



(10) **DE 10 2012 210 914 B4** 2018.09.13

(12) **Patentschrift**

(21) Aktenzeichen: **10 2012 210 914.5**
 (22) Anmeldetag: **27.06.2012**
 (43) Offenlegungstag: **10.01.2013**
 (45) Veröffentlichungstag
 der Patenterteilung: **13.09.2018**

(51) Int Cl.: **G06F 15/177 (2006.01)**
G06F 15/173 (2006.01)
H04L 12/50 (2006.01)
H04L 29/14 (2006.01)

Innerhalb von neun Monaten nach Veröffentlichung der Patenterteilung kann nach § 59 Patentgesetz gegen das Patent Einspruch erhoben werden. Der Einspruch ist schriftlich zu erklären und zu begründen. Innerhalb der Einspruchsfrist ist eine Einspruchsgebühr in Höhe von 200 Euro zu entrichten (§ 6 Patentkostengesetz in Verbindung mit der Anlage zu § 2 Abs. 1 Patentkostengesetz).

(30) Unionspriorität:
13/177,680 **07.07.2011** **US**

(73) Patentinhaber:
Lenovo Enterprise Solutions (Singapore) Pte. Ltd., Singapore, SG

(74) Vertreter:
Schweiger & Partners, 80687 München, DE

(72) Erfinder:
Armstrong, William J., Rochester, Minn., US;
Borkenhagen, John M., Rochester, Minn., US;
Crippen, Martin J., Research Triangle Park, N.C., US;
Desai, Dhruv M., Research Triangle Park, N.C., US;
Engbrechtsen, David R., Rochester, Minn., US;
Hillier III, Philip R., Rochester, Minn., US;
Holland, William G., Research Triangle

Park, N.C., US; Hughes, James E., Research Triangle Park, N.C., US; O'Connor, James A., Poughkeepsie, N.Y., US; Patel, Pravin S., Research Triangle Park, N.C., US; Tri, Steven M., Rochester, Minn., US

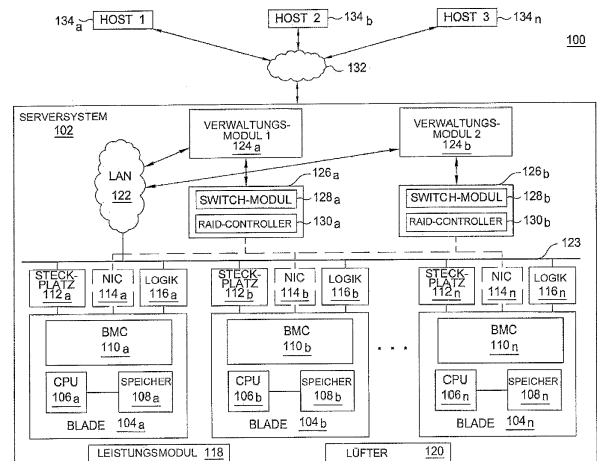
(56) Ermittelter Stand der Technik:

CHEN, Kai, et al. Generic and automatic address configuration for data center networks. In: ACM SIGCOMM Computer Communication Review. ACM, 2010. S. 39-50. doi: 10.1145/1851275.1851190

NIRANJAN MYSORE, Radhika, et al. Portland: a scalable fault-tolerant layer 2 data center network fabric. In: ACM SIGCOMM Computer Communication Review. ACM, 2009. S. 39-50. doi: 10.1145/1592568.1592575

(54) Bezeichnung: **Switch-Fabric-Management**

(57) Hauptanspruch: System, das Folgendes umfasst:
 eine Mittelplatine, die eine Fabric-Verbindung für eine Switch-Fabric umfasst;
 eine oder mehrere Serverkarten, die funktionsmäßig mit der Mittelplatine verbunden sind, wobei jede Serverkarte einen oder mehrere Computerprozessoren und einen Speicher umfasst;
 eine Vielzahl von Switch-Modulen, die funktionsmäßig mit der Mittelplatine verbunden sind, wobei jedes Switch-Modul dafür konfiguriert ist, den Netzverkehr für mindestens eine der einen oder mehreren Serverkarten zu schalten; und
 einen Verwaltungscontroller, der dafür konfiguriert ist, die Vielzahl von Switch-Modulen über die Fabric-Verbindung zu verwalten.



Beschreibung

HINTERGRUND

[0001] Obwohl frühe Computerarchitekturen eigenständige Einzelcomputer (häufig als Personal Computer (PCs) bezeichnet) verwendet haben, nutzen leistungsstärkere moderne Computersysteme oftmals mehrere Computer, die in einem gemeinsamen Gehäuse miteinander verbunden sind. Ein beispielhaftes gemeinsames Gehäuse ist bekannt als Blade-Gehäuse, welches eine Vielzahl von Server-Blades umfasst, die über ein gemeinsames Backbone im Blade-Gehäuse verbunden sind. Jeder Server-Blade ist eine steckbare Platine, die mindestens einen Prozessor, einen integrierten Speicher und eine Ein-/Ausgangs- (E/A-)Schnittstelle umfasst. Die Vielzahl von Server-Blades sind für einen Datenaustausch untereinander und eine Nutzung der gemeinsamen Ressourcen, wie zum Beispiel Speichereinheiten, Bildschirme, Eingabeeinheiten usw., konfiguriert. Außerdem können ein oder mehrere Blade-Gehäuse ein Blade-System bilden, das oftmals einem einzelnen Unternehmen und/oder einer speziellen Aufgabe, wie der Kreditbearbeitung, Lohnkostenverwaltung usw., zugeordnet ist.

[0002] Die Druckschrift NIRANJAN MYSORE, Radhika et al. Portland: a scaleable fault-tolerant layer 2 data center network fabric offenbart ein Datencenternetzwerk, das einen lokalen zentralisierten fabric manager aufweist.

KURZDARSTELLUNG

[0003] Eine Ausführungsform der Erfindung stellt ein System bereit, das eine Mittelplatine (midplane), eine oder mehrere Serverkartentypen, eine Vielzahl von Switch-Modulen und einen Verwaltungscontroller umfasst. Die Mittelplatine umfasst eine Fabric-Verbindung für eine Switch-Fabric. Die eine oder die mehreren Serverkarten sind funktionsmäßig mit der Mittelplatine verbunden, wobei jede Serverkarte einen oder mehrere Computerprozessoren und einen Speicher umfasst. Die Vielzahl der Switch-Module ist auch funktionsmäßig mit der Mittelplatine verbunden, wobei jedes Switch-Modul dafür konfiguriert ist, den Netzverkehr für mindestens eine der Serverkarten zu schalten. Der Verwaltungscontroller ist dafür konfiguriert, die Vielzahl von Switch-Modulen über die Fabric-Verbindung zu verwalten.

[0004] Eine andere Ausführungsform der Erfindung stellt einen Verwaltungscontroller bereit, der einen Computerprozessor und eine Firmware zur Speicherverwaltung umfasst, die bei Ausführung im Computerprozessor eine Operation ausführt. Die Operation umfasst die Verwaltung einer Vielzahl von Switch-Modulen, die funktionsmäßig mit der Mittelplatine eines Serversystems verbunden sind. Die Mittelplatine

umfasst eine Fabric-Verbindung für eine Switch-Fabric. Die Vielzahl der Switch-Module wird über die Fabric-Verbindung verwaltet. Die Mittelplatine ist funktionsmäßig mit einer oder mehreren Serverkarten verbunden. Jedes Switch-Modul ist dafür konfiguriert, den Netzverkehr für mindestens eine der Serverkarten zu verwalten.

[0005] Eine weitere Ausführungsform der Erfindung stellt ein computergestütztes Verfahren bereit, das die Bereitstellung eines Serversystems einschließlich einer Mittelplatine, einer oder mehrerer Serverkarten und einer Vielzahl von Switch-Modulen umfasst. Die Mittelplatine umfasst eine Fabric-Verbindung für eine Switch-Fabric. Alle Serverkarten und alle Switch-Module sind jeweils funktionsmäßig mit der Mittelplatine verbunden. Jedes Switch-Modul ist dafür konfiguriert, den Netzverkehr für mindestens eine der einen oder mehreren Serverkarten zu schalten. Die Operation umfasst auch die Verwaltung der Vielzahl von Switch-Modulen über die Fabric-Verbindung.

Figurenliste

[0006] Eine ausführlichere Beschreibung von Ausführungsformen der oben kurz zusammengefassten Erfindung unter Bezugnahme auf die angefügten Zeichnungen zeigt die Art und Weise, in der die oben genannten Aspekte erreicht werden und im Einzelnen verstanden werden können.

[0007] Es muss jedoch beachtet werden, dass die angefügten Zeichnungen nur typische Ausführungsformen dieser Erfindung darstellen und daher nicht als Beschränkung für den Schutzbereich der Erfindung betrachtet werden dürfen, da die Erfindung weitere gleichermaßen wirksame Ausführungsformen zulassen kann.

Fig. 1 ist ein Blockschaltbild einer Datenverarbeitungsumgebung, die gemäß einer Ausführungsform der Erfindung mehrere Hosts mit Zugang zu einem Serversystem aufweist.

Fig. 2 zeigt eine Konfiguration, in der Zwischenstecker-Karten gemäß einer Ausführungsform der Erfindung in einem Serversystem funktionsmäßig mit Serverkarten verbunden sind.

Fig. 3 zeigt eine Konfiguration, in der eine Zwischenstecker-Karte gemäß einer Ausführungsform der Erfindung in einem Serversystem funktionsmäßig mit zwei Serverkarten verbunden ist.

Fig. 4 zeigt ein Serversystem, das gemäß einer Ausführungsform der Erfindung dafür konfiguriert ist, die Auswirkungen einer Reparatur an einem Switch-Modul zu verringern.

Fig. 5 zeigt auch ein Serversystem, das gemäß einer Ausführungsform der Erfindung dafür konfiguriert ist, die Auswirkungen einer Reparatur an einem Switch-Modul zu verringern.

Fig. 6 zeigt eine Switch-Fabric für ein Serversystem gemäß einer Ausführungsform der Erfindung.

Fig. 7 zeigt ein Serversystem, das gemäß einer Ausführungsform der Erfindung eine Mittelplatine aufweist, die mit einer Vielzahl von Zwischenstecker-Karten verbunden ist.

Fig. 8 zeigt ein Serversystem, das gemäß einer Ausführungsform der Erfindung mehrere Gestellrahmen (frames) umfasst.

Fig. 9 zeigt ein Serversystem, das gemäß einer Ausführungsform der Erfindung mehrere Gestellrahmen umfasst, wobei jeder Gestellrahmen vier Gehäuse enthält.

Fig. 10 zeigt ein Serversystem, das gemäß einer Ausführungsform der Erfindung so ausgestaltet ist, dass es eine Zwischenstecker-Verbindung umfasst.

Fig. 11 zeigt eine Konfiguration eines Serversystems, das gemäß einer Ausführungsform der Erfindung vermeidet, dass ein Switch-Modul ein Single Point of Failure (SPOF) in einem Paar informationstechnischer Speichereinheiten (Speicher-ITEs) ist.

Fig. 12 zeigt eine Konfiguration eines Paares miteinander verbundener Zwischenstecker-Karten gemäß einer Ausführungsform der Erfindung.

Fig. 13 zeigt eine Konfiguration eines Serversystems, das gemäß einer Ausführungsform der Erfindung mehrere Schaltkarten umfasst.

Fig. 14 zeigt eine logische Ansicht einer Konfiguration eines Serversystems, das gemäß einer Ausführungsform der Erfindung mehrere Schaltkarten umfasst.

Fig. 15 zeigt eine Konfiguration eines Serversystems, das gemäß einer Ausführungsform der Erfindung mehrere Schaltkarten umfasst.

Fig. 16 zeigt ein Serversystem, das gemäß einer Ausführungsform der Erfindung einen Verwaltungscontroller und eine unabhängige Verwaltungsverbindung umfasst.

Fig. 17 zeigt ein Serversystem, das gemäß einer Ausführungsform der Erfindung zwar einen Verwaltungscontroller, aber keine unabhängige Verwaltungsverbindung umfasst.

Fig. 18 ist ein Flussdiagramm, das gemäß einer Ausführungsform der Erfindung ein Verfahren zur Verringerung der Verdrahtung in einem Serversystem darstellt, welches einen Verwaltungscontroller umfasst.

Fig. 19 ist ein Flussdiagramm, das gemäß einer Ausführungsform der Erfindung ein Verfahren zum Verwalten einer Switch-Fabric darstellt.

[0008] Einige Ausführungsformen der Erfindung verringern die Verdrahtung in einem Serversystem, das einen Verwaltungscontroller umfasst, der dazu konfiguriert ist, eine Switch-Fabric für das Serversystem zu verwalten. Eine Switch-Fabric bezieht sich im Sinne der vorliegenden Beschreibung auf eine Netztopologie, in der die Netzknoten über einen oder mehrere Netz-Switches miteinander verbunden sind. Bei einer Ausführungsform umfasst das Serversystem eine Mittelplatine, die ihrerseits eine Fabric-Verbindung für die Switch-Fabric umfasst. Das Serversystem kann auch eine oder mehrere Serverkarten umfassen, die funktionsmäßig mit der Mittelplatine verbunden sind. Jede Serverkarte umfasst einen oder mehrere Computerprozessoren und einen Speicher. Bei einer Ausführungsform kann jede Serverkarte ein Server-Blade sein, der auch als Blade-Server oder Blade bezeichnet wird. Das Serversystem kann auch Switch-Module umfassen, die funktionsmäßig mit der Mittelplatine verbunden sind, wobei jedes Switch-Modul den Netzverkehr für mindestens eine Serverkarte schaltet. Der Verwaltungscontroller kann dazu konfiguriert werden, die Switch-Module über die Fabric-Verbindung, nicht aber über eine andere Verbindung zwischen den Switch-Modulen zu verwalten. Die Verdrahtung im Serversystem wird verringert, da keine andere Verbindung erforderlich ist, um die Switch-Module zu verwalten. Auf diese Weise können die Kosten und/oder der Aufwand der Ausstattung (packaging) des Serversystems verringert werden.

[0009] Außerdem können Ausführungsformen der Erfindung die Auswirkung eines Switch-Ausfalls in einer Switch-Fabric verringern. Ein Serversystem kann zum Beispiel eine erste Zwischenstecker-Karte und ein erstes Switch-Modul umfassen, die zwischen der Mittelplatine und der einen oder den mehreren Serverkarten angeordnet sind, und die Mittelplatine funktionsmäßig mit einer oder mehreren Serverkarten verbinden. Obwohl die Mittelplatine in Bezug auf die erste Zwischenstecker-Karte beschrieben wird, kann sie so konfiguriert sein, dass sie mit einer Vielzahl von Zwischenstecker-Karten verbunden ist. Die erste Zwischenstecker-Karte kann von der Mittelplatine aus während des Betriebs ausgetauscht werden, und die eine oder die mehreren Serverkarten können von der ersten Zwischenstecker-Karte aus während des Betriebs ausgetauscht werden. Wenn bei der Ausstattung des Serversystems die hier beschriebenen Techniken verwendet werden, kann die Anzahl der Single Points of Failure (SPOFs) und/oder Single Points of Repair (SPORs) verringert oder auf ein Mindestmaß herabgesetzt werden. Die SPOFs gelten als beseitigt, wenn das Serversystem trotz des Ausfalls einer beliebigen Komponente weiterarbeiten kann. Die SPORs gelten als vermieden, wenn das Serversystem bei der Reparatur oder dem Austausch

einer beliebigen (ausgefallenen) Komponente weiterarbeiten kann.

