede Speicherplatte die Benutzerdaten vor und nach einem Speicherplattenausfall gemäß sowohl einer Spiegel- als auch einer Paritätsredundanz speichert.

8. Ein Datenspeichersystem gemäß Anspruch 7, bei dem:
das RAID-Verwaltungssystem den virtuellen Speicherplatz (**40**) der RAID-Ebene ferner in einen virtuellen Speicherplatz (**50**) der Anwendungsebene abbildet; und
das RAID-Verwaltungssystem eine Kapazität des virtuellen Speicherplatzes (**50**) der Anwendungsebene berechnet, die einem Benutzer überlassen werden kann, ohne eine Kapazität einer Speicherplatte einer größten Größe zu umfassen.

9. Ein Datenspeichersystem gemäß Anspruch 8, bei dem das RAID-Verwaltungssystem ein Überlassen von zusätzlichem virtuellem Speicherplatz (**50**) der Anwendungsebene über die berechnete Kapazität hinaus verhindert, um zu gewährleisten, daß die ausreichende Quantität des virtuellen Speicherplatzes (**40**) der RAID-Ebene zum Rekonstruieren der Benutzerdaten und Wiederherstellen der Redundanz zur Verfügung gestellt werden kann.

Es folgen 6 Blatt Zeichnungen

Anhängende Zeichnungen

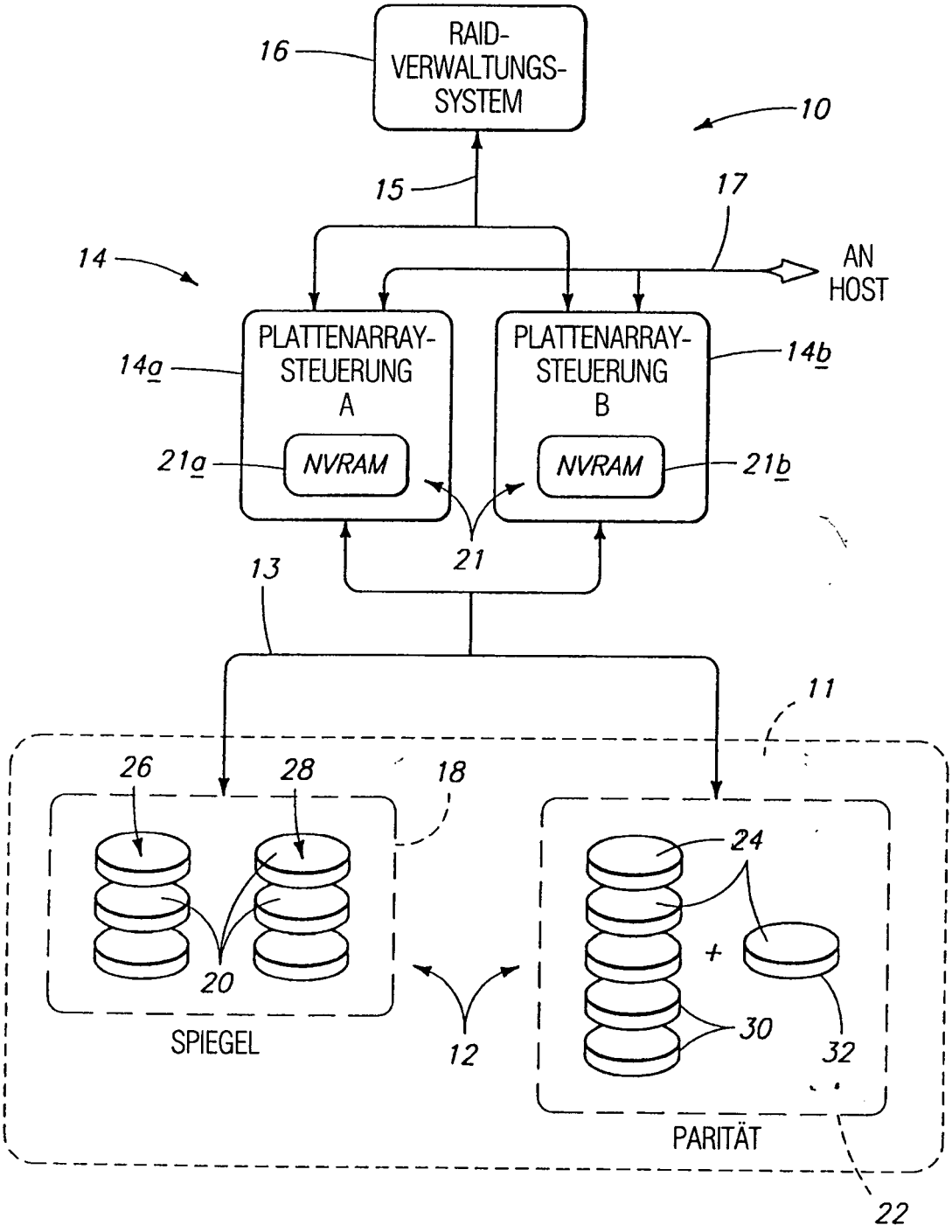

Fig 1

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

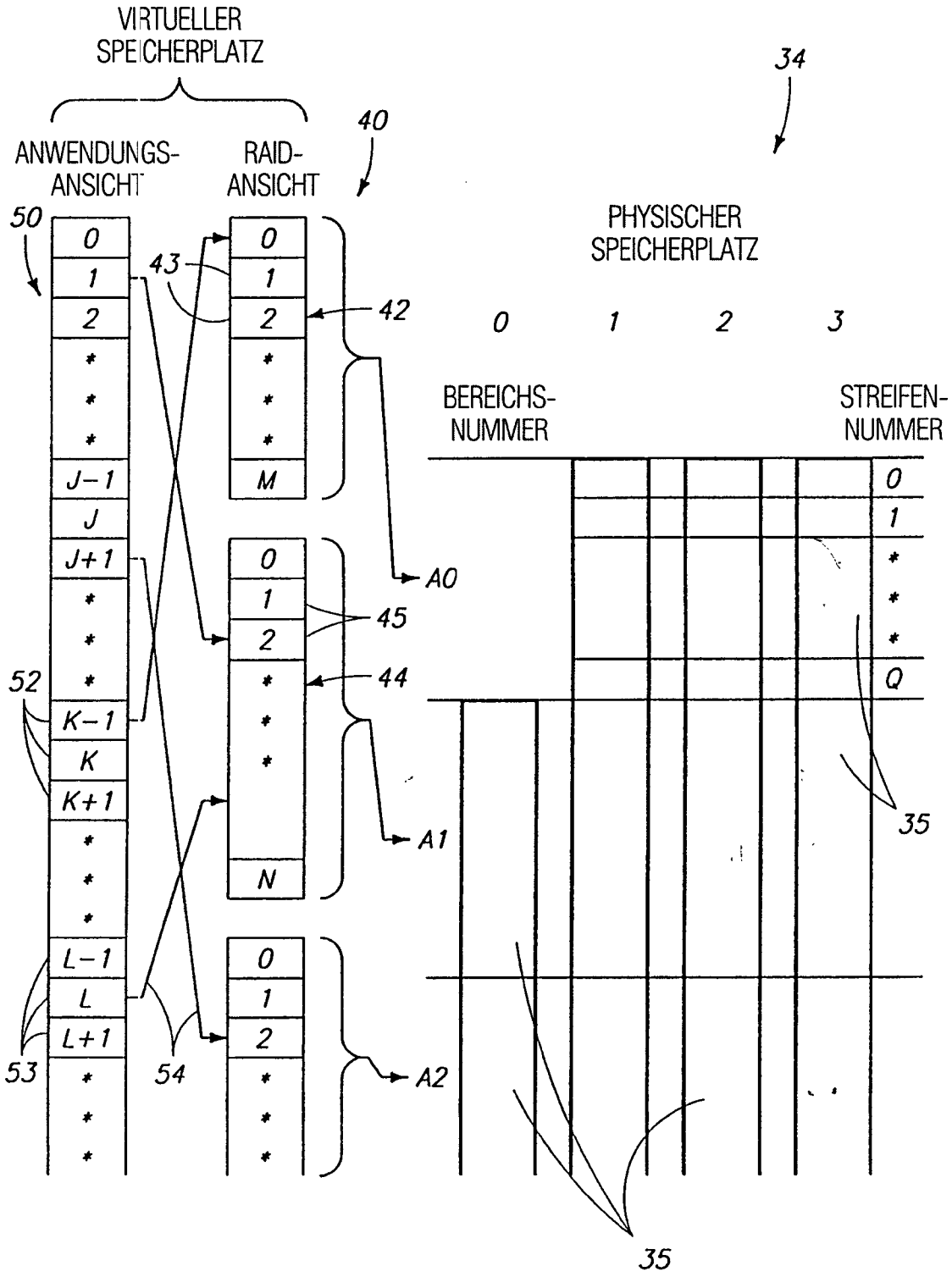
VIRTUELLER SPEICHERPLATZ
ANWENDUNGS-ANSICHT
RAID-ANSICHT
40
34
50
43
42
PHYSISCHER SPEICHERPLATZ
0
1
2
3
BEREICHS-NUMMER
STREIFEN-NUMMER
A0
45
44
A1
52
53
54
A2
35
35

