



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 698 31 944 T2** 2006.07.06

(12) **Übersetzung der europäischen Patentschrift**

(97) **EP 0 983 548 B1**

(51) Int Cl.⁸: **G06F 11/14** (2006.01)

(21) Deutsches Aktenzeichen: **698 31 944.3**

(86) PCT-Aktenzeichen: **PCT/US98/09887**

(96) Europäisches Aktenzeichen: **98 923 431.5**

(87) PCT-Veröffentlichungs-Nr.: **WO 1998/053400**

(86) PCT-Anmeldetag: **14.05.1998**

(87) Veröffentlichungstag
der PCT-Anmeldung: **26.11.1998**

(97) Erstveröffentlichung durch das EPA: **08.03.2000**

(97) Veröffentlichungstag
der Patenterteilung beim EPA: **19.10.2005**

(47) Veröffentlichungstag im Patentblatt: **06.07.2006**

(30) Unionspriorität:
858231 19.05.1997 US

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

(73) Patentinhaber:
EMC Corp., Hopkinton, Mass., US

(72) Erfinder:
CLIFTON, J., Richard, Cary, US; CHATTERJEE, Sanjoy, Raleigh, US; LARSON, P., John, Chapel Hill, US; RICHART, R., Joseph, Raleigh, US; SAGAN, E., Cyril, Raleigh, US

(74) Vertreter:
**Kuhnen & Wacker Patent- und
Rechtsanwaltsbüro, 85354 Freising**

(54) Bezeichnung: **VORRICHTUNG UND VERFAHREN ZUR SICHERUNG EINES PLATTENSPEICHERSYSTEM**

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

HINTERGRUND DER ERFINDUNG

Bereich der Erfindung

[0001] Die vorliegende Erfindung betrifft im Allgemeinen Datensicherungssysteme zur Verwendung mit Computern und im Besonderen eine Vorrichtung und ein Verfahren zum Sichern eines Speichersystems auf Band, während das Speichersystem in Gebrauch ist.

Beschreibung des Stands der Technik

[0002] Datenverarbeitungssysteme verwenden gewöhnlich wenigstens ein Plattenlaufwerk zum Speichern von Daten und Programmen. Der Hostrechner ruft vom Plattensystem die betreffenden Informationen ab, die gerade benötigt werden, und sendet neue oder aktualisierte Informationen oder Daten oder Programme, die der Hostrechner eventuell in Zukunft benötigen wird, aber zu diesem Zeitpunkt nicht im internen Hauptspeicher braucht, zur Speicherung an das Plattensystem.

[0003] Viele Organisationen wie große Unternehmen und staatliche Körperschaften haben extrem große Datenbanken für Informationen, die für sie leicht für schnellen Zugriff und rasches Modifizieren verfügbar sein müssen. Diese Datenbanken können unter einigen Umständen einem Terabyte von Daten entsprechen oder ein Terabyte von Daten überschreiten und erfordern große Datenspeichersysteme, die mehrere Plattenlaufwerke oder Plattenlaufwerkverbände enthalten, die als ein einziges großes logisches Speichersystem organisiert sind. Oft gibt es einen Hostrechner oder Server, der ausschließlich und hauptsächlich der Abwicklung von Datenbanktransaktionen gewidmet ist. Die diversen Benutzer der Datenbank senden ihre Datenbankanforderungen an den Datenbankserver und erhalten die angeforderten Daten aus dem Datenbankserver über ein Netzwerk.

[0004] Organisationen, die derart große Datenbanken benutzen, müssen meist von Zeit zu Zeit aus geschäftlichen, rechtlichen oder archivierungstechnischen Gründen Sicherungskopien ihrer Datenbanken erstellen. Auch kann es sein, dass, obwohl moderne Plattensysteme im Allgemeinen hoch zuverlässige Vorrichtungen sind, einige Organisationen es wünschen, ihre Datenbanken als Schutz gegen die Möglichkeit eines Speichersystemausfalls sichern zu lassen.

[0005] Es ist daher allgemein üblich, regelmäßig eine teilweise oder vollständige Sicherung aller Daten im Plattensystem durchzuführen. Dies wird meist durch Kopieren der Daten, die gesichert werden sollen, auf Magnetbänder durchgeführt. Die Bänder

werden dann für eine vom Systembenutzer bestimmte Zeitspanne aufbewahrt und können nötigenfalls zum Wiederherstellen eines bekannten Zustands des Systems verwendet werden.

[0006] Eine Anzahl von kommerziellen Dienstprogrammen sind zum Durchführen von Sicherungsvorgängen erhältlich. Im typischen Fall sind diese Dienstprogramme zum Ablufen auf dem Datenbankserver vorgesehen. In einigen Fällen kann das Dienstprogramm auf einem anderen Rechnersystem laufen, das über das Lokalnetz (LAN) mit dem Datenbankserver kommuniziert. Das hat insofern Nachteile, als das Sichern einer Terabyte-Datenbank über ein Lokalnetz sehr langsam wäre und, ob das Sicherungsdienstprogramm nun auf dem Sicherungsserver läuft oder auf einem anderen Rechner am Lokalnetz, die Beteiligung des Datenbankservers während des Sicherungsprozesses erforderlich ist. Die Einbeziehung des Datenbankservers leitet Verarbeitungsenergie von den Hauptaufgaben des Servers weg und kann die Reaktionszeit gegenüber Systembenutzern verschlechtern oder die zum Vervollständigen der Sicherung erforderliche Zeit verlängern.

[0007] Ein weiteres Problem bei dem Stand der Technik ist, dass Organisationen allgemein einen „Snapshot“ (Momentaufnahme) ihrer Datenbank wünschen, wie sie zu einem bestimmten Zeitpunkt besteht. Eine Methode zum Gewährleisten der Datenkonsistenz während des Sicherns ist, den Zugriff auf die Daten während des Sicherungsvorgangs für Benutzer zu beschränken. Da die Sicherung extrem großer Datenbanken manchmal Stunden dauern kann, ist es für die Organisationen oft unakzeptabel, dass ihre Datenbanken während der Dauer der Sicherung für ihre Benutzer nicht verfügbar sind.

[0008] Systeme vom Stand der Technik wurden in der Bemühung entwickelt, dieses Problem zu lösen und es Benutzern weiterhin zu erlauben, in die Datenbank zu schreiben, während die Sicherung läuft. Beispielsweise offenbart US-Patent 5 535 381 ein System, bei dem eine Copy-on-Write-Methode zum verzögerten Kopieren (COW) verwendet wird, um „Original“-Daten vor dem Ausführen eines Schreibbefehls in separate Pufferspeicher zu sichern. Die COW-Daten werden auf dem Sicherungsband in separaten Banddatensätzen gespeichert, sodass das auf dem Band gespeicherte Bild kein Duplikat des Originalbildes ist und reintegriert werden muss, um das Originaldatenbild wieder zu erstellen.

[0009] US-Patent 5 487 160 offenbart eine weitere COW-Methode, bei der Originaldaten auf einem Reserveplattenlaufwerk gespeichert werden, bis die Sicherung abgeschlossen ist, und dann der Inhalt des Reservelaufwerks in Bausch und Bogen auf das Sicherungsband übertragen wird. Auch hier ist das Bild auf dem Sicherungsband fragmentiert und erfordert

einen Reintegrationsprozess zum Wiederausammensetzen des Originalbildes.