[0010] Im Folgenden wird auf die Ausführungsformen der Erfindung Bezug genommen. Dennoch versteht sich von selbst, dass die Erfindung nicht auf die hier beschriebenen spezifischen Ausführungsformen beschränkt ist. Vielmehr werden beliebige Kombinationen der folgenden Merkmale und Elemente, egal ob zu verschiedenen Ausführungsformen gehörig oder nicht, als Umsetzung und Anwendung der Erfindung betrachtet. Obwohl mit den Ausführungsformen der Erfindung Vorteile gegenüber anderen möglichen Lösungen und/oder dem Stand der Technik erreicht werden können, ist außerdem zu beachten, dass die Tatsache, ob durch eine bestimmte Ausführungsform ein spezieller Vorteil erreicht wird oder nicht, keine Beschränkung für die Erfindung darstellt. Daher dienen die folgenden Aspekte Merkmale, Ausführungsformen und Vorteile nur der Veranschaulichung und sind nicht als Elemente oder Einschränkungen der beigefügten Ansprüche zu verstehen, es sei denn, dies wird ausdrücklich in einem oder mehreren Ansprüchen angegeben. Gleichermaßen darf der Bezug auf „die Erfindung“ nicht als Verallgemeinerung der hier offenbarten Erfindungsgegenstände aufgefasst werden und sollte nicht als Element oder Einschränkung der beigefügten Ansprüche betrachtet werden, es sei denn, dies wird ausdrücklich in einem oder mehreren Ansprüchen angegeben.

[0011] Wie der Fachmann zu würdigen weiß, können Aspekte der vorliegenden Erfindung als System, Verfahren oder Computerprogrammprodukt ausgeführt werden. Entsprechend können Aspekte der vorliegenden Erfindung die Form einer vollständigen Hardware-Ausführungsform, einer vollständigen Software-Ausführungsform (einschließlich Firmware, residenter Software, Mikrocode usw.) oder einer Ausführungsform annehmen, die Software und Hardware-Aspekte kombiniert, die hier allgemein als „Schaltkreis“, „Modul“ oder „System“ bezeichnet werden können. Außerdem können Aspekte der vorliegenden Erfindung die Form eines Computerprogrammprodukts annehmen, das in einem oder mehreren computerlesbaren Medien enthalten ist, auf denen ein computerlesbarer Programmcode enthalten ist.

[0012] Jede beliebige Kombination eines oder mehrerer computerlesbarer Medien kann verwendet werden. Das computerlesbare Medium kann ein computerlesbares Signalmedium oder ein computerlesbares Speichermedium sein. Ein computerlesbares Speichermedium kann unter anderem beispielsweise ein System, eine Vorrichtung oder eine Einheit zur elektronischen, magnetischen, optischen, elektromagnetischen, Infrarot- oder Halbleiterspeicherung sein oder jede geeignete Kombination davon. Konkrete Beispiele für computerlesbare Speichermedien wür-

den (in einer unvollständigen Liste) Folgendes umfassen: eine elektrische Ein- oder Zweidrahtverbindung, eine tragbare Computerdiskette, eine Festplatte, einen RAM (Random Access Memory), einen ROM (Read-Only Memory), einen EPROM (Erasable Programmable Read-Only Memory) oder Flash-Speicher, einen Lichtwellenleiter, eine CD-ROM (Compact Disc Read-Only Memory), eine optische Speichereinheit, eine magnetische Speichereinheit oder jede geeignete Kombination davon. Im Zusammenhang mit dieser Offenbarung kann ein computerlesbares Speichermedium jedes verfügbare Medium sein, auf dem ein Programm zur Verwendung durch oder in Verbindung mit einem System, einer Vorrichtung oder einer Einheit zur Befehlsausführung enthalten sein oder gespeichert werden kann.

[0013] Ein computerlesbares Signalmedium kann ein übertragenes Datensignal umfassen, das einen computerlesbaren Programmcode enthält, wie beispielsweise im Basisband oder als Teil einer Trägerwelle. Ein solches übertragenes Signal kann eine von einer Vielzahl von Formen annehmen, unter anderem eine elektromagnetische oder optische Form oder jede geeignete Kombination davon. Ein computerlesbares Signalmedium kann jedes computerlesbare Medium sein, das kein computerlesbares Speichermedium ist und welches ein Programm für eine Verwendung durch oder in Verbindung mit einem System, einer Vorrichtung oder einer Einheit zur Befehlsausführung austauschen, übertragen oder transportieren kann.

[0014] Programmcode, der auf einem computerlesbaren Medium enthalten ist, kann unter Verwendung eines geeigneten Mediums übertragen werden, unter anderem über Funk- oder Festnetz, Lichtwellenleiter, HF usw. oder jede geeignete Kombination davon.

[0015] Computerprogrammcode zum Durchführen von Operationen zu Aspekten der vorliegenden Erfindung kann in jeder beliebigen Kombination aus einer oder mehreren Programmiersprachen geschrieben werden, einschließlich objektorientierter Programmiersprachen, wie zum Beispiel Java, Smalltalk, C++ oder ähnlicher Programmiersprachen, und herkömmlicher prozeduraler Programmiersprachen, wie zum Beispiel der „C“-Programmiersprache oder ähnlicher Programmiersprachen. Der Programmcode kann entweder vollständig auf dem Computer des Benutzers, teilweise auf dem Computer des Benutzers, als unabhängiges Softwarepaket, teilweise auf dem Computer des Benutzers und teilweise auf einem fernen Computer oder vollständig auf einem fernen Computer oder Server ausgeführt werden. Im letztgenannten Szenario kann der ferne Computer mit dem Computer des Benutzers entweder über ein beliebiges Netz verbunden werden, einschließlich eines lokalen Netzes (LAN - Local Area Network) oder eines Fernnetzes (WAN - Wide Area Network), oder

die Verbindung kann zu einem externen Computer (zum Beispiel, über das Internet unter Verwendung eines Internet-Diensteanbieters) hergestellt werden.

[0016] Aspekte der vorliegenden Erfindung werden nachfolgend unter Bezugnahme auf Flussdiagrammdarstellungen und/oder Blockschaltbilder zu Verfahren, Vorrichtungen (Systemen) und Computerprogrammprodukten gemäß Ausführungsformen der Erfindung beschrieben. Selbstverständlich kann jeder Block der Flussdiagrammdarstellungen und/oder Blockschaltbilder und Kombinationen von Blöcken in den Flussdiagrammdarstellungen und/oder Blockschaltbildern durch Computerprogrammbefehle umgesetzt werden. Diese Computerprogrammbefehle können einem Prozessor eines Universalcomputers, Spezialcomputers oder einer anderen programmierbaren Datenverarbeitungsvorrichtung bereitgestellt werden, um eine Maschine zu erzeugen, in der die vom Prozessor des Computers oder der anderen programmierbaren Datenverarbeitungsvorrichtung ausgeführten Befehle Mittel zur Umsetzung der in dem Block oder den Blöcken des Flussdiagramms und/oder Blockschaltbilds spezifizierten Funktionen/Aktionen erzeugen.

[0017] Diese Computerprogrammbefehle können auch in einem computerlesbaren Medium gespeichert werden, das einen Computer, eine andere programmierbare Datenverarbeitungsvorrichtung oder andere Einheiten beauftragen kann, in einer bestimmten Weise zu funktionieren, sodass die im computerlesbaren Medium gespeicherten Befehle einen Herstellungsartikel erzeugen, der Anweisungen enthält, die die in dem Block oder den Blöcken des Flussdiagramms und/oder Blockschaltbilds spezifizierten Funktionen/Aktionen umsetzen.

[0018] Die Computerprogrammbefehle können auch in einen Computer, eine andere programmierbare Datenverarbeitungsvorrichtung oder andere Einheiten geladen werden, um eine Reihe von Betriebschritten zu erzeugen, die auf dem Computer, der anderen programmierbaren Vorrichtung oder den anderen Einheiten ausgeführt werden sollen, um einen computergestützten Prozess zu erzeugen, durch den die Befehle, die im Computer oder in der anderen programmierbaren Vorrichtung ausgeführt werden, Prozesse für das Umsetzen der in dem Block oder den Blöcken des Flussdiagramms und/oder Blockschaltbilds spezifizierten Funktionen/Aktionen bereitstellen.

[0019] Die Flussdiagramme und Blockschaltbilder in den Figuren veranschaulichen die Architektur, Leistungsmerkmale und Betriebsweise möglicher Umsetzungen von Systemen, Verfahren und Computerprogrammprodukten gemäß verschiedenen Ausführungsformen der vorliegenden Erfindung. In dieser Hinsicht kann jeder Block der Flussdiagramme oder

Blockschaltbilder ein Modul, ein Segment oder einen Teil von Code darstellen, der ein oder mehrere ausführbare Befehle zur Umsetzung der festgelegten, logischen Funktion(en) umfasst. Es sollte auch beachtet werden, dass bei einigen alternativen Umsetzungen die im Block angegebenen Funktionen in einer anderen als der in der Figur angegebenen Reihenfolge vorkommen. Zum Beispiel können zwei Blöcke, die nacheinander dargestellt sind, im Wesentlichen gleichzeitig ausgeführt werden, oder manchmal können diese Blöcke abhängig von den jeweiligen Leistungsmerkmalen auch in umgekehrter Reihenfolge ausgeführt werden. Es muss auch erwähnt werden, dass jeder Block der Blockschaltbilder und/oder der Flussdiagrammdarstellung und Kombinationen von Blöcken in den Blockschaltbildern und/oder der Flussdiagrammdarstellung durch spezielle hardwaregestützte Systeme umgesetzt werden können, welche die festgelegten Funktionen oder Aktionen oder Kombinationen von Spezialhardware- und Computeranweisungen ausführen.

[0020] Fig. 1 ist ein Blockdiagramm einer Datenverarbeitungsumgebung 100, die gemäß einer Ausführungsform der Erfindung mehrere Hosts mit Zugang zu einem Serversystem 102 aufweist. Obwohl aus Übersichtsgründen nur drei Hosts 134a,b,n dargestellt sind, wird der Fachmann verstehen, dass zusätzliche Hosts Zugriff auf das Serversystem 102 haben können. Die Hosts 134a,b,n sind über eine Netzfabric 132 mit dem Serversystem 102 verbunden. Je nach Ausführungsform kann jeder Host 134a,b,n als Client handeln, der Zugang zu den vom Serversystem 102 gebotenen Leistungsmerkmalen hat und/oder entsprechende Serverleistungsmerkmale außerhalb des Serversystems 102 bereitstellen kann. Die Netzfabric 132 kann ein Telekommunikationsnetz und/oder ein Fernnetz (WAN) sein. Bei einer besonderen Ausführungsform bildet das Internet die Netzfabric 132. Das Serversystem 102 umfasst ein Gehäuse, in dem die Server-Blades 104a,b,n untergebracht sind. Die Server-Blades 104a,b,n sind mit einer Mittelplatine 123 verbunden, die mechanische und logische Verbindungen (z.B., Austausch von Daten und Steuersignalen) zwischen den Server-Blades 104a,b,n bereitstellt. Obwohl drei Server-Blades 104a,b,n dargestellt sind, wird der Fachmann verstehen, dass zusätzliche Server-Blades mit der Mittelplatine 123 verbunden werden können. Obwohl die Ausführungsformen hier in Bezug auf Blade-Systeme beschrieben werden, sind außerdem auch andere Formfaktoren oder physische Konfigurationen (z.B. Rack-Systeme) im weitesten Sinne vorgesehen.

[0021] Obwohl die Ausführungsformen hier in Bezug auf die Server-Blades 104a,b,n, beschrieben werden, die mit der Mittelplatine 123 verbunden sind, wird der Fachmann außerdem erkennen, dass die Server-Blades generell auch mit beliebigen Leiterplatten (PCB - Printed Circuit Board) verbunden werden kön-

nen, die als Gehäuse-Backbone dienen, wie zum Beispiel Rückwandplatine, Steuerplatine, usw. Obwohl die Ausführungsformen hier in Bezug auf das Serversystem **102** beschrieben werden, das ein einziges Gehäuse hat, wird der Fachmann außerdem erkennen, dass das Serversystem **102** bei anderen Ausführungsformen auch mehrere Gehäuse umfassen kann. Zum Beispiel kann das Serversystem **102** bei einer alternativen Ausführungsform ein Blade-System sein, das mindestens zwei Blade-Gehäuse umfasst, die jeweils eine Vielzahl von Blades aufweisen.

[0022] Bei einer Ausführungsform umfasst das Serversystem **102** ein oder mehrere Verwaltungsmodule **124**. Bei der dargestellten Ausführungsform umfasst das Serversystem **102** ein primäres Verwaltungsmodul **124a** und ein Sicherungs-Verwaltungsmodul **124b**. Jedes Verwaltungsmodul **124** ist in der Lage, mehrere Server-Blades **104** zu verwalten. Während des normalen Betriebs ist eines der Verwaltungsmodule **124** funktionsmäßig über ein lokales Netz (LAN) **122**, die Mittelplatine **123** und einen Baseboard Management Controller (BMCs) **110** in jedem Server-Blade **104** mit den Server-Blades **104** verbunden, um einen Inband-Verwaltungspfad zu bilden. Bei einer Ausführungsform dient die Netz-Fabric **132** als Erweiterung für das LAN **122**. Das LAN **122** und der BMC **110** werden weiter unten ausführlicher beschrieben.

[0023] Bei einer Ausführungsform ist die Mittelplatine **123** mitten im Gehäuse des Serversystems **102** angebracht und enthält Schaltkreise und Steckplätze **112**, in die zusätzliche elektronische Einheiten oder Karten, einschließlich Server-Blades **104**, eingefügt werden können. Die Mittelplatine **123** umfasst mindestens einen Bus für einen sicheren internen Inband-Datenaustausch über die BMCs **110** sowie zwischen den Verwaltungsmodulen **124** und den Server-Blades **104** und/oder zwischen den Server-Blades **104** selbst.

[0024] Bei einer Ausführungsform wird beim Einfügen eines Server-Blade **104** in einen spezifischen Steckplatz **112** eine physische Adresse für den Server-Blade **104** festgelegt. Es sei angenommen, der Server-Blade **104a** wird zum Beispiel in den Steckplatz **112a** eingefügt. Bei einer Ausführungsform erkennt die Steuerlogik **116a** die Anwesenheit des Server-Blades **104a** im Steckplatz **112a**. Die Steuerlogik **116a** kann sich gemäß dem RS485-Standard der Electronics Industry Association (EIA) für die Datenübertragung verhalten. Bei anderen Ausführungsformen kann die Steuerlogik **116a** kompatibel sein mit dem Inter-IC-Standard von Phillips (Inter-Integrated Circuit oder I²C) oder mit einem Ethernet-Netzstandard. Als Antwort auf das Einfügen des Server-Blades **104a** in den Steckplatz **112a** weist die Steuerlogik **116a**, die in Verbindung mit dem Verwaltungsmodul **124a** arbeitet, dem Server-Blade **104a** auf einem Bus in der Mittelplatine **123** eine physische Adresse

zu. Wie gezeigt, ist jedem Server-Blade **104** eine entsprechende Steuerlogik **116** zugeordnet, die funktionsmäßig mit der Mittelplatine **123** verbunden ist. Bei einer alternativen Ausführungsform können mehrere Server-Blades **104** gemeinsam eine einzige Steuerlogik **116** nutzen.

[0025] Bei einer Ausführungsform wird jedem Server-Blade **104** auf der Mittelplatine **123** eine eindeutige Internet-Protokoll- (IP-)Adresse zugeteilt. Dies geschieht, da die Mittelplatine **123** einen Datenaustausch unter Verwendung des IP-Adressprotokolls unterstützen kann, bei welchem jeder Einheit, die funktionsmäßig mit der Mittelplatine **123** verbunden ist, von einer (nicht dargestellten) Logik, die sich innerhalb oder außerhalb des Gehäuses des Serversystems **102** befindet, eine IP-Adresse zugewiesen wird. Zum Beispiel kann ein DHCP-Server (Dynamic Host Configuration Protocol) verwendet werden, um dem Server-Blade **104a** eine IP-Adresse zuzuweisen. Die Datenübertragung mit dem Server-Blade **104a** erfolgt danach über einen dem Server-Blade **104a** zugeordneten Netzschneittstellencontroller (NIC - Network Interface Controller) **114a**. Der NIC **114a** kann eine beliebige Netzkommunikationseinheit sein, die es dem Server-Blade **104a** erlaubt, über das LAN **122** und/oder die Netz-Fabric **132** einen Datenaustausch mit anderen Server-Blades **104b,n** und/oder Computern auszuführen.