Fig 4

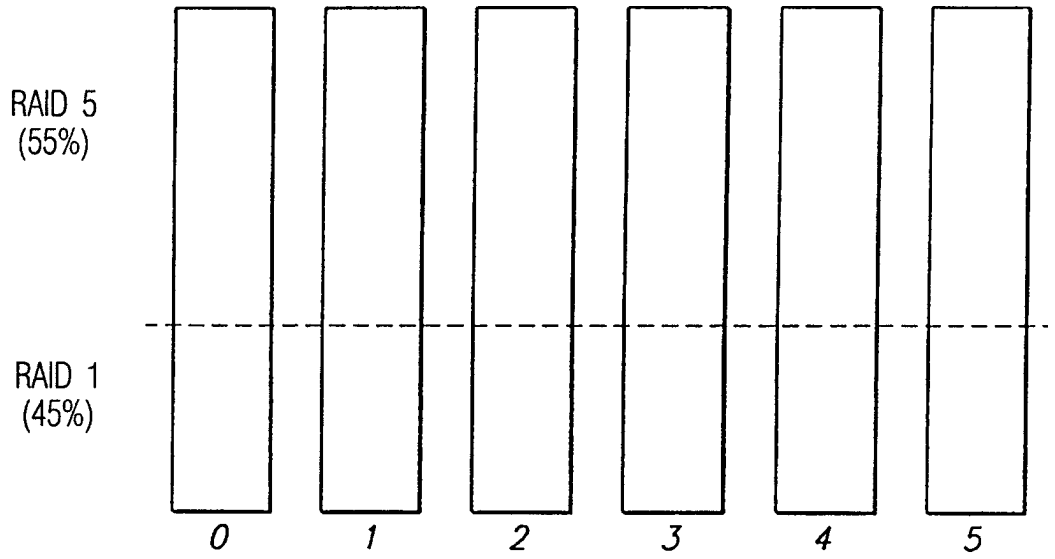
RAID 5
(55%)
RAID 1
(45%)
0
1
2
3
4
5
PLATTEN