[0010] EP-Patent 0566967 offenbart ein Verfahren zum Erstellen eines Snapshot-Bildes eines gewünschten Abschnitts der Daten auf einer Plattenperipherie in einem Datenverarbeitungssystem. Es erstellt eine Bitmap für eine Datensicherung von der Plattenperipherie auf eine Bandperipherie. Wenn ein verzögerter Kopierprozess in der Spurumlaufspeichereinheit ausgeführt wird, werden die relevanten Daten als Nebendateien in einen Speicher geschrieben. Es wird eine separate Bitmap von Daten geführt, die aktualisiert wurden und in Nebendateien gespeichert werden. Ein Problem bei diesem Verfahren ist, dass die die Speicherstelle von Blöcken betreffenden Daten wieder in die primäre Bitmap zurück geschrieben werden müssen und die Größe der Bitmapdatei je nach der Menge der als Nebendateien in den Speicher geschriebenen Daten groß sein kann.

[0011] Die vorliegende Erfindung löst diese Probleme und Nachteile, indem sie Benutzern während des Sicherungsprozesses uneingeschränkter Zugang zum System erlaubt, während sie ein Snapshot-Sicherungsbild auf Band erstellt, das nicht rekonstruiert werden muss.

ZUSAMMENFASSUNG DER ERFINDUNG

[0012] Die vorliegende Erfindung, wie sie von den angehängten unabhängigen Ansprüchen definiert wird, betrifft ein Verfahren und eine Vorrichtung zum Sichern einer Datenbank oder eines Speichersystems auf einem Bandsicherungssystem unter Verwendung von im Handel erhältlichen Software-Dienstprogrammen.

[0013] Es ist eine Aufgabe der Erfindung, ein Sicherungsband zu erstellen, das einen Snapshot der Datenbank enthält.

[0014] Es ist eine weitere Aufgabe der Erfindung, schnelles konsistentes Sichern der Datenbank zu erlauben, während Benutzer weiterhin Zugriff haben.

[0015] Es ist ein Merkmal der Erfindung, dass ein separates Sicherungsgerät zum Abwickeln der Übertragung der Datenbank von dem Plattensystem auf das Bandsystem verwendet wird.

[0016] Es ist ein weiteres Merkmal der Erfindung, dass Blöcke der Original-Datenbankdaten vor der Übertragung auf das Sicherungsbandsystem vorübergehend gespeichert und wieder in das Datenbankbild integriert werden, bevor die Übertragung auf das Sicherungsbandsystem stattfindet.

[0017] Es ist ein Vorteil der Erfindung, dass standardmäßige im Handel erhältliches Siche-

rungs-Dienstprogramme eingesetzt werden können.

[0018] Es ist ein weiterer Vorteil der Erfindung, dass der Hostprozessor nicht an der Sicherung teilnehmen muss.

[0019] Es ist ein weiterer Vorteil der Erfindung, dass das Sicherungsbild auf Band keine spätere Reintegration oder Rekonstruktion erfordert.

[0020] Andere Merkmale und Vorteile der vorliegenden Erfindung sind für normal fachkundige Personen nach Bezugnahme auf die ausführliche Beschreibung der bevorzugten Ausgestaltung und Zeichnungen hierin verständlich.

KURZE BESCHREIBUNG DER ZEICHNUNGEN

[0021] [Fig. 1](#) ist ein Blockdiagramm eines erfindungsgemäßen Datenverarbeitungssystems zum Durchführen einer Sicherung.

[0022] [Fig. 2](#) zeigt die funktionelle Organisation des Datenverarbeitungssystems von [Fig. 1](#).

[0023] [Fig. 3](#) zeigt den Fluss von COW **231**.

[0024] [Fig. 4](#) zeigt den Fluss von ODCU (Originaldaten-Cache-Dienstprogramm) **243**.

[0025] [Fig. 5](#) zeigt den Fluss von SRU (spezielles Lesedienstprogramm) **242**.

[0026] [Fig. 6](#) zeigt die Organisation von MM (modifizierte Map) **245**.

BESCHREIBUNG DER BEVORZUGTEN AUSGESTALTUNG

[0027] Die bevorzugte Ausgestaltung und das bevorzugte Verfahren werden bezüglich dem Sichern einer in einem Plattenspeichersystem befindlichen großen Datenbank auf eine Bandperipherie besprochen. Normal fachkundige Personen werden verstehen, dass die hierin beschriebene(n) Vorrichtung und Methoden nicht auf Datenbanksicherungen begrenzt sind, sondern auch leicht zum Durchführen einer Sicherung eines beliebigen ausgewählten Abschnitts der Informationen in einem Speichersystem oder zum Durchführen einer vollständigen Sicherung aller Informationen in Speicher **130** angepasst werden könnten.

[0028] In [Fig. 1](#) sind mehrere Benutzer **101** über Lokalnnetz **110** mit Server **120** verbunden. Speicher **130** kommuniziert über Bus **162**, entweder SCSI oder Fibre-Channel, mit Server **120**. In einer bevorzugten Ausgestaltung eines Systems, bei dem Datenbank **131** in der Größenordnung von einem Terabyte Daten ist, könnte Server **120** beispielsweise ein Data Gene-

ral AViiON Modell **6600** Rechner sein, auf dem das Microsoft NT Betriebssystem läuft, und Speicher **130** könnte ein oder mehrere intelligente RAID-Festplattenverbände sein, z.B. Data General CLARiiON Plattenverbände, die als ein einzelner logischer Platten-speicher organisiert sind. Speicher **130** enthält einen oder mehrere interne Prozessoren, die in [Fig. 1](#) kollektiv als Speicherprozessor (SP) **132** abgebildet sind. SP **132** kann Programme ausführen und auf Befehle von Server **120** reagieren. SP **132** bietet die Fähigkeit von Speicher **130** zum Durchführen von Tasks wie dem unten beschriebenen verzögerten Kopieren (COW) ohne die Beteiligung von Server **120**. Speicher **130** speichert verschiedene Daten und Programme, darunter Datenbank **131**, die von Server **120** verwendet wird. Datenbank **131** könnte von einem von mehreren kommerziellen Datenbankanbietern bereitgestellt sein, z.B. Oracle, Microsoft oder Informix.

[0029] Sicherungsgerät (BA) **140** ist mit Speicher **130**, Sicherungsspeichereinheit (BSU) **150** und Lokalnetz (LAN) **110** verbunden. BA **140** kann ein beliebiger Rechner nach Wahl der Benutzer sein, der mit Speicher **130**, BSU **150** und Lokalnetz **110** verbunden werden und arbeiten kann und ausreichend Verarbeitungskapazität zum Ausführen des Sicherungs-Dienstprogramms und zum Handhaben der Sicherungsdaten hat. BA **140** kann typisch ein kleinerer Rechner als Server **120** sein, beispielsweise ein Data General AViiON Modell **3600**. BA **140** ist über eine Standardnetzverbindung **164** mit LAN **110** und über SCSI- oder Fibre-Channel-Busse **163** und **165** mit Speicher **130** und BSU **150** verbunden. BA **140** kann – muss aber nicht – die gleiche Betriebssystemsoftware wie Server **120** ausführen. BA **140** weist Prozessor **142** und Speichersystem **141** auf, das aus RAM-Speicher **144** und Plattenlaufwerk **143** besteht. Wie unten noch ausführlich besprochen wird, speichert Speichersystem **141** vorübergehend Originaldatenblöcke aus Speicher **130** und verwandte Informationen. BSU **150** könnte jedes beliebige Bandsystem großer Kapazität sein, beispielsweise ein Data General DLT Array-System.