[0026] Bei einer Ausführungsform ist ein integriertes Modul **126a** funktionsmäßig mit dem NIC **114a** verbunden. Die integrierten Module **126a** können paarweise (z.B. mit dem integrierten Modul **126b**) verwendet werden, um eine Redundanz bereitzustellen. Bekanntlich bezieht sich Small Computer System Interface (SCSI) auf eine Reihe von Standards für die physische Verbindung und die Datenübertragung zwischen Computern und Peripheriegeräten. Bei einer Ausführungsform umfassen die integrierten Module **126** Switch-Module **128**, wie zum Beispiel ein SAS-Switch-Modul (Serial-Attached SCSI). Die Switch-Module **128** stellen für die Server-Blades **104** eine Konnektivität mit Ethernet, Fibre Channel over Ethernet (FCoE), SAS usw. bereit. Bei einer Ausführungsform ist jedes Switch-Modul **128** ein Switch-Chip. Je nach Ausführungsform können die integrierten Module **126** außerdem redundante RAID-Controller **130** (RAID-Redundant Array of Independent Discs, Controller für Arrays unabhängiger Platten) umfassen. Jeder RAID-Controller **130** ist mit RAID-Einheiten verbunden, wie zum Beispiel mit Speichereinheiten in einer RAID-Konfiguration. Die RAID-Einheiten können in einem oder mehreren Server-Blades **104** untergebracht sein. Die RAID-Controller **130** und die RAID-Einheiten können gemeinsam als RAID-Teilsystem des Serversystems **102** betrachtet werden.

[0027] Bei einer Ausführungsform kann jede Speichereinheit eine persistente Speichereinheit sein. Außerdem kann jede Speichereinheit eine Kombination aus festen und/oder austauschbaren Speichereinheiten sein, wie zum Beispiel Festplattenlaufwerken, Diskettenlaufwerken, Bandlaufwerken, austauschbaren Speicherarten, Solid-State-Laufwerken oder optischen Speichern. Der Speicher **108** und die Speichereinheit können Teil eines virtuellen Adressraums sein, der mehrere primäre und sekundäre Speichereinheiten übergreift.

[0028] Bei einer Ausführungsform kann jeder Server-Blade **104** mindestens eine zentrale Verarbeitungseinheit (CPU) **106** und einen Speicher **108** aufweisen. Die CPU **106** steht stellvertretend für eine einzelne CPU, mehrere CPUs, eine einzelne CPU mit mehreren Prozessorkernen und Ähnliches. In ähnlicher Weise kann der Speicher **108** ein RAM sein. Obwohl der Speicher **108** als einzelne Einheit dargestellt ist, muss dies so verstanden werden, dass der Speicher **108** eine Vielzahl von Modulen umfassen kann und dass der Speicher **108** auf mehreren Ebenen vorhanden sein kann, von Hochgeschwindigkeitsregistern und Cache-Speichern bis hin zu größeren DRAM-Chips mit geringerer Geschwindigkeit. Der Speicher **108** kann ein Flash-Nur-Lese-Speicher („Flash-ROM“ oder „Flashspeicher“) sein, der gelöscht und in Speichereinheiten, die als „Blocks“ bezeichnet werden, neu programmiert werden kann. Der Speicher **108** kann auch nichtflüchtigen, elektrisch löschbaren programmierbaren Nur-Lese-Speicher (EEPROM) aufweisen, der ähnlich wie ein Flashspeicher ist, nur dass der EEPROM auf Byteebene gelöscht und neu beschrieben wird und gewöhnlich eine geringere Kapazität aufweist. Jeder Server-Blade **104** kann als Prozessor-Blade oder als Speicher-Blade eingerichtet werden. Ein Prozessor-Blade umfasst eine oder mehrere Verarbeitungseinheiten, während ein Speicher-Blade mehrere integrierter Speichereinheiten wie zum Beispiel Plattenlaufwerke umfasst.

[0029] Wenn der Server-Blade **104** direkt vom Hersteller versandt wurde, kann der Speicher **108** bei einer Ausführungsform mit Firmware vorgebrannt sein, einschließlich eines Basic-Input/Output-Systems (BIOS) und Software zum Überwachen des Server-Blades **104**. Die Überwachung kann die Steuerung der Speichereinheiten, die Überwachung und Steuerung der Spannungen im ganzen System, die Ermittlung des Einschaltstatus des Server-Blades **104**, die Anforderung von Zugriff auf eine gemeinsam genutzte Tastatur, Video, Maus, CD-ROM und/oder Diskettenlaufwerke, die Überwachung des auf dem Server-Blade **104** laufenden Betriebssystems (OS - Operating System) usw. umfassen. Beispiele der Betriebssysteme umfassen UNIX, Versionen des Betriebssystems Microsoft Windows® und Varianten des Betriebssystems Linux®. Generell kann jedes Betriebs-

system verwendet werden, das die hier offenbarten Funktionen unterstützt.

[0030] Bei einer Ausführungsform sind die Verwaltungsmodule **124** dazu fähig, die Anwesenheit, Anzahl, Art und Revisionsstufe aller Server-Blades **104**, Stromversorgungsmodule **118** und Mittelplatten **123** im System zu erkennen. Die Verwaltungsmodule **124** können auch direkt den Betrieb aller Server-Blades **104** und Stromversorgungsmodule **118** steuern. Die Verwaltungsmodule **124** können auch direkt den Betrieb von Lüftern **120** und anderen Komponenten im Gehäuse des Serversystems **102** steuern. Die direkte Betriebssteuerung hat zur Folge, dass der Betrieb gesteuert wird, ohne das BIOS in den Server-Blades **104** zu verwenden. Bei einer alternativen Ausführungsform können die Verwaltungsmodule **124** das BIOS zum indirekten Steuern des Betriebs der Lüfter **120** und anderer Komponenten im Gehäuse des Serversystems **102**.

[0031] Bei einer Ausführungsform umfasst jeder Server-Blade **104** einen Baseboard Management Controller (BMC) **110**, der die lokale Überwachungssteuerung des dem BMC **110** zugeordneten Server-Blades **104** bereitstellt. Jeder BMC **110** ist konfiguriert für den Datenaustausch mit einem Verwaltungsmodul **124** entweder über den Datenaustauschpfad des LAN **122** (d.h. über ein Inband-Netz) oder alternativ durch die Verwendung von Switch-Modulen **128** und NICs **114** (d.h. über ein Out-of-band-Netz). Die Verwaltungsmodule **124** können zum Austauschen von Daten mit jedem Server-Blade **104** eine Vielfalt an Datenaustauschpfaden im LAN **122** nutzen, wie zum Beispiel einen RS485-Pfad, einen LAN-Pfad und einen I²C-Pfad.

[0032] Bei einer Ausführungsform ist das LAN **240** ein Inband-Netz, das sich gemäß dem RS485-Standard der Electronics Industry Association (EIA) für die Datenübertragung verhält. Die Verwaltungsmodule **124** (z.B. entweder das primäre Verwaltungsmodul **124a** oder das Sicherungs-Verwaltungsmodul **124b**, wenn das primäre Verwaltungsmodul **124a** außer Betrieb ist) nutzen das LAN **122** zum Datenaustausch mit dem BMC **110**, der eine Logik zum Koordinieren des Datenaustauschs mit den Server-Blades **104** über die Steckplätze **112** umfasst.

[0033] Bei einer Ausführungsform kann das LAN **122** dafür konfiguriert sein, eine Datenübertragung zwischen den Server-Blades **104** und den Verwaltungsmodulen **124** in Bezug auf die fernen BIOS-Einstellungen und die BIOS-Verwaltung zu erlauben. Die Server-Blades **104** können die BMCs **110** als Proxys für den Datenaustausch mit den Verwaltungsmodulen **124** durch das RS485-Protokoll verwenden. In ähnlicher Weise können die Verwaltungsmodule die BMCs **110** als Proxys für den Datenaustausch mit den Server-Blades **104** durch das RS485-Protokoll

verwenden. Bei einer alternativen Ausführungsform kann separat eine RS485-Verbindung zwischen jedem Server-Blade 104 und den Verwaltungsmodulen **124** hergestellt werden. Zusätzlich können auch andere Datenübertragungsprotokolle und -pfade über die Switch-Module **128** genutzt werden, wie zum Beispiel I²C, TCP/IP, Ethernet, FCoE usw.

[0034] Je nach Ausführungsform kann das Serversystem **102** funktionsmäßig auch mit einer Eingabeeinheit und/oder einer Ausgabeeinheit verbunden sein. Die Eingabeeinheit kann jede beliebige Einheit sein, die Eingaben für das Serversystem **102** bereitstellt. Zum Beispiel können eine Tastatur, ein Tastenfeld, ein Lichtstift, ein Touchscreen, eine Rollkugel oder eine Spracherkennungseinheit, ein Audio/Video-Player und Ähnliches verwendet werden. Die Ausgabeeinheit kann eine beliebige Einheit sein, die Ausgaben des Serversystems **102** für einen Benutzer bereitstellt. Die Ausgabeeinheit kann zum Beispiel ein herkömmlicher Bildschirm oder eine Lautsprechergruppe sein sowie ihre entsprechenden Schnittstellenkarten, d.h. Videokarten und Soundkarten. Außerdem können die Eingabeeinheit und die Ausgabeeinheit kombiniert werden. Zum Beispiel kann ein Bildschirm mit einem integrierten Touchscreen, ein Bildschirm mit einer integrierten Tastatur oder eine Spracherkennungseinheit mit einem Text-Sprachwandler kombiniert werden.

[0035] Fig. 2 zeigt eine Konfiguration **200**, in der Zwischenstecker-Karten gemäß einer Ausführungsform der Erfindung in einem Serversystem funktionsmäßig mit Serverkarten verbunden sind. Je nach Ausführungsform kann die Zwischenstecker-Karte auch mit den Serverkarten und/oder der Mittelplatine verbunden sein. Bei einer Ausführungsform wird das Serversystem zur Verringerung der Auswirkungen einer Reparatur an einem ausgefallenen Switch-Modul so konfiguriert, dass es eine Mittelplatine **123** und Serverkarten **202** umfasst, wobei die Mittelplatine und die Serverkarten **202** funktionsmäßig über eine oder mehrere Zwischenstecker-Karten **204** miteinander verbunden sind. Jede Serverkarte **202** kann einem Server-Blade 104 entsprechen. Jede Zwischenstecker-Karte **204** kann während des Betriebs ausgetauscht werden und umfasst ein oder mehrere Switch-Module **206**. Die Switch-Module **206** schalten den Netzverkehr für eine oder mehrere Serverkarten **202**, die funktionsmäßig mit der jeweiligen Zwischenstecker-Karte verbunden sind.

[0036] Bei einer Ausführungsform ist das Serversystem so konfiguriert, dass es den Ausfall des Switch-Moduls **206** erkennt. Nachdem der Ausfall des Switch-Moduls **206** erkannt wurde, kann das Serversystem am Bildschirm anzeigen, dass eine Reparatur am Switch-Modul **206** ausgeführt werden muss. Zum Beispiel kann die Anzeige eine Ausgabe in einem Fenster einer grafischen Benutzeroberfläche

(GUI) sein oder eine per E-Mail an einen Benutzer gesendete Benachrichtigung. Je nach Ausführungsform kann das Erkennen und/oder die Ausgabe durch eine beliebige Komponente des Serversystems erfolgen, wie zum Beispiel die Serverkarten **202**, das Switch-Modul **206** und/oder die im Serversystem enthaltene Firmware. Bei einer Ausführungsform umfasst das Serversystem beispielsweise Verwaltungs-Firmware, die den Zustand des Serversystems überwacht und einen Ausfall des Switch-Moduls **206** erkennt.

[0037] Dementsprechend kann beim Ausfall eines Switch-Moduls **206** die Zwischenstecker-Karte **204**, die das Switch-Modul **206** enthält, durch eine Zwischenstecker-Karte ausgetauscht werden, die ein funktionsfähiges Switch-Modul aufweist. Aufgrund der Fähigkeit, Zwischenstecker-Karten, Serverkarten und/oder die Mittelplatine während des Betriebs auszutauschen, kann die Zwischenstecker-Karte **204** ausgetauscht werden, ohne dass ein Abschalten oder ein Warmstart des Serversystems und/oder der Switch-Fabric erforderlich ist. Die Zwischenstecker-Karte, die ein funktionsfähiges Switch-Modul aufweist, kann dann über das Konfigurationstool wieder in die Switch-Fabric integriert werden. Je nach Ausführungsform kann das Konfigurationstool auf dem Serversystem oder auf einem anderen über die Netz-Fabric **132** mit dem Serversystem verbundenen Computer ausgeführt werden.

[0038] Folglich sind während der Dauer des Austauschs der Zwischenstecker-Karte mit dem ausgefallenen Switch-Modul nur das ausgefallene Switch-Modul und die zugehörigen Serverkarten nicht von der Netz-Fabric aus erreichbar. Während der Dauer des Austauschs der Zwischenstecker-Karte mit dem ausgefallenen Switch-Modul bleiben andere Switch-Module und/oder Serverkarten, die funktionsmäßig mit der Mittelplatine verbunden sind, erreichbar. Dementsprechend ist die Auswirkung der Reparatur des ausgefallenen Switch-Moduls räumlich auf die Serverkarten beschränkt, die dem ausgefallenen Switch-Modul zugeordnet sind. Mit anderen Worten: Die einzigen Netzknoten, die während der Reparatur nicht von der Netz-Fabric erreicht werden können, sind die Netzknoten, die den Serverkarten zugeordnet sind, welche funktionsmäßig mit dem ausgefallenen Switch-Modul verbunden sind.

[0039] Fig. 3 zeigt eine Konfiguration **300**, bei der eine Zwischenstecker-Karte gemäß einer Ausführungsform der Erfindung in einem Serversystem funktionsmäßig mit zwei Serverkarten verbunden ist. Wie gezeigt umfasst die Zwischenstecker-Karte **204** das Switch-Modul **206** und zwei konvergente Netzadapter (Converged Network Adapters, CNAs) **302**. Jede der zwei Serverkarten **202** kann auch zwei CPUs **106** und einen CNA **304** umfassen. Bei einer Ausführungsform sind die CPUs **106** mit FCoE mit

CNAs verbunden, die sowohl Fibre-Channel-Hostbusadapter (HBA) als auch Ethernet-NIC-Leistungsmerkmale aufweisen. Die CNAs können einen oder mehrere physische Ethernet-Anschlüsse umfassen und von den CPUs **106** dafür konfiguriert sein, eine maschinennahe Verarbeitung und/oder SCSI-Protokolleistungsmerkmale auszulagern, was traditionell von den Fibre-Channel-Hostbusadaptern ausgeführt wird. Wie oben beschrieben schaltet das Switch-Modul **206** den Netzverkehr für die Serverkarten **202**. Wenn das Switch-Modul **206** ausfällt, erlaubt die Konfiguration **300**, dass das Switch-Modul **206** ausgetauscht wird, ohne dass ein Warmstart der Serverkarten anderer Zwischenstecker-Karten, die mit der Mittelplatine verbunden sind, erforderlich ist.