Fig 5

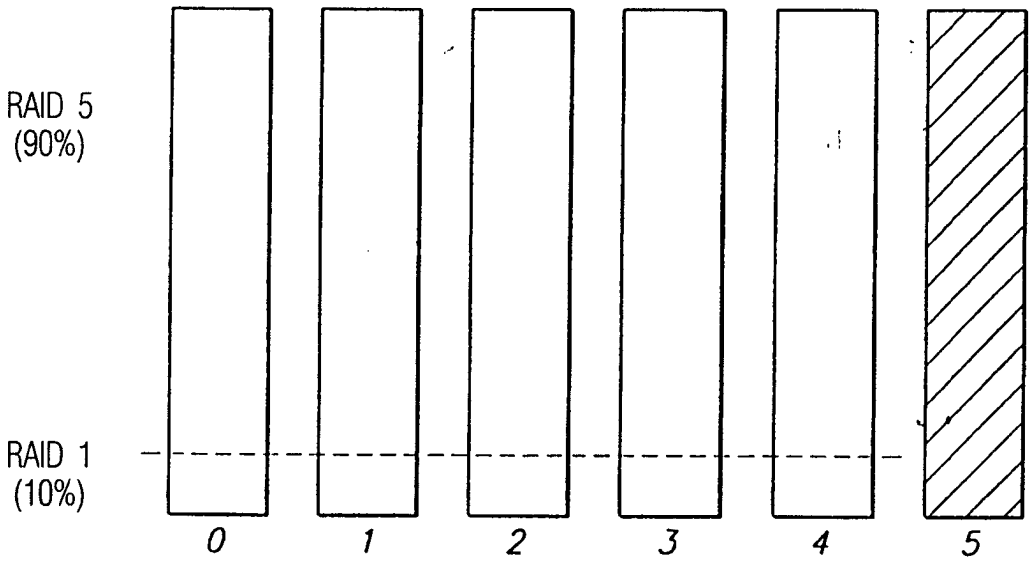
RAID 5
(90%)
RAID 1
(10%)
0
1
2
3
4
5
PLATTEN