[0030] In der Ausgestaltung von [Fig. 1](#) kommuniziert BU (Sicherungsdienstprogramm) **141** mit Server **120** über LAN **110**, um die Vorbereitung auf den Sicherungsprozess und das Einleiten des Sicherungsprozesses zu koordinieren. Unmittelbar vor dem Einleiten der Sicherung muss die Datenbank von Server **120** in einen Wartezustand versetzt werden. Dies erfordert im typischen Fall das vorübergehende Ausschließen neuer Datenbanktransaktionen, bis alle ausstehenden Transaktionen abgeschlossen worden sind und alle notwendigen Datenbankinformationen von Server **120** in Speicher **130** geschrieben worden sind. Dies gewährleistet, dass die Datenbank in einem stabilen konsistenten Zustand ist. Die Unterbrechung der normalen Datenbankverarbeitungsabläufe

wird meist schnell und mit minimaler Störung für die Benutzer erreicht. Sobald das System in dem Wartezustand ist, kann die Sicherung gestartet werden und Benutzer können das System weiter normal benutzen, während die Sicherung stattfindet.

[0031] In [Fig. 2](#) wird die funktionelle Organisation einer bevorzugten Ausgestaltung eines erfindungsgemäßen Sicherungssystems abgebildet. Speicher **130** enthält Copy-on-Write-Programm (COW-Programm) **231** zur Ausführung durch SP **132** während des Sicherungsvorgangs. COW **231** kontrolliert auf Schreibbefehle an Speicher **130** von Server **120** und suspendiert, wenn ein Schreibbefehl erfasst wird, die Ausführung des Schreibvorgangs, bis eine Kopie der aktuell an dieser Speicheradresse in Speicher **130** befindlichen Daten hergestellt wurde. In der hierin beschriebenen Ausgestaltung von Speicher **130** werden Daten im Speicher **130** adressiert, kopiert und in Blöcken von 512 Bytes verschoben.

[0032] Die modifizierte Map (MM) **245** und der Originaldaten-Cache (ODC) **246** repräsentieren Bereiche des Speichersystems **141**. In einer bevorzugten Ausgestaltung sind MM **245** und ODC **246** in Plattenlaufwerk **143** befindliche Dateien und MM **245** ist in den Speicher eingebündelt (memory-mapped), um die Zugriffsgeschwindigkeit zu erhöhen. Wie unten noch ausführlicher besprochen wird, ist die Aufgabe des ODC **246**, aus Speicher **130** kopierte Blöcke von „Originaldaten“ vorübergehend aufzunehmen, bis sie zum Konstruieren des in BSU **150** zu speichernden Snapshot-Datenbankbildes benötigt werden.

[0033] MM **245** enthält einen Eintrag für jeden während der Sicherung in ODC **246** gespeicherten Block. Die Übertragung des assoziierten Blocks von ODC **246** zu BSU **150** oder das Überschreiben des Blocks in ODC **246** verursacht kein Entfernen des Eintrags in MM **245**. Der Inhalt und die Organisation der MM **245** ist in [Fig. 6](#) abgebildet und wird unten besprochen.

[0034] BU **241**, spezielles Lesedienstprogramm (SRU) **242** und Originaldaten-Cache-Dienstprogramm (ODCU) **243** sind in Prozessor **142** laufende Programme. BU **241** ist ein im Handel erhältliches Sicherungsdienstprogramm nach Wahl des Systembenutzers, beispielsweise Legato Networker oder Ch eyenne ARCserver. Der Betrieb von SRU **242** und ODCU **243** ist für BU **241** transparent und BU **241** braucht zur Verwendung in dem offenbarten System nicht modifiziert zu werden. BU **241** glaubt, dass es direkt mit Speichers **130** kommuniziert und dass es die volle Kontrolle über die Sicherung hat.

[0035] SRU **242** und ODCU **243** kommunizieren mit und steuern MM **245** und ODC **246**. SRU **242** führt die folgenden Aufgaben durch: (1) Weiterleiten von Datenleseanforderungen von BU **241** an Speicher

130, (2) Zurückhalten der angeforderten Daten vom Speicher **130**, (3) Ablegen der erhaltenen Daten in einem RAM-144-Pufferspeicher, während die Speicher-130-adressen der erhaltenen Blöcke mit MM **245** verglichen werden, um zu ermitteln, ob irgendwelche früher aus diesen Adressen gelesenen Datenblöcke sich bereits in ODC **246** befinden, (4) falls sich einer oder mehrere derartige bereits gelesene Blöcke in ODC **246** befinden, Ersetzen der aktuell im Pufferspeicher befindlichen Blöcke aus den gleichen Adressen durch jene Blöcke aus ODC **246**, (5) Modifizieren von MM **245**, um anzuzeigen, dass jene Blöcke zu BU **241** übertragen wurden, und (6) Weiterleiten des Inhalts des wieder abgeglichenen Pufferspeichers an BU **241**.

[0036] In einer bevorzugten Ausgestaltung filtert SRU **242** auch die Schreibbefehle aus BU **241**. Schreibbefehle an Subsysteme von BA **140** werden zugelassen, aber Versuche von BU **241**, während des Sicherungsprozesses in Speicher **130** zu schreiben, falls sie vorkommen, werden nicht erlaubt.

[0037] ODCU **243** führt die folgenden Aufgaben durch: (1) Überwachen von COW-Datenübertragungen aus Speicher **130**, (2) Ermitteln, ob die aus Speicher **130** erhaltenen COW-Blöcke Teil des Bereichs von Speicher **130** sind, der gesichert wird, (3) wenn ja, Vergleichen der Adresse von jedem der erhaltenen COW-Blöcke mit dem Inhalt von MM **245**, um zu ermitteln, ob bereits ein Eintrag für einen der erhaltenen Sicherungsblöcke existiert, (4) Erstellen eines Eintrags in MM **245** für jeden erhaltenen Original-COW-Block und (5) Speichern von Original-COW-Blöcken in ODC **246**.

[0038] In [Fig. 6](#) ist die funktionelle Organisation von MM **245** abgebildet. Beim Einleiten der Sicherung ist MM **245** leer. Wenn die Sicherung gestartet wurde, veranlasst jeder Schreibbefehl von Server **120** zu einem Datenblock in Speicher **130**, dass COW **231** zuerst die existierenden Daten aus dieser Adresse kopiert und sie als einen COW-Block an ODCU **243** sendet. Wenn ODCU **243** einen COW-Block aus Datenbank **131** erhält und die Originaldaten aus diesem Block nicht bereits im ODC **246** sind, erstellt ODCU **243** einen Eintrag in MM **245** für diesen Block. Im Verlauf der Sicherung sammelt MM **245** eine Mehrzahl von Einträgen 1-N, wobei jeder Eintrag einen während der Sicherung von Speicher **130** nach ODC **246** kopierten Originaldatenblock repräsentiert. Nur der erste von einer Adresse erhaltene COW-Block wird in ODC **246** gespeichert. Spätere COW-Blöcke mit der gleichen Adresse werden von ODCU **243** ignoriert.