[0040] Fig. 4 zeigt ein Serversystem **400**, das gemäß einer Ausführungsform der Erfindung dafür konfiguriert ist, die Auswirkungen einer Reparatur an einem Switch-Modul zu verringern. Wie gezeigt, umfasst das Serversystem **400** einen logischen Server **402**, der über eine informationstechnische Prozessoreinheit (Prozessor-ITE) **404** und eine informationstechnische E/A-Einheit (E/A-ITE) **406** konfiguriert wird. ITE bezieht sich hier generell auf eine beliebige Einheit, die so konfiguriert ist, dass sie funktionsmäßig mit der Mittelplatine **123** verbunden ist. Bei einer alternativen Ausführungsform kann der logische Server **402** auch über eine Speicher-ITE **408** konfiguriert werden. Die E/A-ITE **406** und die Speicher-ITE **408** sind so konfiguriert, dass sie eine oder mehrere Prozessor-ITEs mit zusätzlicher E/A-Kapazität bzw. Speicher-Kapazität versorgen. Je nach Ausführungsform kann jede ITE **404**, **406**, **408** als Teil eines oder mehrerer Server-Blades integriert oder als eigenständige Karte mit der Mittelplatine **123** verbunden werden. Die Prozessor-ITE **404** umfasst eine oder mehrere virtuelle Maschinen **410**, einen Hypervisor **412**, Speicher **414**, Prozessoren **416** und Festplattenlaufwerke **418**. Die E/A-ITE **406** umfasst eine gemeinsam genutzte E/A-ITE-Komponente **422** und E/A-Adapter **424**, während die Speicher-ITE **408** eine gemeinsam genutzte Speicher-ITE-Komponente **426** und Solid-State-Laufwerke **428** umfasst.

[0041] Bei einer Ausführungsform umfassen der Server-Blade **404** und die ITEs **406**, **408** außerdem jeweils ein Switch-Modul **206**. Jedes Switch-Modul **206** kann ein Switch-Chip und auf einer (nicht dargestellten) Zwischenstecker-Karte enthalten sein, die zwischen der Mittelplatine **123** und der Prozessor-ITE **404** und/oder der ITE **406**, **408** angeordnet ist. Gemeinsam stellen die Switch-Module **206** eine Switch-Fabric **432** bereit. Ein Ausfall des Switch-Moduls **206**, der Prozessor-ITE **404** - angezeigt durch ein X-Symbol **430** - hat nur Auswirkungen auf die Prozessor-ITE **404** und nicht auf andere ITEs, die funktionsmäßig mit der Mittelplatine **123** verbunden sind. Dementsprechend bleiben andere logische Server, die über die E/A-ITE **406** und/oder Speicher-ITE **408** konfigu-

riert wurden, funktionsfähig, und die Konnektivität in der Switch-Fabric **432** bleibt weitgehend funktionsfähig - d.h. mit Ausnahme der Konnektivität zur Prozessor-ITE **404**. Je nach Ausführungsform kann die funktionsfähig gebliebene Konnektivität in der Switch-Fabric **432** auch eine redundante Konnektivität in der Switch-Fabric **432** umfassen. Außerdem kann das Switch-Modul **206** ausgetauscht werden, ohne dass sich dies auf andere ITEs, logische Server und/oder die Switch-Fabric **432** auswirkt. Folglich kann sich die Verfügbarkeit der Switch-Fabric **432** verbessern.

[0042] Fig. 5 zeigt auch ein Serversystem **500**, das gemäß einer Ausführungsform der Erfindung dafür konfiguriert ist, die Auswirkungen einer Reparatur an einem Switch-Modul zu verringern. Wie gezeigt, umfasst das Serversystem **500** einen ersten logischen Server **402**, der über eine erste Prozessor-ITE **404** und die E/A-ITE **406** konfiguriert wird. Das Serversystem **500** umfasst auch einen zweiten logischen Server **502**, der über eine zweite Prozessor-ITE **504** und die E/A-ITE **406** konfiguriert wird. Bei einer alternativen Ausführungsform kann der erste logische Server **402** und/oder der zweite logische Server **502** auch über die Speicher-ITE **408** konfiguriert werden. Die ITEs **404**, **406**, **408**, **504** umfassen jeweils ein Switch-Modul **206**. Ein Ausfall des Switch-Moduls **206**, der Prozessor-ITE **404** (angezeigt durch ein X-Symbol **506**) wirkt sich nur auf die Prozessor-ITE **404**, nicht aber auf die zweite Prozessor-ITE **504** aus. Dementsprechend bleiben die zweite Prozessor-ITE **504**, die E/A-ITE **406** und die Speicher-ITE **408** funktionsfähig und behalten die Konnektivität zur Switch-Fabric während des Ausfalls und/oder des Austauschs des Switch-Moduls **206** bei.

[0043] Fig. 6 zeigt eine Switch-Fabric **432** für ein Serversystem gemäß einer Ausführungsform der Erfindung. Wie gezeigt, umfasst die Switch-Fabric **432** eine Vielzahl von Switch-Modulen **206**, wobei jeweils eine auf einer entsprechenden Zwischenstecker-Karte **204** untergebracht ist. Jede Zwischenstecker-Karte **204** verbindet zwei Serverkarten **202** funktionsmäßig mit der Switch-Fabric **432**. Je nach Ausführungsform können die Switch-Module in einer Switch-Fabric über eine in einer Mittelplatine angebrachte Verdrahtung, eine bezüglich der Mittelplatine externe Verkabelung oder eine Kombination davon miteinander verbunden sein. Außerdem können ein oder mehrere Switch-Module **206** außer mit den Serverkarten **202** auch mit anderen Einheiten **602** verbunden sein. Zu Beispielen für die Einheiten **602** gehören Netzeinheiten, Speichereinheiten und E/A-Einheiten. Folglich hat der Ausfall oder der Austausch eines Switch-Moduls **206** nur Auswirkungen auf die Serverkarten **202**, die mit dem Switch-Modul **206** verbunden sind, andererseits aber keine Auswirkung auf den Rest der Switch-Fabric **432** und/oder die anderen Serverkarten.

[0044] Fig. 7 zeigt ein Serversystem 700, das gemäß einer Ausführungsform der Erfindung eine Mittelplatine 123 aufweist, die mit einer Vielzahl von Zwischenstecker-Karten 204 verbunden ist. Jede Zwischenstecker-Karte 204 umfasst ein Switch-Modul 206 und verbindet ein oder zwei Serverkarten 202 funktionsmäßig mit der Mittelplatine 123. Die Mittelplatine umfasst eine Fabric-Verdrahtung, welche die Switch-Module 206 verbindet, um eine Switch-Fabric zu bilden. Das Serversystem 700 ist so konfiguriert, dass die Serverkarten 202 von den Zwischenstecker-Karten 204 aus während des Betriebs ausgetauscht werden können. Das Serversystem 700 ist außerdem so konfiguriert, dass die Zwischenstecker-Karten 204 von der Mittelplatine 123 aus während des Betriebs ausgetauscht werden können. Dementsprechend erlauben die Ausstattung und die Möglichkeit, Elemente des Serversystems 700 während des Betriebs austauschen zu können, dass ein fehlerhaftes Switch-Modul 206 ausgetauscht werden kann, während die Auswirkungen auf das Serversystem 700 und/oder die Switch-Fabric verringert oder auf ein Mindestmaß herabgesetzt werden.

[0045] Fig. 8 zeigt ein Serversystem 800, das gemäß einer Ausführungsform der Erfindung mehrere Gestellrahmen 804 umfasst. Jeder Gestellrahmen ein oder mehrere Gehäuse 802, die funktionsmäßig über Gehäuse-Verbindungskarten 806 und die zugehörige Verkabelung 808 miteinander verbunden sind. Die Gehäuse 802 können außerdem funktionsmäßig gestellübergreifend über Gestellrahmen-Verbindungskarte 810 und die zugehörige Verkabelung 812 miteinander verbunden sein. Jedes Gehäuse 802 enthält eine Mittelplatine 123, die gemäß den hier offengelegten Techniken über eine Zwischenstecker-Karte 204 mit einer oder zwei Serverkarten 202 verbunden ist. Jede Zwischenstecker-Karte 204 umfasst ein Switch-Modul 206 zum Schalten des Netzverkehrs für die Serverkarten 202. Die Mittelplatten 123 umfassen eine Fabric-Verdrahtung zur Verbindung der Switch-Module 206. Die Switch-Module 206, die Fabric-Verdrahtung, die Gehäuse-Verbindungskarten 806 und die zugehörige Verkabelung 808 sowie die Gestellrahmen-Verbindungskarte 810 und die zugehörige Verkabelung 812 bilden gemeinsam eine Switch-Fabric für das Serversystem 800. Die Switch-Fabric für das Serversystem 800 umfasst mit anderen Worten gehäuse- und gestellrahmenübergreifende Verbindungen. Folglich bedeutet der Ausfall und/oder der Austausch eines Switch-Moduls 206 (angezeigt durch ein X-Symbol 814) lediglich das Entfernen einer zugehörigen Serverkarte 202 aus der Switch-Fabric. Der Betrieb der anderen Serverkarten und/oder die Konnektivität der Switch-Fabric bleibt während des Ausfalls und/oder des Austauschs des Switch-Moduls 206 erhalten.

[0046] Fig. 9 zeigt ein Serversystem 900, das gemäß einer Ausführungsform der Erfindung mehrere

Gestellrahmen umfasst, wobei jeder Gestellrahmen vier Gehäuse 802 aufweist. Die Gehäuse innerhalb jedes Gestellrahmens können funktionsmäßig über eine gehäuseübergreifende Verkabelung 904 miteinander verbunden sein. Gehäuse verschiedener Gestellrahmen können funktionsmäßig über eine gestellübergreifende Verkabelung 906 miteinander verbunden sein. Jedes Gehäuse enthält eine Mittelplatine, die gemäß den hier offengelegten Techniken eine Fabric-Verdrahtung, mindestens eine Zwischenstecker-Karte, die ein Switch-Modul aufweist, und mindestens eine Serverkarte umfasst. Die Switch-Module, die Fabric-Verdrahtung, die gehäuseübergreifende Verkabelung 904 und die gestellübergreifende Verkabelung 906 sowie alle zugehörigen Linkkarten bilden gemeinsam eine Switch-Fabric für das Serversystem 900. Das Serversystem 900 ist dadurch so konfiguriert, dass die Verfügbarkeit der Switch-Fabric und/oder des Serversystems 900 während eines Ausfalls und/oder eines Austauschs eines Switch-Moduls verbessert wird.

[0047] Fig. 10 zeigt ein Serversystem 1000, das gemäß einer Ausführungsform der Erfindung so ausgestattet ist, dass es eine Zwischenstecker-Verbindung umfasst. Die Ausstattung des Serversystems 1000 kann zwischen mindestens einer ersten Zwischenstecker-Karte und einer zweiten Zwischenstecker-Karte eine Zwischenstecker-Verbindung 1006 umfassen. Und jede Zwischenstecker-Karte umfasst ein Switch-Modul 206, das mit zwei Serverkarten 202 verbunden ist. Die Zwischenstecker-Verbindung 1006 kann eine Verkabelung zwischen einem Netzadapter der ersten Zwischenstecker-Karte und einem Netzadapter der zweiten Zwischenstecker-Karte umfassen. Diese Verkabelung kann ihrerseits außerhalb der Mittelplatine liegen. Jeder Netzadapter kann ein CNA 302 der jeweiligen Zwischenstecker-Karte oder ein CNA 304 der entsprechenden Serverkarte sein. Das Serversystem 1000 kann auch Switch-Module 1002 umfassen, die für Konnektivität zu externen Serversystemen und/oder Speichercontrollern sorgen. Die Switch-Module 1002 können funktionsmäßig über einen oder mehrere CNAs 1004 mit den Switch-Modulen 206 verbunden sein.

[0048] Wenn ein Switch-Modul 206₁ der ersten Zwischenstecker-Karte ausfällt, kann folglich ein Switch-Modul 206₂ der zweiten Zwischenstecker-Karte - zusätzlich zum Schalten des Netzverkehrs der Serverkarten 202₃, 202₄ der zweiten Zwischenstecker-Karte - den Netzverkehr für die Serverkarten 202₁, 202₂ der ersten Zwischenstecker-Karte schalten. Auf diese Weise wird durch die Ausstattung des Serversystems mit der Zwischenstecker-Verbindung 1006 vermieden, dass das Switch-Modul 206₁ der ersten Zwischenstecker-Karte ein SPOF ist. Mit anderen Worten, die Serverkarten 202₁, 202₂ der ersten Zwischenstecker-Karte behalten ihre Konnektivität mit der Switch-Fabric und/oder deren Redundanz bei,

und zwar selbst bei einem Ausfall des Switch-Moduls 206₁ der ersten Zwischenstecker-Karte.

[0049] Bei einer Ausführungsform kann die Zwischenstecker-Verbindung 1006 außerdem eine Verkabelung zwischen den CNAs 302₃, 302₄ der zweiten Zwischenstecker-Karte und dem Switch-Modul 206₁ der ersten Zwischenstecker-Karte umfassen. Auf diese Weise wird - zusätzlich zur Vermeidung, dass das Switch-Modul 206₁ ein SPOF ist - vermieden, dass das Switch-Modul 206₂ der zweiten Zwischenstecker-Karte ein SPOF ist. Folglich wird vermieden, dass sowohl das Switch-Modul 206₁ der ersten Zwischenstecker-Karte als auch das Switch-Modul 206₂ der zweiten Zwischenstecker-Karte SPOFs sind.

[0050] Bei einer Ausführungsform ist der CNA 302₄ über eine Verkabelung mit dem Switch-Modul 206₁ verbunden und stellt Redundanz für die zweite Zwischenstecker-Karte bereit. Wenn das Switch-Modul 206₂ der zweiten Zwischenstecker-Karte ausfällt, kann folglich das Switch-Modul 206₁ der ersten Zwischenstecker-Karte - zusätzlich zum Schalten des Netzverkehrs der Serverkarten 202₁, 202₂ der ersten Zwischenstecker-Karte - den Netzverkehr für die Serverkarten 202₃, 202₄ der zweiten Zwischenstecker-Karte schalten.

[0051] Generell wird durch die Ausstattung des Serversystems **1000** mit der zwischen Paaren von Zwischenstecker-Karten eingefügten Zwischenstecker-Verbindung **1006** vermieden, dass die Switch-Module auf jeder Zwischenstecker-Karten SPOFs sind. Jedes Paar Zwischenstecker-Karten kann zwei Zwischenstecker-Karten umfassen, die gemäß einer vordefinierten Achse in einem Gehäuse des Serversystems **1000** in unmittelbarer Nachbarschaft zueinander angebracht sind. Bei einer alternativen Ausführungsform ist jedes Paar Zwischenstecker-Karten in einem Einzelrack im Serversystem **1000** untergebracht. Die vordefinierte Achse kann eine x-Achse, eine y-Achse, eine z-Achse oder jede beliebige Achse umfassen, die geeignet ist, die relativen Positionen der Zwischenstecker-Karten im Gehäuse des Serversystems **1000** zu beschreiben.