Fig 6

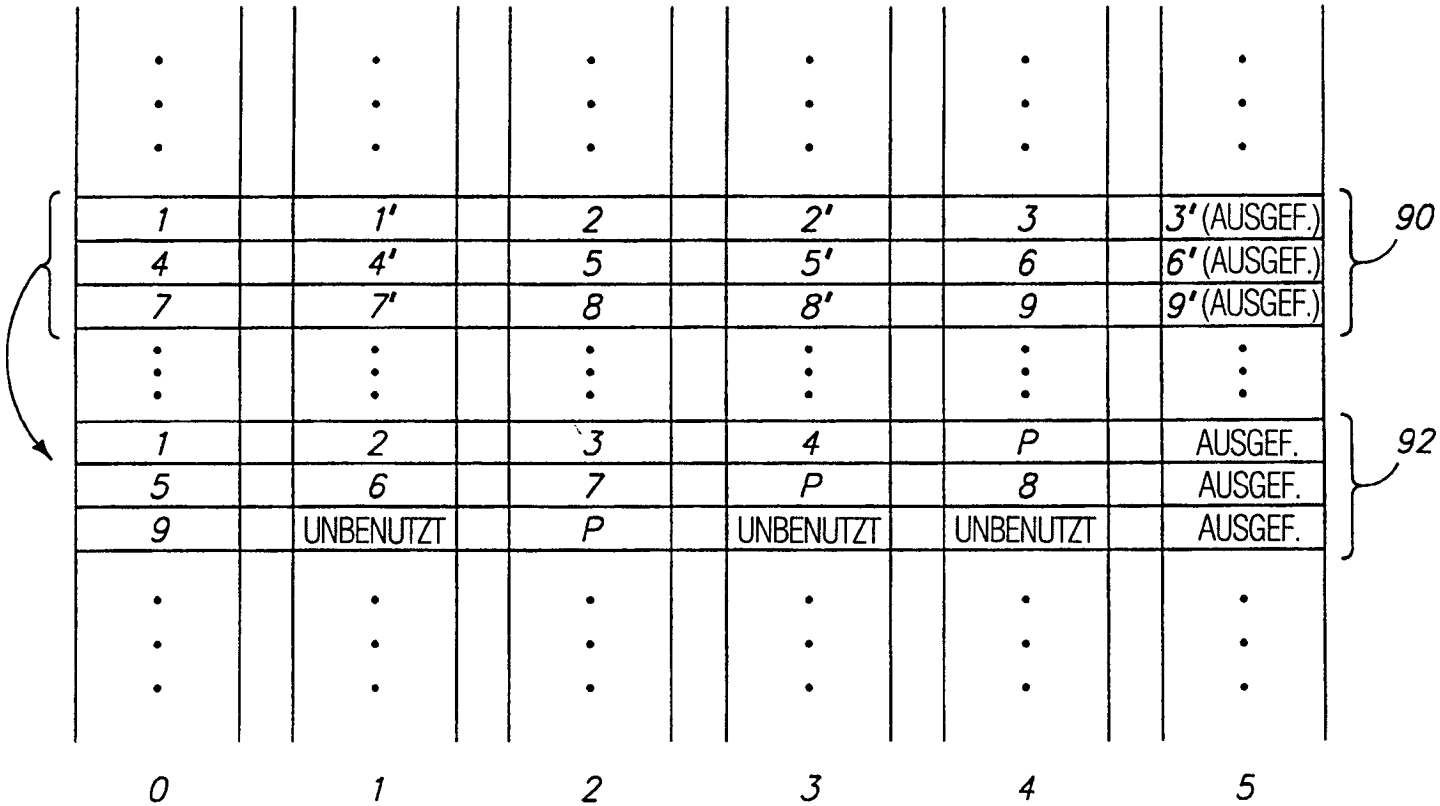
1
1'
2
2'
3
3' (AUSGEF.)
4
4'
5
5'
6
6' (AUSGEF.)
7
7'
8
8'
9
9' (AUSGEF.)
90
1
2
3
4
P
AUSGEF.
5
6
7
P
8
AUSGEF.
9
UNBENUTZT
P
UNBENUTZT
UNBENUTZT
AUSGEF.
92
0
1
2
3
4
5