[0039] Jeder Eintrag 1-N in MM **245** umfasst vier Informationsfelder. In der hierin beschriebenen Ausgestaltung besteht die Speicheradresse jedes Blocks in Speicher **130** aus einem eindeutigen Identifikator der

spezifischen physikalischen oder logischen Komponente in Speicher **130**, in der sich dieser betreffende Datenblock befindet, und einem Offset zu dem spezifischen 512-Byte-Block in der Komponente. Diese zwei Adresskomponenten werden in den Feldern **610** und **620** von MM **245** gespeichert und stellen einen eindeutigen Identifikator für jeden Datenblock in Speicher **130** bereit. Feld **630** enthält den Zustand des assoziierten Datenblocks. Jeder Block in ODC **246** befindet sich in einem von zwei Zuständen: von BU **141** gelesen oder von BU **141** noch nicht gelesen. Und schließlich enthält Feld **640** den Offset in den ODC **246**, wo sich der assoziierte Datenblock befindet.

[0040] Wenn Datenbank **131** in einen Wartezustand versetzt worden ist, wird die Sicherungsoperation eingeleitet. [Fig. 3](#) illustriert den Betrieb von COW **231**. In Schritt **310** überwacht COW **231** alle Befehle von Server **120** an Speicher **130**, um Schreibbefehle zu erfassen. Wenn in Schritt **320** ein Schreibbefehl an Speicher **130** erfasst wird, wird der Befehl in Schritt **330** suspendiert und die relevanten Daten werden in Schritt **340** kopiert. Ein Schreibbefehl von Server **120** könnte einen einzelnen Datenblock oder eine Anzahl von Datenblöcken adressieren. Alle von dem Schreibbefehl adressierten Blöcke werden in Schritt **340** kopiert. Der kopierte Datenblock bzw. die kopierten Datenblöcke wird/werden in Schritt **350** an BA **140** gesendet. Nach dem Kopieren des Blocks oder der Blöcke kann der Schreibbefehl von Server **120** durchgeführt werden.

[0041] Wie fachkundige Personen gut verstehen, kann COW **231**, falls erwünscht, mit zusätzlicher Funktionalität ausgebildet sein, je nach der verfügbaren Verarbeitungskapazität und Speicherkapazität von SP **132**. In der hierin beschriebenen Ausgestaltung wird eine relativ unkomplizierte COW-231-Methode besprochen. COW **231** macht eine Kopie aller Schreibvorgänge nach Speicher **130** und suspendiert den gesamten Schreibvorgang, bis alle Blöcke kopiert worden sind. In einer alternativen Ausgestaltung könnte COW **231** eine relativ große Schreibforderung, d.h. eine Schreibforderung, die eine relativ große Anzahl von Blöcken betrifft, in eine Anzahl kleinerer inkrementeller Schreibvorgänge unterteilen. COW **231** könnte dann jeden inkrementellen Schreibvorgang suspendieren, bis der Abschnitt der von diesem inkrementellen Schreibvorgang adressierten Blöcke kopiert worden ist. In einer weiteren Ausgestaltung könnte COW **231** eine Map der spezifischen zu sichernden Adressen führen und nur dann eine Kopie herstellen, wenn Schreibvorgänge für diese spezifischen Speicherbereiche vorgesehen sind. In noch einer weiteren alternativen Ausgestaltung könnte COW **231** eine Tabelle der Adressen von Blöcken in Speicher **130** führen, die bereits kopiert wurden, und verzögertes Kopieren (Copy-on-Write) nur für jene Blöcke durchführen, die während der Siche-

zung nicht schon einmal kopiert wurden.

[0042] **Fig. 4** illustriert die Art und Weise, wie ODCU **243** die in Schritt **350** zu BA **140** gesendeten COW-Blöcke handhabt. Beim Einleiten des Sicherns leitet ODCU **243** Schritt **410** ein, die Überwachung auf COW-Blöcke von Speicher **130**. Wenn ein Block oder mehrere Blöcke in Schritt **420** aus Speicher **130** erhalten werden, werden die Adressen der erhaltenen Blöcke in Schritt **430** überprüft, um zu ermitteln, ob die erhaltenen Daten Teil des Abschnitts von Speicher **130** sind, der gesichert wird. Wenn die erhaltenen Daten nicht Teil des Bereichs von Speicher **130** sind, der gesichert wird, werden die Blöcke in Schritt **440** ignoriert und ODCU **243** setzt sein Überwachen fort. Wenn die erhaltenen Daten Teil des Bereichs von Speicher **130** sind, der gesichert wird, wird MM **245** überprüft, um zu ermitteln, ob ein Eintrag für den eindeutigen Identifikator von einem der erhaltenen Blöcke bereits in MM **245** vorliegt, was bedeutet, dass die Originaldaten von diesem Block bereits in ODC **246** gespeichert wurden. Wenn in MM **245** bereits ein Eintrag für eine Blockadresse existiert, dann repräsentieren die gerade erhaltenen Daten einen nachfolgenden Schreibvorgang zu demselben Block während der Sicherung. In diesem Fall wird der erhaltene Block ignoriert und die Überwachung in Schritt **410** fortgesetzt.

[0043] Wenn die Adresse eines erhaltenen Blocks in Schritt **460** in MM **245** nicht gefunden wird, dann wird in Schritt **470** für diesen Block ein neuer Eintrag in MM **245** erstellt. Zum Durchführen einer Sicherung einer einzelnen Datenbank ist es unnötig, Datenblöcke nach ihrer Sicherung in ODC **246** zu behalten. ODCU **243** kann daher den ODC-Offset eines Blocks wiederverwenden, der bereits zu BSU **150** verschoben wurde, wie von Zustand **630** angezeigt wird, wenn kein derartiger Offset verfügbar ist. Wenn kein zuvor benutzter ODC-Offset verfügbar ist, weist ODCU **243** einen neuen ODC-Offset zu. Als eine alternative Ausgestaltung kann die Wiederverwendung von ODC-Offsets verboten werden, wodurch sichergestellt wird, dass jedem Eintrag in MM **245** während der Sicherung ein eindeutiger ODC-Offset zugewiesen wird und alle ODC-Datenblöcke in ODC **246** behalten werden. Schließlich wird der erhaltene Block in Schritt **480** in ODC **246** gespeichert.

[0044] **Fig. 5** illustriert den Betrieb von SRU **242**. Wie oben angedeutet, aus der Sicht von BU **241**, gibt BU **241** Lesebefehle an Speicher **130** aus und erhält die angeforderten Daten zurück, die es dann wiederum an BSU **150** sendet. Tatsächlich ist SRU **242** zwischen BU **241** und Speicher **130** angeordnet. In Schritt **510** überwacht SRU **242** auf Lesebefehle von BU **241** an Speicher **130**. Lesebefehle werden in Schritt **540** an Speicher **130** weitergeleitet. Wenn der Lesebefehl nicht an Speicher **130** gerichtet ist, wird der Lesebefehl zur Durchführung in Schritt **530** an

sein Ziel weitergeleitet. Speicher **130** ruft die angeforderten Daten in Schritt **540** ab und sendet sie zum Sicherungsgerät zurück.