[0052] Obwohl hier Ausführungsformen in Bezug auf miteinander verbundene Paare von Zwischenstecker-Karten beschrieben werden, sind außerdem auch andere Ausführungsformen im weitesten Sinne vorgesehen. Bei einer alternativem Ausführungsform können zum Beispiel drei oder mehr Zwischenstecker-Karten in einer Verkettung miteinander verbunden werden. Bei diesem Beispiel umfasst die Zwischenstecker-Verbindung: (i) Verkabelung zwischen einem CNA der ersten Zwischenstecker-Karte und einem Switch-Modul der zweiten Zwischenstecker-Karte, (ii) Verkabelung zwischen einem CNA der zweiten Zwischenstecker-Karte und einem Switch-Modul der dritten Zwischenstecker-Karte und (iii) Verka-

belung zwischen einem CNA der dritten Zwischenstecker-Karte und einem Switch-Modul der ersten Zwischenstecker-Karte. Bei einer alternativen Ausführungsform umfasst die Zwischenstecker-Verbindung eine Verkabelung zwischen dem CNA der dritten Zwischenstecker-Karte und dem Switch-Modul der zweiten Zwischenstecker-Karte (statt der ersten Zwischenstecker-Karte). Um zusätzliche Redundanz zu schaffen, kann jede Zwischenstecker-Karte außerdem mit mehreren anderen Zwischenstecker-Karten verbunden sein. Die Zwischenstecker-Verbindung kann zum Beispiel Folgendes umfassen: (i) eine Verkabelung zwischen einem ersten CNA der dritten Zwischenstecker-Karte und einem Switch-Modul der ersten Zwischenstecker-Karte und (ii) eine Verkabelung zwischen einem zweiten CNA der dritten Zwischenstecker-Karte und einem Switch-Modul der zweiten Zwischenstecker-Karte. Folglich ist beim Auftreten eines Ausfalls von Switch-Modulen in der zweiten bzw. dritten Zwischenstecker-Karte die erste Zwischenstecker-Karte so konfiguriert, dass sie den Netzverkehr für Serverkarten schaltet, die mit der dritten Zwischenstecker-Karte verbunden sind. Der Fachmann wird erkennen, dass mit den hier offenbarten Techniken eine beliebige vordefinierte Anzahl von Zwischenstecker-Karten miteinander verbunden werden kann.

[0053] Fig. 11 zeigt eine Konfiguration **1100** eines Serversystems, das gemäß einer Ausführungsform der Erfindung vermeidet, dass ein Switch-Modul **206** ein Single Point of Failure (SPOF) in einem Paar informationstechnischer Speichereinheiten (Speicher-ITEs) ist. Die Konfiguration **1100** umfasst eine erste Speicher-ITE 1102₁ und eine zweite Speicher-ITE 1102₂. Jede Speicher-ITE 1102₁, 1102₂ kann über eine Zwischenstecker-Karte 204, die ein Switch-Modul **206** umfasst, mit der Mittelplatine verbunden werden. Bei einer alternativen Ausführungsform ist jede Speicher-ITE mit der Mittelplatine verbunden und umfasst das Switch-Modul **206**. Wie gezeigt umfasst jede Speicher-ITE eine Vielzahl von Komponenten, einschließlich einer Funktionskarte **1108**, zweier Ausgangsverzweigungs-Karten **1104** und einer Speichereinheit **1106**. Die Funktionskarte **1108** jeder Speicher-ITE **1102** kann so konfiguriert werden, dass sie die Leistungsmerkmale der jeweiligen Speicher-ITE **1102** anpasst. Zum Beispiel kann die Funktionskarte **1108** so konfiguriert werden, dass die Speicher-ITE als RAID-ITE, Network-attached Storage (NAS) und/oder Dateicachespeicher-ITE usw. angepasst wird. Die Ausgangsverzweigungs-Karten **1104** jeder Speicher-ITE 1102 stellen eine Netzkonnektivität für die Speicher-ITE **1102** bereit und/oder verbessern die Verfügbarkeit der jeweiligen Speicher-ITE **1102**. Jede Ausgangsverzweigungs-Karte **1104** umfasst eine oder mehrere Ausgangsverzweigungs-Komponenten 1112. Jede Speicher-ITE **1102** kann außerdem eine Speicherverbindung 1110₁, 1110₂ umfassen, welche die Komponenten der jeweiligen

Speicher-ITE **1102** funktionsmäßig miteinander verbindet. Bei einer Ausführungsform sorgen die Speicherverbindungen **1110₁**, **1110₂** für SAS-Konnektivität (Serial-Attached SCSI) zwischen den Komponenten der Speicher-ITEs **1102**. Bei alternativen Ausführungsformen sorgen die Speicherverbindungen für FCoE- oder SATA-Konnektivität (Serial-ATA).

[0054] Bei einer Ausführungsform umfasst die Konfiguration **1100** außerdem eine Switch-Verbindung zusammen mit einer Verkabelung **1114** zwischen den Speicherverbindungen **1110** der Speicher-ITEs **1102**. Die Verkabelung **1114** kann eine Ausgangsverzweigungs-Karte **1104₁** der ersten Speicher-ITE **1102₁** mit einer Ausgangsverzweigungs-Karte **1104₄** der zweiten Speicher-ITE **1102₂** verbinden. Je nach Ausführungsform kann die Switch-Verbindung zur Bereitstellung einer größeren Bandbreite eine Verkabelung **1116** zwischen einer zusätzlichen Ausgangsverzweigungs-Karte in jeder Speicher-ITE **1102** umfassen. Wenn das Switch-Modul **206₁** der ersten Speicher-ITE **1102₁** ausfällt, kann folglich ein Switch-Modul **206₂** der zweiten Speicher-ITE **1102₂** - zusätzlich zum Schalten des Netzverkehrs für die zweite Speicher-ITE **1102₂** - den Netzverkehr für die erste Speicher-ITE **1102₁** schalten.

[0055] Durch die Ausstattung der Speicher-ITEs **1102** mit der Verkabelung **1114** zwischen den Speicherverbindungen **1110** wird unter Verwendung der Speicherverbindungen **1110** der Speicher-ITEs **1102** vermieden, dass eines der Switch-Module **206₁**, **206₂** ein SPOF ist.

[0056] Fig. 12 zeigt eine Konfiguration **1200** eines Paares Zwischenstecker-Karten **204**, wobei die Konfiguration **1200** gemäß einer Ausführungsform der Erfindung eine Zwischenstecker-Verbindung umfasst. Statt die Speicherverbindungen zur Vermeidung von SPOFs zu verwenden, umfasst die Konfiguration **1200** für die Vermeidung von SPOFs eine Verkabelung zwischen den zu verschiedenen Zwischenstecker-Karten gehörigen CNAs. Jede Zwischenstecker-Karte **204** umfasst ein Switch-Modul **206** und verbindet zwei Serverkarten **202** funktionsmäßig mit einer Mittelplatine. Das Switch-Modul **206** auf jeder Zwischenstecker-Karte **204** ist dafür konfiguriert, den Netzverkehr für die mit der jeweiligen Zwischenstecker-Karte **204** verbundenen Serverkarten **202** zu schalten. Jede Zwischenstecker-Karte **204** umfasst außerdem einen oder mehrere CNAs **302**. Jede Serverkarte **202** umfasst eine oder mehrere CPUs **106**. Je nach Ausführungsform umfasst jede Serverkarte **202** außerdem einen oder mehrere CNAs **304**. Die Konfiguration **1200** der Zwischenstecker-Karten **202** kann auch eine Verkabelung **1202** zwischen den Zwischenstecker-Karten **202** umfassen. Die Verkabelung **1202** kann die CNAs **304₁**, **304₂** der mit der ersten Zwischenstecker-Karte **204₁** verbundenen Serverkarten **202** mit dem Switch-Modul **206₂** der zwei-

ten Zwischenstecker-Karte **204₂** verbinden. Die Verkabelung **1202** kann auch die CNAs **304₃**, **304₄** der mit der zweiten Zwischenstecker-Karte **204₂** verbundenen Serverkarten **202** mit dem Switch-Modul **206₁** der ersten Zwischenstecker-Karte **204₁** verbinden.

[0057] Wenn das Switch-Modul **206₁** der ersten Zwischenstecker-Karte **204₁** ausfällt, kann das Switch-Modul **206₂** der zweiten Zwischenstecker-Karte **204₂** - zusätzlich zum Schalten des Netzverkehrs der mit der zweiten Zwischenstecker-Karte **204₂** verbundenen Serverkarten **202** - den Netzverkehr für die mit der ersten Zwischenstecker-Karte **204₁** verbundenen Serverkarten **202** verwalten. Auf diese Weise vermeidet die Konfiguration **1200**, dass eines der Switch-Module **206** ein SPOF ist. Mit anderen Worten: Die Serverkarten **202**, die mit jeder Zwischenstecker-Karte verbunden sind, behalten ihre Konnektivität mit der Switch-Fabric und/oder deren Redundanz bei, und zwar selbst bei einem Ausfall eines der Switch-Module **206**.

[0058] Wie oben beschrieben, lässt sich durch die Ausstattung des Serversystems mit einer oder mehreren mit der Mittelplatine verbundenen Serverkarten vermeiden, dass das Switch-Modul ein SPOF ist. In diesem Fall umfasst die Mittelplatine eine Fabric-Verbindung für eine Switch-Fabric. Die Mittelplatine kann eine oder mehrere Platinen mit Gehäuseverbindungselementen (CIE - Chassis Interconnect Element) umfassen, die ihrerseits ein oder mehrere Verarbeitungssteilsysteme umfassen. Obwohl die Ausführungsformen hier unter der Annahme beschrieben werden, dass die Schaltkarten und die CIE-Platinen separate Komponenten sind, können je nach Ausführungsform die Leistungsmerkmale der CIE-Platinen teilweise oder vollständig in die Schaltkarten integriert werden. Die Schaltkarten und/oder die CIE-Platinen können mit einer ersten Teilfläche der Mittelplatine verbunden werden, und eine oder mehrere Serverkarten können mit einer zweiten Teilfläche der Mittelplatine verbunden werden.

[0059] Bei einer Ausführungsform können die Schaltkarten und/oder die CIE-Platinen entlang einer ersten Achse ausgerichtet werden, und die Serverkarten können entlang einer zweiten Achse ausgerichtet werden. Außerdem steht die erste Achse zumindest im Wesentlichen senkrecht zur zweiten Achse. Die Schaltkarten können zum Beispiel vertikal mit der ersten Teilfläche der Mittelplatine verbunden sein, und die Serverkarten können horizontal mit der zweiten Teilfläche der Mittelplatine verbunden sein, oder umgekehrt. Je nach Ausführungsform umfasst die Switch-Fabric eine Verdrahtung, die jede Schaltkarte mit jeder Serverkarte verbindet, und/oder eine Verdrahtung, die jede Schaltkarte mit jeder anderen Schaltkarte verbindet. Diese Verdrahtung stellt redundante Pfade bereit und reduziert und/oder vermeidet SPORs und/oder SPOFs in der Switch-Fabric.

Außerdem kann die Verbindung der Schaltkarten und der Serverkarten mit der Mittelplatine auf senkrecht zu einander stehenden Achsen die erforderliche Verdrahtung (zumindest in einigen Fällen) vereinfachen und/oder verringern.

[0060] Fig. 13 zeigt eine Konfiguration 1300 eines Serversystems, das gemäß einer Ausführungsform der Erfindung mehrere Schaltkarten 1302 umfasst. Wie gezeigt, umfasst die Konfiguration 1300 eine Mittelplatine 123, Serverkarten 202 und CIE-Platinen 1304. Die Serverkarten 202 sind horizontal mit der Mittelplatine 123 verbunden, und die Schaltkarten 1302 und die CIE-Platinen 1304 sind vertikal mit der Mittelplatine 123 verbunden. Die Schaltkarten 1302 und/oder die Serverkarten 202 können von der Mittelplatine 123 aus während des Betriebs ausgetauscht werden. Jede Schaltkarte 1302 kann ein oder mehrere Switch-Module 206 umfassen, und jede Serverkarte 202 kann eine oder mehrere CPUs umfassen.

[0061] Je nach Ausführungsform kann die Switch-Fabric eine Verdrahtung zwischen jeder Schaltkarte 1302 oder jedem Switch-Modul 206 und jeder Serverkarte 202 oder jeder CPU umfassen.

[0062] Die Switch-Fabric kann auch eine Verdrahtung zwischen jeder Schaltkarte 1302 oder jedem Switch-Modul 206 und jeder anderen Schaltkarte 1302 oder jedem anderen Switch-Modul 206 umfassen. Folglich stellt die Konfiguration 1300 redundante Pfade zwischen Elementen der Switch-Fabric bereit und vermeidet damit, dass ein Switch-Modul 206 und/oder eine Schaltkarte 1302 ein SPOF in der Switch-Fabric ist. Insbesondere die Serverkarten 202 behalten beim Ausfall eines Switch-Moduls 206 oder einer Schaltkarte 1302 die Konnektivität mit der Switch-Fabric bei. Außerdem hat auch eine Reparatur an einem ausgefallenen Switch-Modul 206 oder einer ausgefallenen Schaltkarte 1302 keine Auswirkung auf die Konnektivität der Serverkarten 202 mit der Switch-Fabric. Die Reparatur kann den Austausch der Schaltkarte 206 gegen eine zweite mit einem funktionsfähigen Switch-Modul ausgestattete Schaltkarte ohne Neustart des Serversystems und/oder der Switch-Fabric umfassen.

[0063] Fig. 14 zeigt eine logische Ansicht 1400 einer Konfiguration eines Serversystems, das gemäß einer Ausführungsform der Erfindung mehrere Schaltkarten umfasst. Wie gezeigt umfasst die logische Ansicht 1400 mehrere Serverkarten 202 und mehrere Switch-Module 206, 1304 eines Serversystems. Das Serversystem kann auch eine Fabric-Verbindung für eine Switch-Fabric umfassen. In der Switch-Fabric sind die Switch-Module 206 in der Nähe der Serverkarten 202 angeordnet und stellen eine Redundanz für das Schalten des Netzverkehrs für die Serverkarten 202 bereit. Die Switch-Module 206 können auch als Nord-Switch-Module bezeichnet werden. Die Switch-Modu-

le 1404 sind in der Nähe der Nord-Switch-Module angeordnet und sorgen für eine Konnektivität zwischen den Nord-Switch-Modulen und dem Rest der Switch-Fabric. Die Switch-Module 1404 können auch als Süd-Switch-Module bezeichnet werden.

[0064] Bei einer Ausführungsform umfasst jede Serverkarte 202 zwei CPUs 106 und zwei CNAs 304. Die Switch-Fabric kann eine Verdrahtung 1404 zwischen jedem Switch-Modul 206 und jeder Serverkarte 202 umfassen. Die Switch-Fabric kann außerdem ein lokale Gestell-Verschaltungseinheit 1402 umfassen, die ihrerseits eine Verdrahtung zwischen den Switch-Modulen 206 und den Switch-Modulen 1404 umfasst. Je nach Ausführungsform kann die lokale Gestell-Verschaltungseinheit 1402 auch eine Verdrahtung von jedem der Switch-Module 206 zu allen anderen Switch-Modulen 206 umfassen, wodurch unter den Switch-Modulen 206 eine Konnektivität aller mit allen bereitgestellt wird. Vorteilhafterweise vermeidet die Konfiguration, dass eines Switch-Module 206 in der Switch-Fabric ein SPOF ist. Sollte zum Beispiel das Switch-Modul 206₁ ausfallen, kann das Switch-Modul 206₂ den Netzverkehr für jede Serverkarte 202 weiterleiten. Dadurch kann jede Serverkarte 202 trotz des Ausfalls im Switch-Modul 206₁ mit der Switch-Fabric in Verbindung bleiben.

[0065] Fig. 15 zeigt eine Konfiguration 1500 eines Serversystems, das gemäß einer Ausführungsform der Erfindung mehrere Schaltkarten 1302 umfasst. Wie gezeigt umfasst jede der Vielzahl von Schaltkarten 1302 zwei Switch-Module 206 oder Nord-Switch-Module. Jedes Switch-Modul verwaltet den Netzverkehr für mindestens eine Serverkarte 202. Die Konfiguration 1500 umfasst eine Verdrahtung 1404 zwischen jedem Server-Switch-Modul 206 und jeder Serverkarte 202 aus einer Untergruppe. Untergruppen der Serverkarten 202 können zum Beispiel eine erste Untergruppe mit den Serverkarten 202₁₋₇ und eine zweite Untergruppe mit den Serverkarten 202₈₋₁₄ umfassen. Die Konfiguration 1500 kann auch in jeder Untergruppe von Switch-Modulen 206 eine (nicht gezeigte) Verdrahtung aller mit allen umfassen. Die Untergruppen der Switch-Module 206 können zum Beispiel eine erste Untergruppe mit den Switch-Modulen 206₁₋₄ und eine zweite Untergruppe mit den Switch-Modulen 206₅₋₈ umfassen. Mit anderen Worten: Die Schaltkarten 1302 und/oder die Serverkarten 202 können hinsichtlich der Verdrahtung in der Switch-Fabric physisch in verschiedene Untergruppen unterteilt sein, wobei die Verdrahtung für eine Konnektivität und/oder Redundanz in der Switch-Fabric sorgt. Die Konfiguration 1500 kann auch eine die Untergruppen der Switch-Module 206 übergreifende zusätzliche Verdrahtung umfassen.