Fig 7

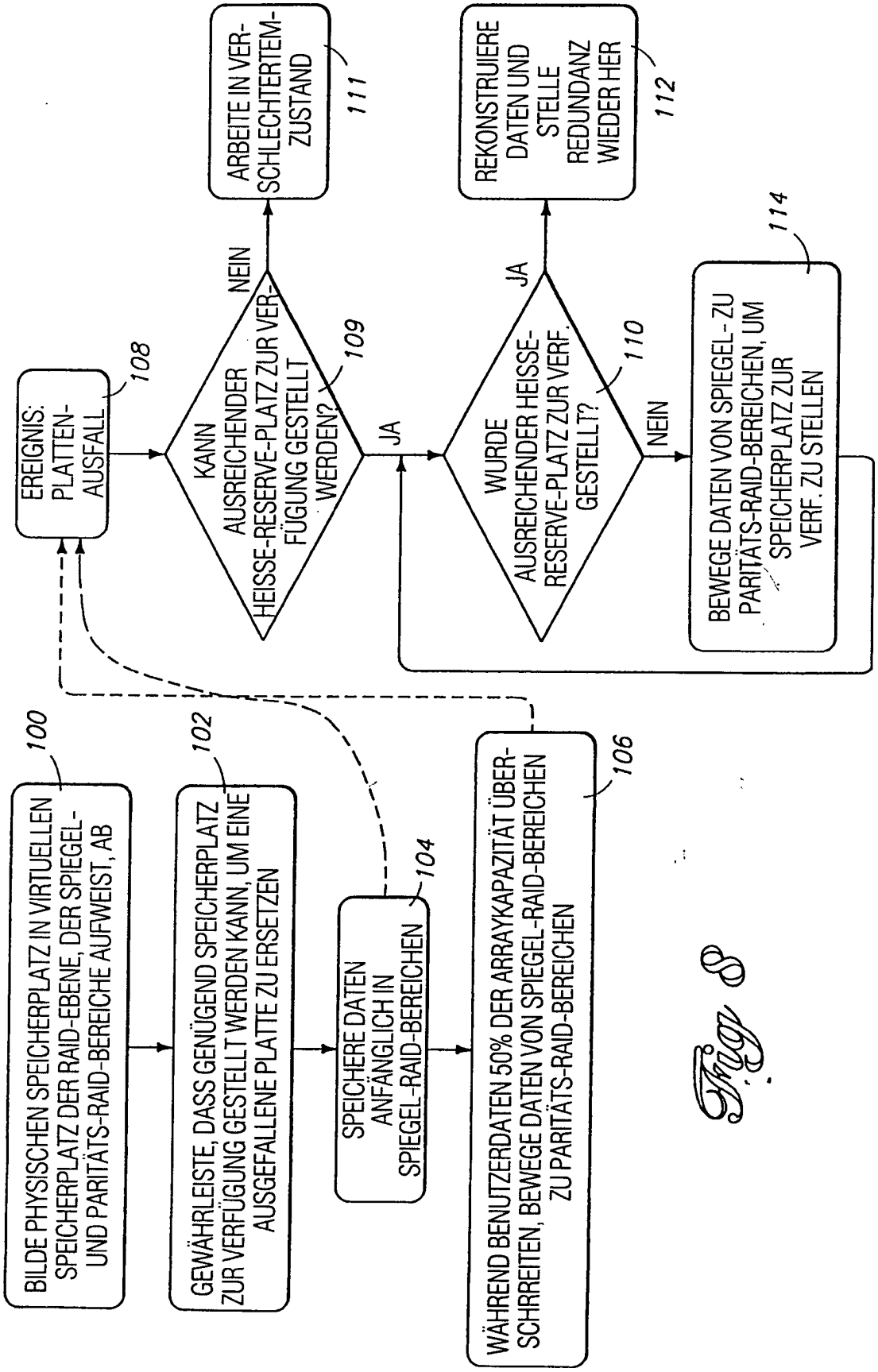
BILDE PHYSISCHEN SPEICHERPLATZ IN VIRTUELLEN SPEICHERPLATZ DER RAID-EBENE, DER SPIEGEL- UND PARITÄTS-RAID-BEREICHE AUFWEIST, AB
100
GEWÄHRLEISTE, DASS GENÜGEND SPEICHERPLATZ ZUR VERFÜGUNG GESTELLT WERDEN KANN, UM EINE AUSGEFALLENE PLATTE ZU ERSETZEN
102
SPEICHERE DATEN ANFÄNGLICH IN SPIEGEL-RAID-BEREICHEN
104
WÄHREND BENUTZERDATEN 50% DER ARRAYKAPAZITÄT ÜBERSCHRREITEN, BEWEGE DATEN VON SPIEGEL-RAID-BEREICHEN ZU PARITÄTS-RAID-BEREICHEN
106
EREIGNIS: PLATTEN-AUSFALL
108
KANN AUSREICHENDER HEISSE-RESERVE-PLATZ ZUR VERFÜGUNG GESTELLT WERDEN?
109
NEIN
ARBEITE IN VERSCHLECHTERTEM ZUSTAND
111
JA
WURDE AUSREICHENDER HEISSE-RESERVE-PLATZ ZUR VERF. GESTELLT?
110
JA
REKONSTRUIERE DATEN UND STELLE REDUNDANZ WIEDER HER
112
NEIN
BEWEGE DATEN VON SPIEGEL- ZU PARITÄTS-RAID-BEREICHEN, UM SPEICHERPLATZ ZUR VERF. ZU STELLEN
114

Fig. 8