[0045] Allgemein bestimmt und wählt das als BU **241** eingesetzte im Handel erhältliche Dienstprogramm die spezifische Datenmenge, die es mit jeder Leseanforderung von Speicher **130** anfordern wird. Wie oben besprochen verwendet Speicher **130** in einer bevorzugten Ausgestaltung einen 512-Byte-Datenblock als seine Standarddateneinheit für Manipulieren und Adressieren. Meist ist die von BU **241** gelesene Datenmenge beträchtlich größer als 512 Bytes. Beispielsweise könnte es sein, dass BU **241** mit jedem Lesebefehl 64 Kbyte anfordert. SRU **242** leitet die Leseanforderung an Speicher **130** weiter und legt, wenn der Speicher die angeforderten Daten zurücksendet, die zurückgesendete Datenmenge in einen Pufferspeicher in RAM **144**. An diesem Punkt können einige der Datenblöcke in dem Pufferspeicher Blöcke sein, die nach dem Einleiten der Sicherung modifiziert wurden. Um sicherzustellen, dass die Sicherung einen echten Snapshot erstellt, müssen alle derartigen modifizierten Datenblöcke in dem Pufferspeicher durch die entsprechenden Originaldatenblöcke aus ODC **246** ersetzt werden, bevor die angeforderten Daten an BU **241** geliefert werden.

[0046] Daher wird in Schritt **560** die in Speicher **130** befindliche eindeutige Adresse jedes Blocks in dem Pufferspeicher mit den in MM **245** gespeicherten Adressen verglichen, um zu ermitteln, ob Originaldaten für diese Adresse in ODC **246** verfügbar sind. In Schritt **570** wird jeder Block in dem Pufferspeicher, der eine Adresse hat, die mit einer in MM **245** gespeicherten Adresse übereinstimmt, verworfen und durch den Originaldatenblock aus ODC **246** ersetzt. In Schritt **580** wird das Zustandsfeld von MM **245** jedes Originaldatenblocks, der in den Pufferspeicher gelegt wurde, aktualisiert, sodass es reflektiert, dass der Block in Vorbereitung auf seine Sicherung in BSU **150** aus ODC **246** gelesen wurde. Die Änderung des Blockzustands in Schritt **580** dient ODCU **243** als Anzeige dafür, dass der von diesem Block in ODC **246** eingenommene Bereich überschrieben und zum Speichern eines anderen Blocks verwendet werden kann. In Schritt **590** wird der Inhalt des Pufferspeichers, der jetzt nur aus Originaldaten besteht, zur Übertragung an BSU **150** an BU **241** weitergeleitet.

[0047] Die oben beschriebene bevorzugte Ausgestaltung stellt ein Mittel zum Herstellen einer Snapshot-Sicherung von Datenbank **131** mithilfe eines separaten Sicherungsgeräts bereit, das nicht erforderlich ist, dass Server **120** am Sicherungsprozess beteiligt ist, und von Speicher **130** nur COW-Fähigkeit erfordert. Es versteht sich, dass die Funktionen von MM **245** und ODC **246** so implementiert werden könnten, dass sie in Speicher **130** liegen, und die Dienstprogramme SRU **242** und ODCU **243** könnten

so implementiert sein, dass sie in SP 132 anstelle von Prozessor 142 ausgeführt werden. In dieser alternativen Implementierung würden die Speicherung von COW-Datenblöcken und der Abgleich von COW-Daten und Datenlesevorgängen in Reaktion auf Leseanforderungen von BU 142 gänzlich innerhalb Speicher 130 erfolgen. Diese Implementierung würde den vom Prozessor 142 geforderten Verarbeitungsaufwand verringern, die Notwendigkeit für das Vorhandensein der Platte 142 verringern oder beseitigen und den Verkehr auf Bus 163 verringern, sie würde aber bewirken, dass die Arbeitslast von SP 132 erhöht würde und Speicher 130 die Speicherung von MM 245 und ODC 246 übernehmen müsste.

[0048] Die besondere Ausgestaltung oben ist in jeder Hinsicht als veranschaulichend und nicht einschränkend zu betrachten. Der Umfang der Erfindung wird von den angehängten Ansprüchen und nicht der vorhergehenden Beschreibung angezeigt.

Patentansprüche

1. Verfahren zum Erstellen einer Snapshot-Aufnahme eines gewünschten Teils der Daten, die sich auf einer Plattenperipherie in einem Datenverarbeitungssystem befinden, das einen mit der Plattenperipherie verbundenen Hostrechner und einen mit dem Hostrechner und der Plattenperipherie verbundenen Sicherungsrechner hat, wobei der Sicherungsrechner einen Originaldaten-Cache und eine Map hat, die für jeden in dem Originaldaten-Cache gespeicherten Datenblock die entsprechende Adresse des genannten Datenblocks in der Plattenperipherie enthält, wobei das Datenverarbeitungssystem ferner eine mit dem Sicherungsrechner verbundene Bandperipherie hat, wobei das Verfahren die folgenden Schritte umfasst:

- a) Lesen von ein oder mehreren Blöcken der gewünschten Daten aus der Plattenperipherie;
- b) Ermitteln in einem ersten Ermittlungsschritt durch Vergleichen der Adressen in der Map mit den Adressen der erhaltenen Blöcke, ob ein oder mehrere Datenblöcke, die zuvor aus der gleichen Plattenperipherieadresse gelesen wurden wie die ein oder mehreren in Schritt a) gelesenen Datenblöcke, in dem Originaldaten-Cache verfügbar sind;
- c) wenn ja, Übertragen der ein oder mehreren zuvor gelesenen Datenblöcke aus dem Originaldaten-Cache auf die Bandperipherie mithilfe der Map, um die Daten aus dem Originaldaten-Cache für die ein oder mehreren in Schritt a) gelesenen Blöcke einzusetzen; und
- d) wenn nicht, Übertragen der ein oder mehreren in Schritt a) aus der Plattenperipherie gelesenen Datenblöcke auf die Bandperipherie.

2. Verfahren nach Anspruch 1 und ferner umfassend den Schritt des Wiederholens aller Schritte a) bis d) des Verfahrens nach Anspruch 1, bis der ge-

samte gewünschte Teil auf die Bandperipherie übertragen worden ist.

3. Verfahren nach Anspruch 1, ferner umfassend: vor dem Schritt b) und wenn mehr als ein Datenblock aus der Plattenperipherie gelesen wurde, Ablegen wenigstens einiger dieser Datenblöcke in einem Pufferspeicher; dann Durchführen eines zweiten Ermittlungsschrittes für jeden in dem Pufferspeicher abgelegten Block: Ermitteln, ob ein zuvor aus der gleichen Plattenperipherieadresse gelesener Datenblock in dem Originaldaten-Cache verfügbar ist; und wenn ja, Ersetzen des derzeit in dem Pufferspeicher befindlichen Datenblocks durch den zuvor gelesenen Datenblock; Übertragen des Inhalts des Pufferspeichers auf die Bandperipherie; Fortfahren mit allen Verfahrensschritten nach Anspruch 1, angefangen mit Schritt b), für alle nicht in dem Pufferspeicher abgelegten Datenblöcke.

4. Verfahren nach Anspruch 3 und ferner umfassend den Schritt des Wiederholens aller Schritte des Verfahrens nach Anspruch 3, bis der gesamte gewünschte Teil auf die Bandperipherie übertragen worden ist.