[0066] Bei einer Ausführungsform umfassen die Schaltkarten 1302_{1,3} außerdem ein oder mehrere Verwaltungsverarbeitungsteilsysteme. Die Ver-

waltungsverarbeitungsteilsysteme umfassen Verwaltungs-Firmware, die dafür konfiguriert ist, den Zustand des Serversystems und/oder der Switch-Fabric-Elemente zu überwachen, die Elemente zu konfigurieren und/oder einen Ausfall der Elemente zu erkennen und darauf zu reagieren. Wie gezeigt, umfassen die Verwaltungsverarbeitungsteilsysteme einen Ein-/Ausgabe-Master-Controller (IoMC - Input/output Master Controller) und ein Gehäuseleistungselement (CSE - Chassis Service Element). Der IoMC verwaltet Elemente in der Switch-Fabric, während das CSE Komponenten im Servergehäuse umfasst. Außerdem können die IoMCs funktionsmäßig über eine IoMC-Verbindung auf den Schaltkarten **1302** miteinander verbunden sein, um eine Redundanz bei der Überwachung und/oder Verwaltung der Switch-Fabric bereitzustellen, wobei die IoMC-Verbindung eine physische Verdrahtung zwischen den IoMCs umfasst. Bei einem Ausfall eines ersten IoMC 1506_1 kann ein zweiter IoMC 1506_2 so konfiguriert werden, dass er anstelle des ersten IoMC 1506_1 Fähigkeiten zum Überwachen und/oder Verwalten der Switch-Fabric aufweist. Vorteilhafterweise vermeidet die Konfiguration **1500**, dass eines der Switch-Module **206** ein SPOF ist, und außerdem, dass einer der IoMC 1506 in der Switch-Fabric ein SPOF ist.

[0067] Bei einer Ausführungsform umfassen die Schaltkarten $1302_{2,4}$ außerdem Switch-Module **1404** oder Süd-Switch-Module. Wie oben beschrieben, sorgen die Süd-Switch-Module für Konnektivität zwischen den Nord-Switch-Modulen und dem Rest der Switch-Fabric. Jede Schaltkarte **1302** umfasst außerdem lokale Links (L-Links) **1502** und Distanz-Links (D-Links) **1504**. Der L-Link **1502** stellt für die Schaltkarte **1302** physische Konnektivität zu einem anderen Gehäuse innerhalb eines einzelnen physischen Gestellrahmens bereit. Der D-Link **1504** stellt für die Schaltkarte **1302** physische Konnektivität zu einem anderen Gehäuse in anderen physischen Gestellrahmen bereit. Folglich umfasst die Switch-Fabric dadurch zusätzliche redundante Pfade für Elemente der Switch-Fabric.

[0068] Fig. 16 zeigt ein Serversystem **1600**, das gemäß einer Ausführungsform der Erfindung einen Verwaltungscontroller **1604** und eine unabhängige Verwaltungsverbindung **1606** umfasst. Wie gezeigt, umfasst das Serversystem **1600** mehrere Gestellrahmen **804**, wobei jeder Gestellrahmen ein oder mehrere Gehäuse **802** umfasst, die funktionsmäßig über die Gehäuse-Verbindungskarten **806** und die zugehörige Verkabelung **808** verbunden sind. Die Gehäuse **802** können außerdem funktionsmäßig gestellrahmenübergreifend über die Gestellrahmen-Verbindungskarte **810** und die zugehörige Verkabelung **812** miteinander verbunden sein. Jedes Gehäuse **802** enthält eine Mittelplatine **123**, die gemäß den hier offenbarten Techniken über eine Zwischenstecker-Karte **204** mit einer oder zwei Serverkarten **202** ver-

bunden ist. Jede Zwischenstecker-Karte **204** umfasst ein Switch-Modul **206** zum Schalten des Netzwerkverkehrs für die Serverkarten **202**. Die Mittelplatinen **123** umfassen eine Fabric-Verdrahtung zum Verbinden der Switch-Module **206**. Die Switch-Module **206**, die Fabric-Verdrahtung, die Gehäuse-Verbindungskarten **806** und die zugehörige Verkabelung **808** sowie die Gestellrahmen-Verbindungskarte **810** und die zugehörige Verkabelung **812** bilden gemeinsam eine Switch-Fabric für das Serversystem **1600**. Der Verwaltungscontroller **1604** kann in einer Verwaltungseinheit **1602** enthalten sein und so konfiguriert werden, dass er die Switch-Module **206** über die Verwaltungsverbindung **1606** unabhängig von der Switch-Fabric verwaltet.

[0069] Bei einer Ausführungsform kann die Verwaltungsverbindung **1606** so konfiguriert werden, dass sie zugeordnete Pfade zwischen dem Verwaltungscontroller **1604** und den Switch-Modulen **206** bereitstellt. Dies ermöglicht dem Verwaltungscontroller **1604**, die Switch-Module **206** zu verwalten. Bei einer Ausführungsform ist die Verwaltungsverbindung ein Out-of-band-Netz, das physisch von der Switch-Fabric isoliert ist. Dieses Out-of-band-Netz dient als zugeordneter Verwaltungskanal für die Verwaltung der Switch-Module **206**. Der Verwaltungscontroller **1604** kann zum Beispiel funktionsmäßig über die zugeordneten Pfade mit mehreren integrierten Verwaltungsmodulen (IMMs) verbunden sein. Jedes IMM kann ein Serviceprozessor sein, der funktionsmäßig mit einem entsprechenden Switch-Modul **206** verbunden ist. Gemeinsam tauschen die IMMs über die Verwaltungsverbindung **1606** Daten mit dem Verwaltungscontroller **1604** aus und/oder koordinieren sie mit ihm, um die Switch-Module **206** der Switch-Fabric zu verwalten.

[0070] Fig. 17 zeigt ein Serversystem **1700**, das gemäß einer Ausführungsform der Erfindung einen Verwaltungscontroller **1604** umfasst, ohne auch eine unabhängige Verwaltungsverbindung **1606** zu umfassen. Bei einer Ausführungsform kann das Serversystem **1700** so konfiguriert werden, dass es das Fabric-Management über die Switch-Fabric statt über die unabhängige Verwaltungsverbindung **1606** durchführt. In diesem Fall kann die Switch-Fabric eine Verdrahtung zwischen dem Verwaltungscontroller **1604** und der Switch-Fabric umfassen. Bei einer Ausführungsform kann die Verdrahtung den Verwaltungscontroller **1604** physisch mit jedem Switch-Modul **206** verbinden. Das Vermeiden der unabhängigen Verwaltungsverbindung bei der Ausstattung des Serversystems **1700** kann die Verdrahtung im Serversystem **1700** zumindest in einigen Fällen verringern, wodurch die Kosten und/oder der Aufwand der Ausstattung des Serversystems verringert werden.

[0071] Bei einer Ausführungsform überwacht der Verwaltungscontroller **1604** vordefinierte Ereignisse,

z.B. auf Änderungen in der Switch-Fabric, wie zum Beispiel ein physisches Hinzufügen oder Entfernen eines Switch-Moduls **206**, Änderungen der Konnektivität eines Switch-Moduls **206**, Ausfall eines Switch-Moduls **206** usw., und reagiert darauf. Bei einer Ausführungsform kann der Verwaltungscontroller **1604** bei der Verwaltung der Switch-Module **206** eine oder mehrere Operationen an den Switch-Modulen **206** durchführen. Der Verwaltungscontroller kann zum Beispiel ein Switch-Modul **206** einschalten (oder ausschalten), Daten mit dem Switch-Modul **206** austauschen, Firmware in einem Switch-Modul **206** aktualisieren, ein Switch-Modul **206** neu starten, ein Switch-Modul **206** initialisieren, ein Switch-Modul **206** überwachen und ein Switch-Modul konfigurieren. Bei einer Ausführungsform können diese Operationen als Reaktion auf eine Benutzeranfrage durchgeführt werden (z.B., von einem Benutzer mit Verwaltungsaufgaben). Die Operationen können auch das Erkennen eines Ausfalls eines Switch-Moduls **206** und als Reaktion darauf das Benachrichtigen anderer Switch-Module im Fabric-Netzverkehr über den Ausfall des Switch-Moduls **206** umfassen. Als Reaktion darauf können die anderen Switch-Module so konfiguriert werden, dass sie den Netzverkehr für das als ausgefallen angezeigte Switch-Modul **206** schalten. Außerdem kann der Verwaltungscontroller **1604** nach dem Austausch des ausgefallenen Switch-Moduls das funktionsfähige Switch-Modul wieder in die Switch-Fabric integrieren. Der Verwaltungscontroller **1604** kann dann das funktionsfähige Switch-Modul damit beauftragen, den Netzverkehr für das ausgefallene (und ausgetauschte) Switch-Modul zu schalten. Der Verwaltungscontroller **1604** kann auch das eine oder die mehreren anderen Switch-Module damit beauftragen, das Schalten des Netzverkehrs für das jetzt ausgetauschte Switch-Modul zu stoppen. Zum Überwachen der Switch-Fabric kann der Verwaltungscontroller **1604** bei einer Ausführungsform Daten mit mehreren Switch-Modulen **206** über die Switch-Fabric austauschen, um aussagekräftige Daten über die Leistung der Switch-Fabric zu erfassen. Der Verwaltungscontroller **1604** kann dann die erfassten Daten dazu verwenden, eine umfassende Übersicht der Auslastungsgrade der Switch-Fabric bereitzustellen.

[0072] Ein Beispiel für ein vordefiniertes Ereignis ist ein interner Fehler, der in einer Komponente des Switch-Moduls **206** auftritt und das Switch-Modul **206** veranlasst, seine Schaltfunktion einzustellen. Bei einer Ausführungsform erkennt der Verwaltungscontroller **1604** den Fehler über die Switch-Fabric, da die benachbarten Switch-Module nicht mehr in der Lage sind, Daten mit dem ausgefallenen Switch-Modul **206** auszutauschen. Folglich kann der Verwaltungscontroller **1604** einem Benutzer anzeigen, dass das ausgefallene Switch-Modul **206** ausgetauscht werden muss. Nachdem das Switch-Modul **206** ausgetauscht wurde, kann der Verwaltungscontroller **1604**

das neue Switch-Modul wieder in die Switch-Fabric integrieren und dem neuen Switch-Modul erlauben, den Netzverkehr für die Switch-Fabric zu schalten.

[0073] Obwohl einige Ausführungsformen hier in Bezug auf ein Serversystem beschrieben werden, das keine unabhängige Verwaltungsverbindung umfasst, sind auch andere Ausführungsformen im weitesten Sinne vorgesehen. Zum Beispiel kann das Serversystem in einer alternativen Ausführungsform außer der Switch-Fabric eine oder mehrere unabhängige Verbindungen umfassen, aber der Verwaltungscontroller verwendet die unabhängigen Verbindungen nicht (oder konfiguriert sie nicht für eine Verwendung) bei der Verwaltung der Switch-Module.

[0074] Fig. 18 ist ein Flussdiagramm, das gemäß einer Ausführungsform der Erfindung ein Verfahren **1800** zur Verringerung der Verdrahtung in einem Serversystem darstellt, welches einen Verwaltungscontroller umfasst. Wie gezeigt, beginnt das Verfahren **1800** mit dem Schritt **1810**, in dem ein Serversystem-Anbieter das Serversystem so ausstattet, dass es eine Mittelplatine umfasst, die ihrerseits eine Fabric-Verbindung für eine Switch-Fabric umfasst. Im Schritt **1820** stattet der Serversystem-Anbieter das Serversystem außerdem so aus, dass es eine oder mehrere Serverkarten umfasst, die funktionsmäßig mit der Mittelplatine verbunden sind, wobei jede Serverkarte einen oder mehrere Computerprozessoren und einen Speicher umfasst. Im Schritt **1830** stattet der Serversystem-Anbieter das Serversystem außerdem so aus, dass es außerdem Switch-Module umfasst, die funktionsmäßig mit der Mittelplatine verbunden sind, wobei jedes Switch-Modul dafür konfiguriert wird, dass es den Netzverkehr für mindestens eine Serverkarte schaltet. Im Schritt **1840** stattet der Serversystem-Anbieter das Serversystem außerdem so aus, dass es einen Verwaltungscontroller umfasst, der so konfiguriert wird, dass er die Switch-Module über die Fabric-Verbindung verwaltet. Das Verfahren **1800** endet nach dem Schritt **1840**.

[0075] Fig. 19 ist ein Flussdiagramm, das gemäß einer Ausführungsform der Erfindung ein Verfahren **1900** zum Verwalten einer Switch-Fabric darstellt. Wie gezeigt, beginnt das Verfahren **1900** mit dem Schritt **1910**, in dem ein Serversystem-Anbieter das Serversystem so ausstattet, dass es eine Mittelplatine, eine oder mehrere Serverkarten und Switch-Module umfasst. Die Mittelplatine umfasst eine Fabric-Verbindung für eine Switch-Fabric, und alle Serverkarten und alle Switch-Module sind funktionsmäßig mit der Mittelplatine verbunden. Außerdem ist jedes Switch-Modul dafür konfiguriert, den Netzverkehr für mindestens eine Serverkarte zu schalten. Das Serversystem umfasst außerdem einen Verwaltungscontroller. Im Schritt **1920** erkennt der Verwaltungscontroller, dass ein Switch-Modul ausgefallen ist. Im Schritt **1930** sendet der Verwaltungscontroller nach

dem Erkennen des Ausfalls des Switch-Moduls über die Fabric-Verbindung eine Anweisung an ein oder mehrere andere Switch-Module, den Netzverkehr für das ausgefallene Switch-Modul zu schalten. Das Verfahren **1900** endet nach dem Schritt **1930**.

[0076] Die Ausführungsformen der Erfindung verringern in vorteilhafter Weise die Verdrahtungsanforderungen bei Serversystemen, die einen Verwaltungscontroller zur Verwaltung einer Switch-Fabric verwenden. Bei einer Ausführungsform kann das Serversystem eine Mittelplatine umfassen, die ihrerseits eine Fabric-Verbindung für die Switch-Fabric umfasst. Das Serversystem umfasst außerdem eine oder mehrere Serverkarten, die funktionsmäßig mit der Mittelplatine verbunden sind, wobei jede Serverkarte einen oder mehrere Computerprozessoren und einen Speicher umfasst. Das Serversystem umfasst auch Switch-Module, die funktionsmäßig mit der Mittelplatine verbunden sind, wobei jedes Switch-Modul so konfiguriert wurde, dass es den Netzverkehr für mindestens eine Serverkarte schaltet. Der Verwaltungscontroller kann dafür konfiguriert werden, die Switch-Module über die Fabric-Verbindung zu verwalten. Da für die Verwaltung der Switch-Module keine andere Verbindung zwischen den Switch-Modulen erforderlich ist, wird damit die Verdrahtung im Serversystem verringert. Dadurch werden die Kosten und/oder der Aufwand der Ausstattung des Serversystems (zumindest in einigen Fällen) verringert.