5. Verfahren nach Anspruch 1, ferner umfassend den Schritt des Überwachens auf einen Schreibbefehl von dem Hostrechner an die Plattenperipherie; Durchführen der folgenden Schritte vor dem Schritt b), wenn ein Schreibbefehl erfasst wird:

- (i) Suspendieren des Schreibbefehls;
- (ii) Kopieren der Daten aus den Plattenperipherieadressen, an die der Schreibbefehl gerichtet ist;
- (iii) Speichern der in Schritt (ii) kopierten Daten in dem Originaldaten-Cache und
- (iv) Ausführen des Schreibvorgangs.

6. Verfahren nach Anspruch 5, ferner umfassend den Schritt des Wiederholens aller Schritte des Verfahrens nach Anspruch 5, bis der gesamte gewünschte Teil auf die Bandperipherie übertragen worden ist.

7. Verfahren nach Anspruch 1, bei dem die Plattenperipherie Copy-on-Write-Fähigkeit hat und das ferner das Durchführen eines Copy-on-Write-Prozesses umfasst, wobei das Verfahren die folgenden Schritte umfasst:

Überwachen auf einen Schreibbefehl von dem Hostrechner an die Plattenperipherie und, falls vor Schritt b) ein Schreibbefehl erfasst wird, Durchführen der folgenden Schritte:

- (i) Suspendieren des Schreibbefehls;
- (ii) Kopieren der Daten von den Adressen, an die der Schreibbefehl gerichtet ist;
- (iii) Durchführen der folgenden Schritte für jeden in Schritt (ii) kopierten Datenblock:

(1) Vergleichen der Adresse des Blocks mit einer Lis-

te von Adressen von Blöcken, die zuvor in einem Originaldaten-Cache gespeichert wurden;

(2) Verwerfen des Blocks, wenn sich die Adresse des Blocks auf der Liste befindet;

(3) wenn die Adresse des Blocks sich nicht auf der Liste befindet, Speichern des Blocks in dem Originaldaten-Cache und Hinzufügen der Adresse des Blocks zu der Liste; und

(4) Ausführen des Schreibvorgangs; und wenn vor dem ersten Ermittlungsschritt kein Schreibbefehl erfasst wird, Schritt b) und alle übrigen Schritte nach Anspruch 1.

8. Verfahren nach Anspruch 7, bei dem der Schritt (iii) (3) den zusätzlichen Schritt des Speicherns eines Zeigers auf die Speicherstelle des Blocks in dem Sicherungsrechner umfasst.

9. Verfahren nach Anspruch 7, bei dem der Schritt (iii) (3) den zusätzlichen Schritt des Speicherns eines Zustandanzeigers für den Block umfasst, wobei der genannte Zustandsanzeiger anzeigt, ob der Block aus dem Sicherungsrechner gelesen wurde.

10. System zum Erstellen einer Snapshot-Aufnahme eines gewünschten Teils der Daten, die sich auf einer Plattenperipherie in einem Computersystem befinden, umfassend:

einen Hostrechner (**120**);

eine Plattenperipherie (**130**), die mit dem Hostrechner verbunden ist und einen internen Prozessor hat;

eine Bandperipherie (**150**) zum Speichern einer Sicherungskopie eines gewünschten Teils der Daten, die sich auf der Plattenperipherie befinden;

einen Sicherungsrechner (**140**), der mit dem Hostrechner und der Plattenperipherie verbunden ist, wobei der Sicherungsrechner einen Originaldaten-Cache und eine Map hat, die für jeden in dem Originaldaten-Cache gespeicherten Datenblock die entsprechende Adresse des genannten Datenblocks in der Plattenperipherie enthält;

wobei der Sicherungsrechner zum Lesen von einem oder mehreren Blöcken der gewünschten Daten aus der Plattenperipherie, zum Ermitteln durch Vergleichen der Adressen in der Map mit den Adressen der erhaltenen Blöcke, ob ein oder mehrere zuvor von der gleichen Plattenperipherieadresse gelesene Datenblöcke in dem Originaldaten-Cache verfügbar sind, und zum Übertragen der ein oder mehreren zuvor gelesenen Datenblöcke aus dem Originaldaten-Cache auf die Bandperipherie mithilfe der Map, um die Daten aus dem Originaldaten-Cache für die ein oder mehreren von dem Sicherungsrechner gelesenen Blöcke einzusetzen, und ansonsten Übertragen der von dem Sicherungsrechner gelesenen Datenblöcke auf die Bandperipherie.

11. System nach Anspruch 10, bei dem die Plattenperipherie Copy-on-Write-Fähigkeit hat, wobei der

Sicherungsrechner Folgendes umfasst: ein Mittel zum Erhalten von Copy-on-Write-Daten von der Plattenperipherie,

ein Mittel zum Speichern der Copy-on-Write-Daten in dem Sicherungsrechner,

ein Mittel zum Anfordern von Sicherungsdaten von der Plattenperipherie,

ein Mittel zum Erhalten von Sicherungsdaten von der Plattenperipherie in Antwort auf eine Anfrage, ein Mittel zum Übertragen von Daten auf die Bandperipherie und

ein Mittel zum Auswählen der erhaltenen Sicherungsdaten oder der gespeicherten Copy-on-Write-Daten zum Übertragen auf die Bandperipherie.

12. System nach Anspruch 10, bei dem der Sicherungsrechner Folgendes hat: ein Daten-Cache-Dienstprogramm mit einem Mittel zum Erhalten von Copy-on-Write-Daten (COW-Daten) von der Plattenperipherie und einem Mittel zum Speichern der COW-Daten in dem Originaldaten-Cache;

ein Sicherungsdienstprogramm mit einem Mittel zum Ausgeben von Lesebefehlen an die Plattenperipherie, einem Mittel zum Erhalten von Daten von der Plattenperipherie und einem Mittel zum Senden von Daten zu der Bandperipherie; und

ein Lesedienstprogramm mit einem Mittel zum Erhalten von Daten von der Plattenperipherie in Antwort auf einen Lesebefehl von dem Sicherungsdienstprogramm;

ein Mittel zum Vergleichen der Plattenperipherieadresse erhaltener Daten mit Plattenperipherieadressen von COW-Daten; ein Mittel, um dem Sicherungsdienstprogramm Daten bereitzustellen; und

ein Mittel zum Auswählen der dem Sicherungsdienstprogramm bereitzustellenden Daten aus den von der Plattenperipherie erhaltenen Daten oder den in dem Originaldaten-Cache gespeicherten COW-Daten.

13. System nach Anspruch 11 oder 12, bei dem der Sicherungsrechner ferner Folgendes umfasst:

ein Mittel zum Speichern der Plattenperipherieadresse der COW-Daten in der Map in der Map des Originaldaten-Caches; und

ein Mittel zum Zuweisen und Speichern, in der Map, eines Zeigers auf die COW-Daten in dem Originaldaten-Cachemittel zum Prüfen der Map auf gespeicherte Adressen, die mit den Adressen von Daten übereinstimmen, die von der Plattenperipherie erhalten wurden.

14. System nach Anspruch 13, bei dem das Mittel zum Auswählen für jede von dem Sicherungsdienstprogramm angeforderte Adresse die erhaltenen Daten auswählt, wenn in der Map keine entsprechende Adresse gefunden wird, und die Originaldaten-Cache-Daten auswählt, wenn in der Map eine entsprechende Adresse gefunden wird.

15. System nach Anspruch 13, bei dem das Da-

ten-Cache-Dienstprogramm ferner ein Mittel hat zum Prüfen der Map vor dem Speichern von COW-Daten in dem Originaldaten-Cache und Speichern der Daten in dem Originaldaten-Cache nur dann, wenn die Map anzeigt, dass nicht bereits Daten von der gleichen Plattenperipherieadresse in dem Originaldaten-Cache gespeichert sind.

16. System nach Anspruch 14, bei dem das Daten-Cache-Dienstprogramm ferner ein Mittel zum Speichern eines Zustandsfeldes in der Map für jeden COW-Eintrag hat, wobei das genannte Zustandsfeld anzeigt, ob die entsprechenden COW-Daten in dem Originaldaten-Cache von dem Lesedienstprogramm-Auswahlmittel ausgewählt wurden oder nicht.

17. System nach Anspruch 16, bei dem das Lesedienstprogramm ferner ein Mittel zum Modifizieren des Inhalts des Zustandsfeldes hat, wenn das Lesedienstprogramm die mit diesem Zustandsfeld assoziierten COW-Daten aus dem Originaldaten-Cache auswählt.

18. System nach Anspruch 13, bei dem das Mittel zum Zuweisen des Zeigers ein Mittel hat zum Neuzuweisen eines zuvor zugewiesenen Zeigers, wenn das mit dem zuvor zugewiesenen Zeiger assoziierte Zustandsfeld anzeigt, dass die mit diesem Zeiger assoziierten COW-Daten von dem Lesedienstprogramm-Auswahlmittel ausgewählt wurden.