[0077] Außerdem können einige der Ausführungsformen der Erfindung auch die Auswirkung eines Switch-Ausfalls in der Switch-Fabric verringern. Die Serverkarten können mit der Mittelplatine verbunden sein, wobei jede Serverkarte von der Mittelplatine aus während des Betriebs ausgetauscht werden kann. Die Switch-Module können in einer oder mehreren Schaltkarten enthalten sein, die ihrerseits auch mit der Mittelplatine verbunden sind, und dadurch sind die eine oder die mehreren Schaltkarten funktionsmäßig mit der einen oder mehreren Serverkarten verbunden. Jede Schaltkarte kann auch von der Mittelplatine aus während des Betriebs ausgetauscht werden. Um redundante Pfade bereitzustellen, kann die Switch-Fabric eine Verdrahtung zwischen jeder Schaltkarte und jeder Serverkarte und/oder eine Verdrahtung zwischen jeder Schaltkarte und jeder anderen Schaltkarte umfassen. Wenn das Switch-Modul einer ersten Schaltkarte ausfällt, beauftragt der Verwaltungscontroller folglich ein zweites Switch-Modul auf der ersten Schaltkarte oder auf einer zweiten Schaltkarte damit, den Netzverkehr für die von dem Switch-Modul der ersten Schaltkarte unterstützten Serverkarten weiterzuleiten. Dementsprechend vermeidet diese Konfiguration des Serversystems, dass das Switch-Modul der ersten Schaltkarte ein SPOF ist.

[0078] Darüber hinaus können Ausführungsformen der Erfindung auch die Auswirkung von Reparaturen an der Switch-Fabric verringern. Bei einer Ausführungsform kann ein Serversystem so konfiguriert sein, dass bei einem Ausfall des Switch-Moduls der ersten Schaltkarte die erste Schaltkarte durch eine dritte Schaltkarte mit einem funktionsfähigen Switch-Modul ausgetauscht werden kann, ohne dass das Serversystem und/oder die Switch-Fabric abgeschaltet oder neu gestartet werden muss. Auf diese Weise verbessert eine solche Konfiguration die Verfügbarkeit des Serversystems und/oder der Switch-Fabric, während sie gleichzeitig die mit Reparaturen einhergehenden Kosten verringert.

Patentansprüche

1. System, das Folgendes umfasst: eine Mittelplatine, die eine Fabric-Verbindung für eine Switch-Fabric umfasst; eine oder mehrere Serverkarten, die funktionsmäßig mit der Mittelplatine verbunden sind, wobei jede Serverkarte einen oder mehrere Computerprozessoren und einen Speicher umfasst; eine Vielzahl von Switch-Modulen, die funktionsmäßig mit der Mittelplatine verbunden sind, wobei jedes Switch-Modul dafür konfiguriert ist, den Netzverkehr für mindestens eine der einen oder mehreren Serverkarten zu schalten; und einen Verwaltungscontroller, der dafür konfiguriert ist, die Vielzahl von Switch-Modulen über die Fabric-Verbindung zu verwalten.
2. System nach Anspruch 1, wobei die Fabric-Verbindung die einzige Verbindung zwischen den Switch-Modulen und dem Verwaltungscontroller ist.
3. System nach Anspruch 1, wobei die Fabric-Verbindung die einzige Verbindung ist, die vom Verwaltungscontroller für die Verwaltung der Vielzahl von Switch-Modulen verwendet wird.
4. System nach Anspruch 1, wobei das Verwalten der Vielzahl von Switch-Modulen mindestens eine der folgenden Funktionen umfasst: Einschalten von, Ausschalten von, Datenaustausch mit, Aktualisierung der Firmware in, Neustart von, Initialisierung von, Überwachung von und Konfiguration von mindestens einem Switch-Modul der Vielzahl von Switch-Modulen.
5. System nach Anspruch 1, wobei das Verwalten der Vielzahl von Switch-Modulen folgendes umfasst: Ausführen einer vordefinierten Maßnahme als Reaktion auf ein Erkennen eines Ausfalls mindestens eines Switch-Moduls.
6. System nach Anspruch 5, wobei die vordefinierte Maßnahme folgendes umfasst:

Senden einer Anweisung an ein oder mehrere Switch-Module, den Netzverkehr für das mindestens eine ausgefallene Switch-Modul zu schalten.

7. System nach Anspruch 1, wobei mindestens eines der Vielzahl von Switch-Modulen so konfiguriert ist, dass es ausgetauscht werden kann, ohne einen Neustart des Systems erforderlich zu machen und ohne einen Neustart der Switch-Fabric erforderlich zu machen.

8. Verwaltungscontroller, der Folgendes umfasst: einen Computerprozessor; und einen Speicher, in dem Verwaltungs-Firmware gespeichert ist, die, wenn sie auf dem Computerprozessor ausgeführt wird, eine Operation ausführt, die Folgendes umfasst:

Verwalten einer Vielzahl von Switch-Modulen, die funktionsmäßig mit einer Mittelplatine eines Serversystems verbunden sind, wobei die Mittelplatine eine Fabric-Verbindung für eine Switch-Fabric umfasst, wobei die Vielzahl von Switch-Modulen über die Fabric-Verbindung verwaltet wird, wobei die Mittelplatine funktionsmäßig mit einer oder mehreren Serverkarten verbunden und wobei jedes Switch-Modul so konfiguriert ist, dass es den Netzverkehr für mindestens eine der Serverkarten schaltet.

9. Verwaltungscontroller nach Anspruch 8, wobei die Fabric-Verbindung die einzige Verbindung zwischen den Switch-Modulen und dem Verwaltungscontroller ist.

10. Verwaltungscontroller nach Anspruch 8, wobei die Fabric-Verbindung die einzige Verbindung ist, die vom Verwaltungscontroller für die Verwaltung der Vielzahl von Switch-Modulen verwendet wird.

11. Verwaltungscontroller nach Anspruch 8, wobei das Verwalten der Vielzahl von Switch-Modulen mindestens eine der folgenden Funktionen umfasst: Einschalten von, Ausschalten von, Datenaustausch mit, Aktualisierung der Firmware in, Neustart von, Initialisierung von, Überwachung von und Konfiguration von mindestens einem Switch-Modul der Vielzahl von Switch-Modulen.

12. Verwaltungscontroller nach Anspruch 8, wobei das Verwalten der Vielzahl von Switch-Modulen folgendes umfasst:

Ausführen einer vordefinierten Maßnahme als Reaktion auf ein Erkennen eines Ausfalls mindestens eines Switch-Moduls.

13. Verwaltungscontroller nach Anspruch 12, wobei die vordefinierte Maßnahme folgendes umfasst: Senden einer Anweisung an ein oder mehrere Switch-Module, den Netzverkehr für das mindestens eine ausgefallene Switch-Modul zu schalten.

14. Verwaltungscontroller nach Anspruch 8, wobei mindestens eines der Vielzahl von Switch-Modulen so konfiguriert ist, dass es ausgetauscht werden kann, ohne einen Neustart des Serversystems erforderlich zu machen und ohne einen Neustart der Switch-Fabric erforderlich zu machen.

15. Computergestütztes Verfahren, das Folgendes umfasst:

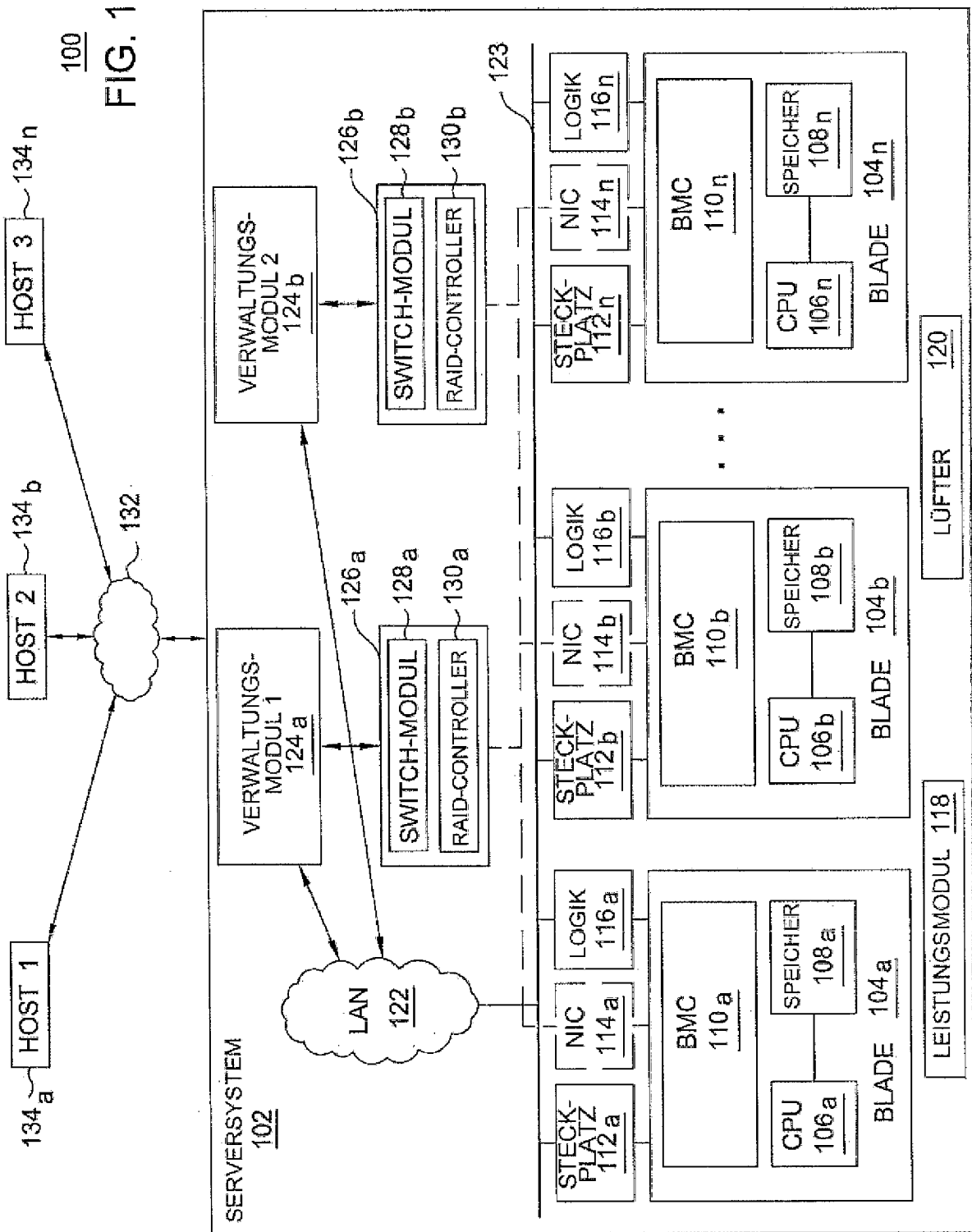
Bereitstellen eines Serversystems, das eine Mittelplatine, eine oder mehrere Serverkarten und eine Vielzahl von Switch-Modulen umfasst, wobei die Mittelplatine eine Fabric-Verbindung für eine Switch-Fabric umfasst, wobei jede Serverkarte und jedes Switch-Modul funktionsmäßig mit der Mittelplatine verbunden und wobei jedes Switch-Modul so konfiguriert ist, dass es den Netzverkehr für mindestens eine der einen oder mehreren Serverkarten schaltet; und Verwalten der Vielzahl von Switch-Modulen über die Fabric-Verbindung und durch den Betrieb von einem oder mehreren Computerprozessoren, wobei die Fabric-Verbindung vorzugsweise die einzige Verbindung zwischen den Switch-Modulen und dem Verwaltungscontroller ist, oder wobei die Fabric-Verbindung vorzugsweise die einzige Verbindung ist, die vom Verwaltungscontroller für die Verwaltung der Vielzahl von Switch-Modulen verwendet wird und/oder, wobei die Verwaltung der Vielzahl von Switch-Modulen mindestens eine der folgenden Funktionen umfasst:

Einschalten von, Ausschalten von, Datenaustausch mit, Aktualisierung der Firmware in, Neustart von, Initialisierung von, Überwachung von und Konfiguration von mindestens einem Switch-Modul der Vielzahl von Switch-Modulen und/oder, wobei das Verwalten der Vielzahl von Switch-Modulen folgendes umfasst: Ausführen einer vordefinierten Maßnahme als Antwort auf ein Erkennen eines Ausfalls mindestens eines Switch-Moduls, wobei die vordefinierte Maßnahme vorzugsweise folgendes umfasst:

Senden einer Anweisung an ein oder mehrere Switch-Module, den Netzverkehr für das mindestens eine ausgefallene Switch-Modul zu schalten.