Es folgen 6 Blatt Zeichnungen

Anhängende Zeichnungen

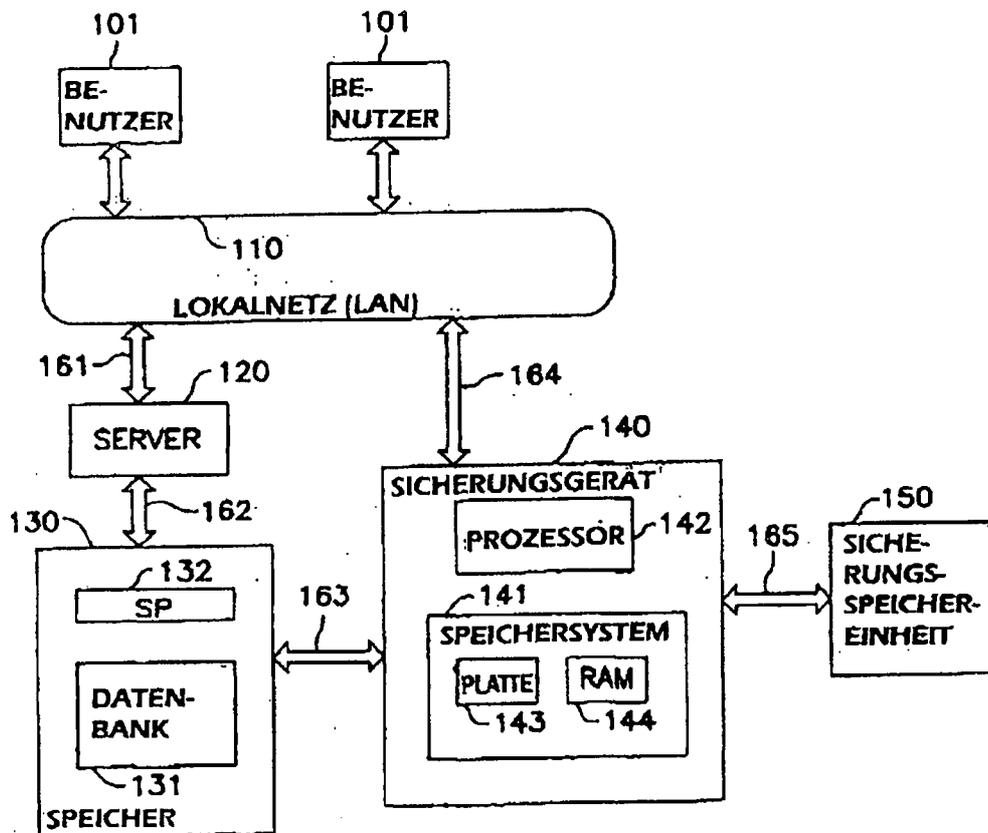


FIG. 1

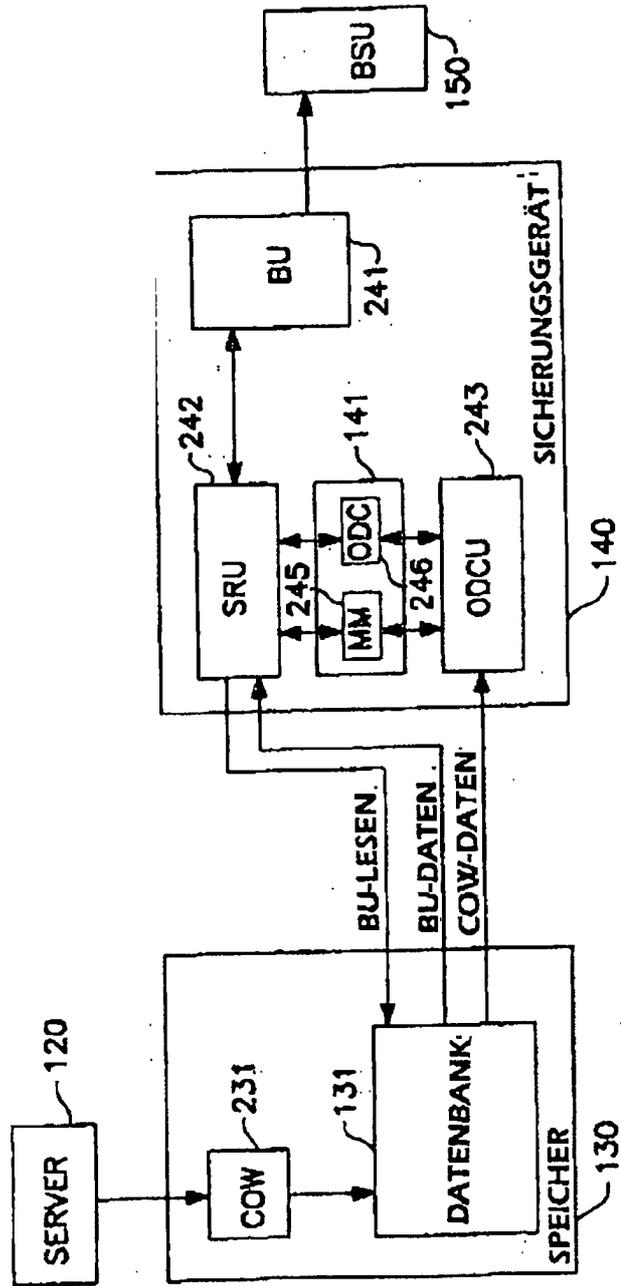


FIG. 2

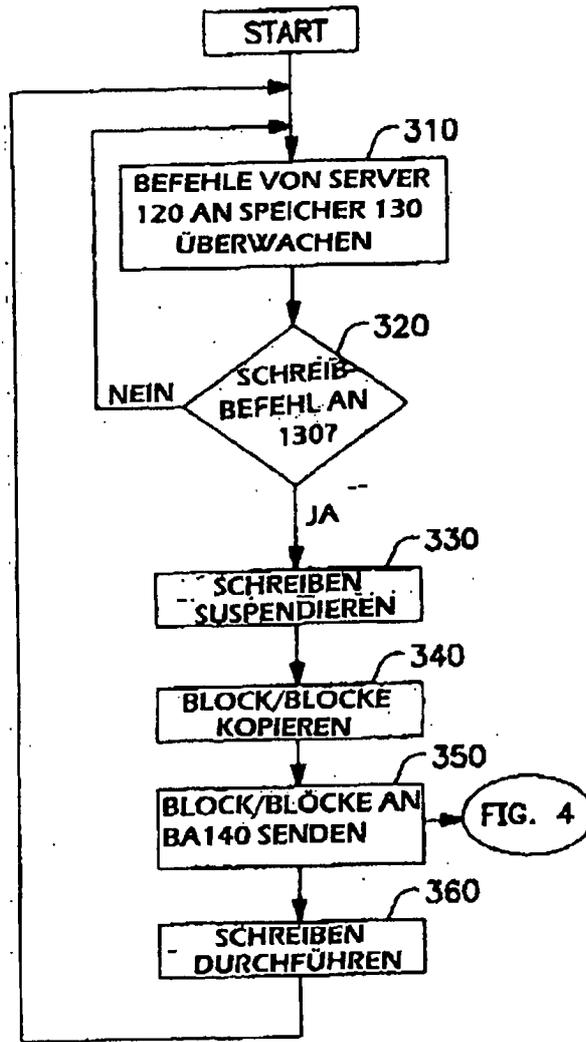


FIG. 3

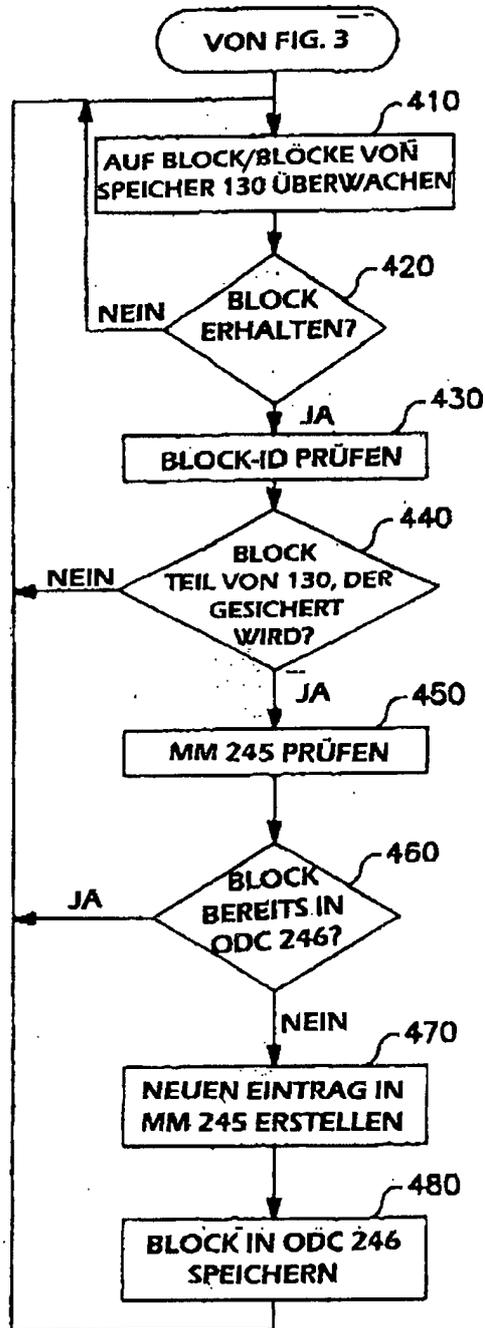


FIG. 4

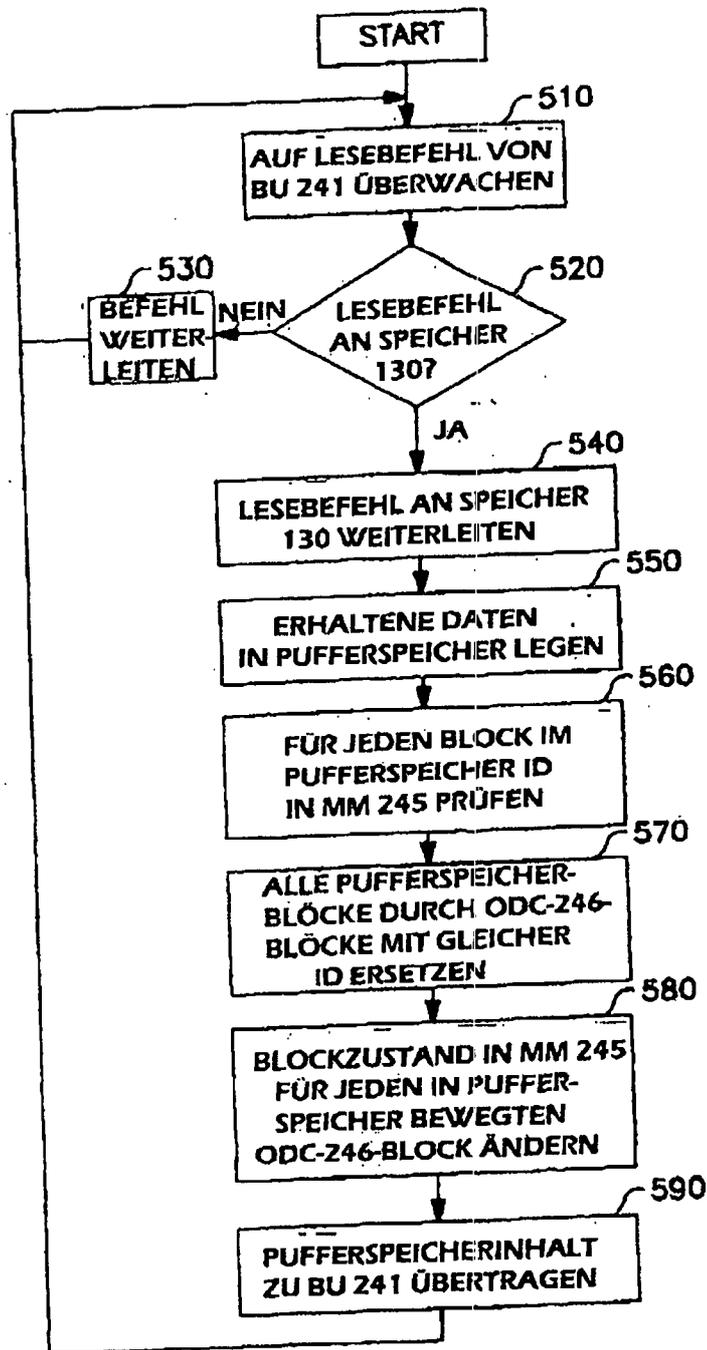


FIG. 5

	610	620	630	640
EINTRAG 1	UI	UI-OFFSET	ZUSTAND	ODC-OFFSET
EINTRAG 2				
EINTRAG N-1				
EINTRAG N				

MODIFIZIERTE MAP 245

FIG. 6