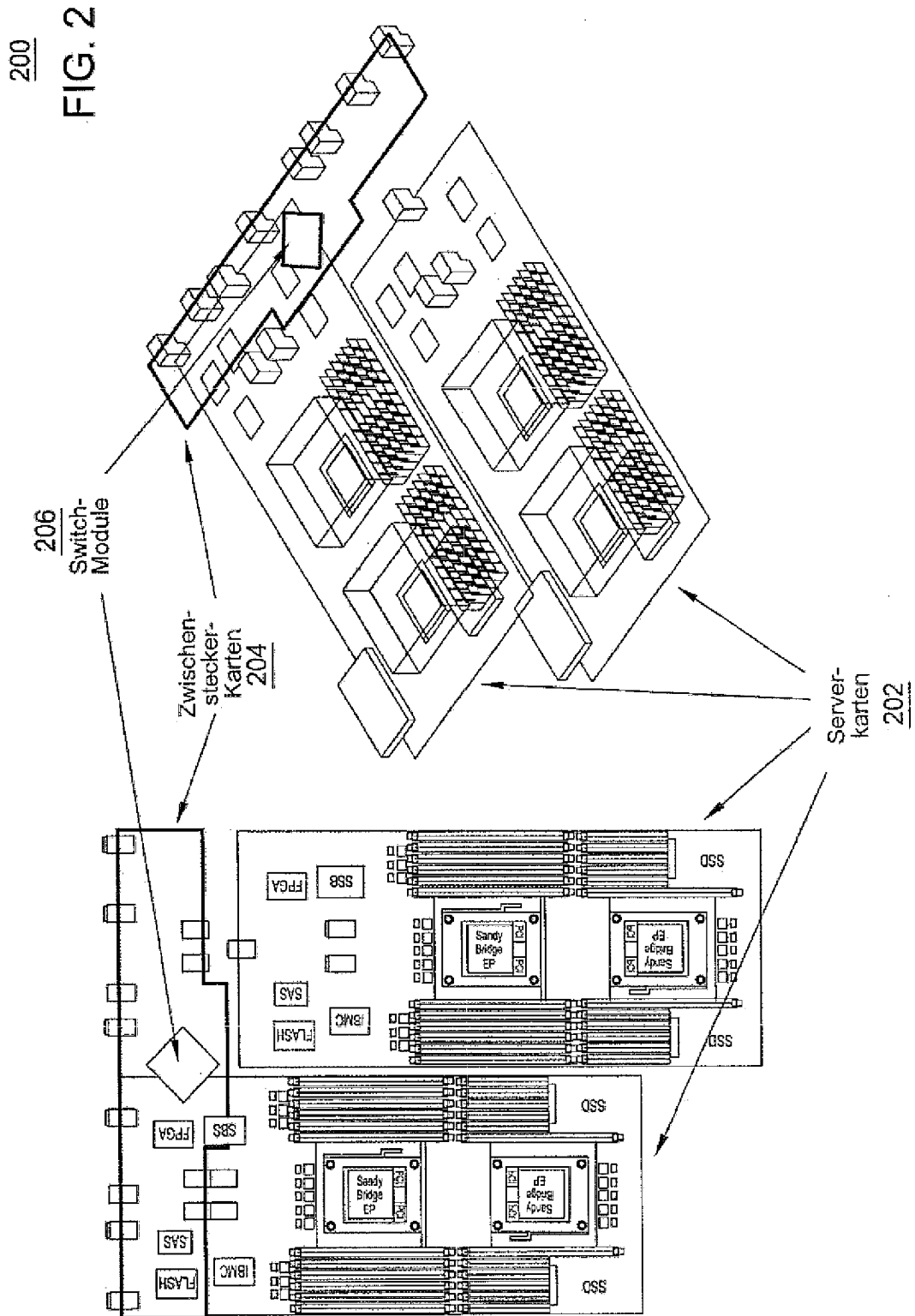
Es folgen 19 Seiten Zeichnungen

Anhängende Zeichnungen

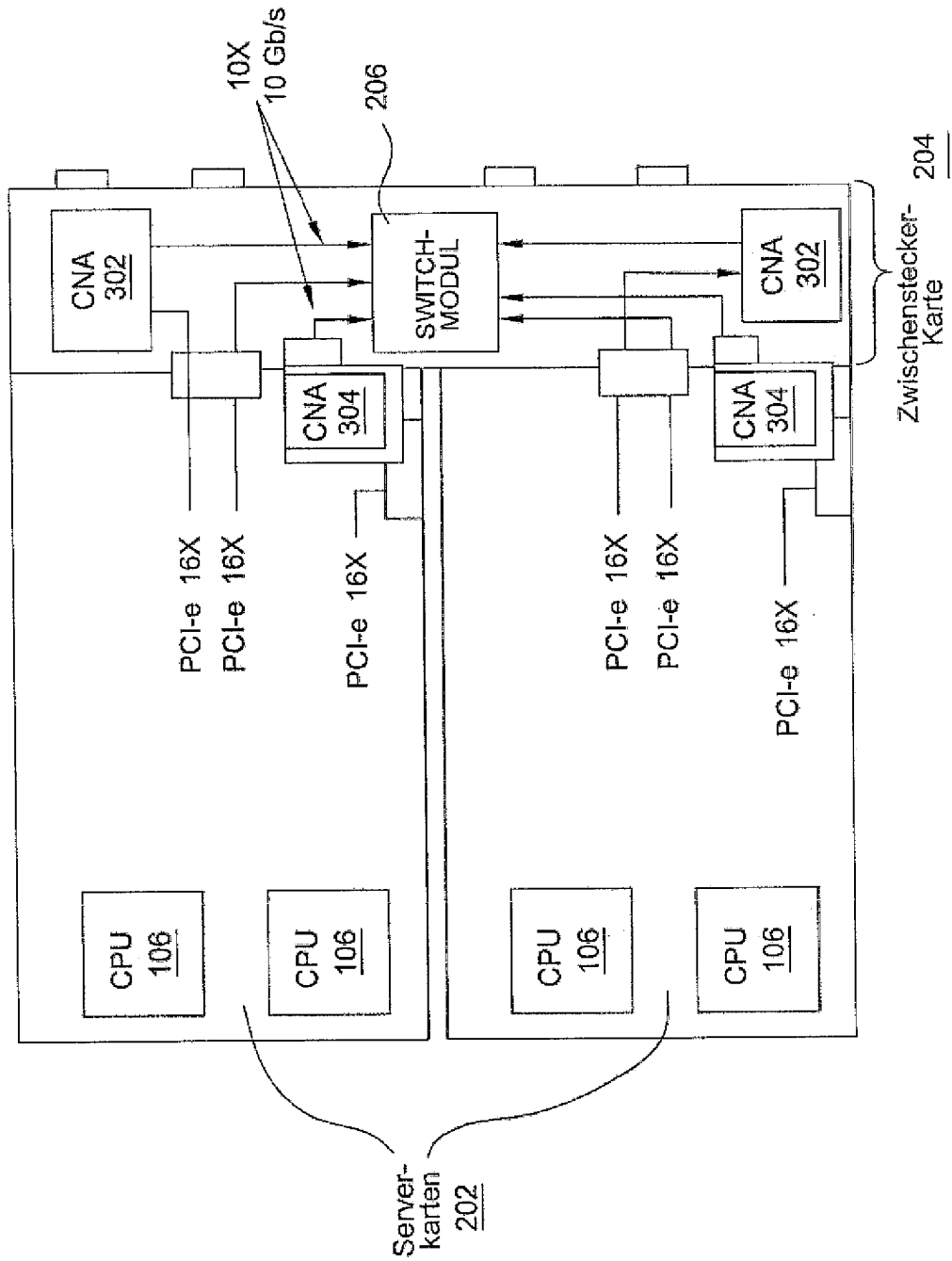


100

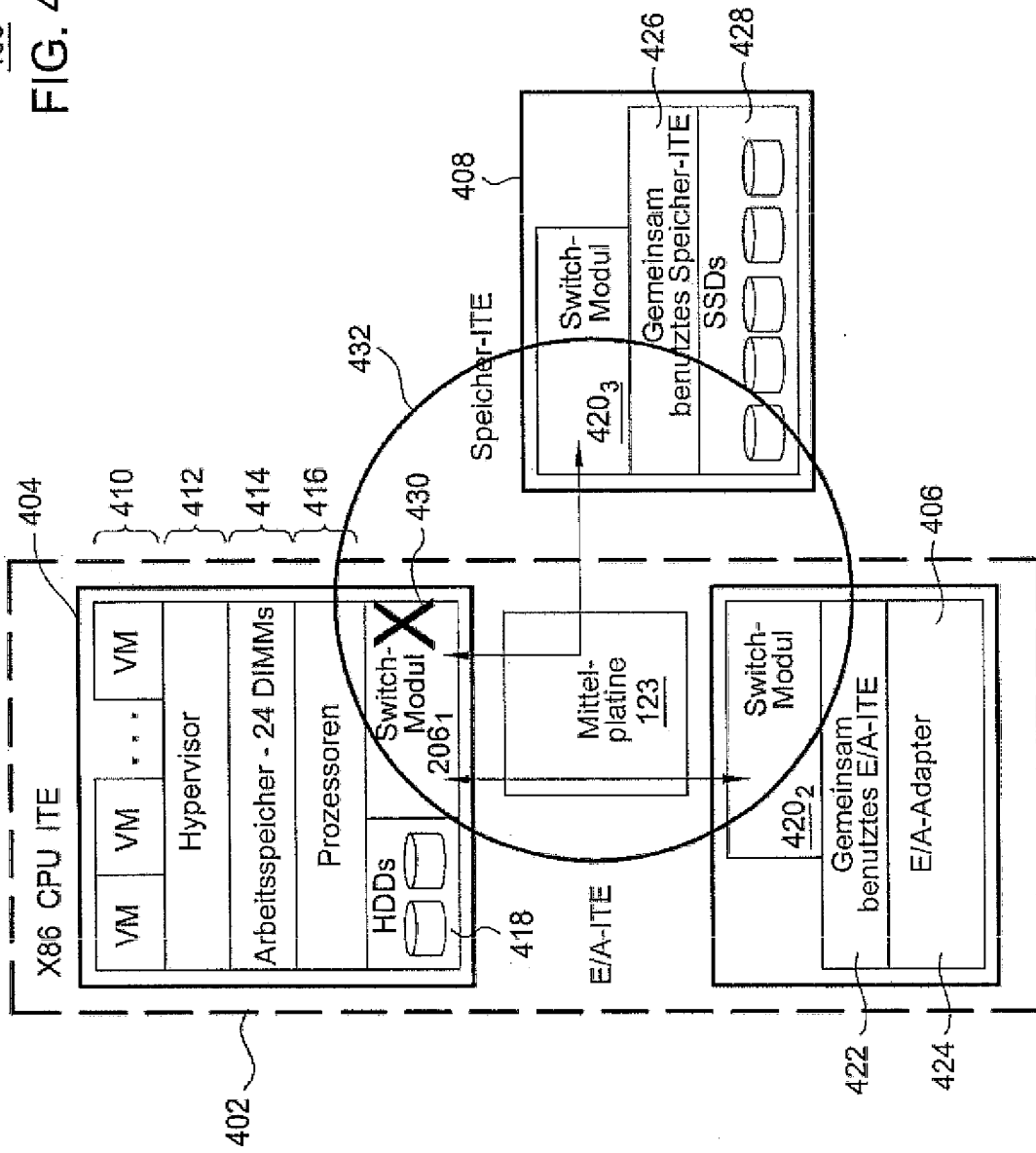
FIG. 1



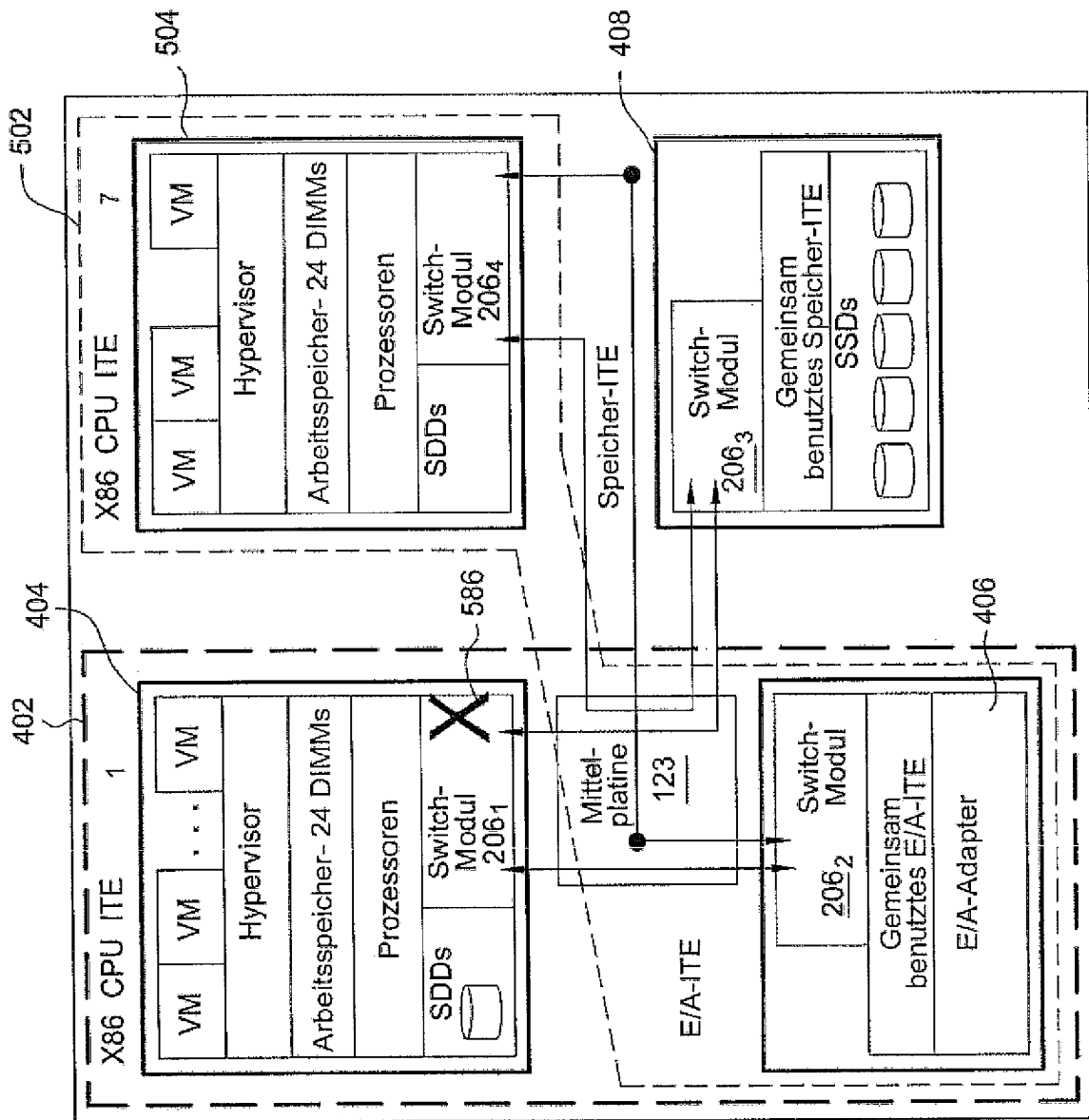
300
FIG. 3



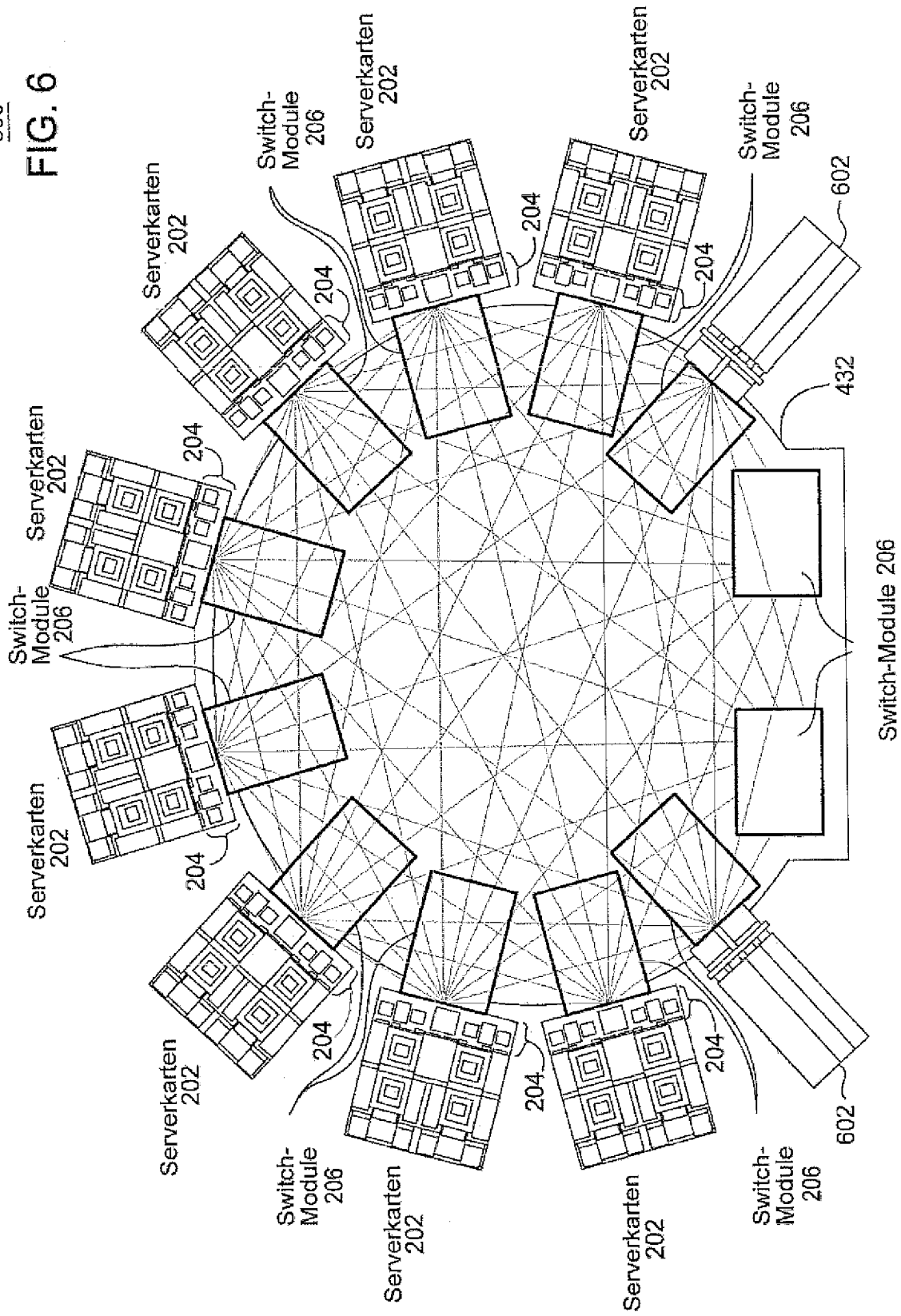
400
FIG. 4



500
FIG. 5

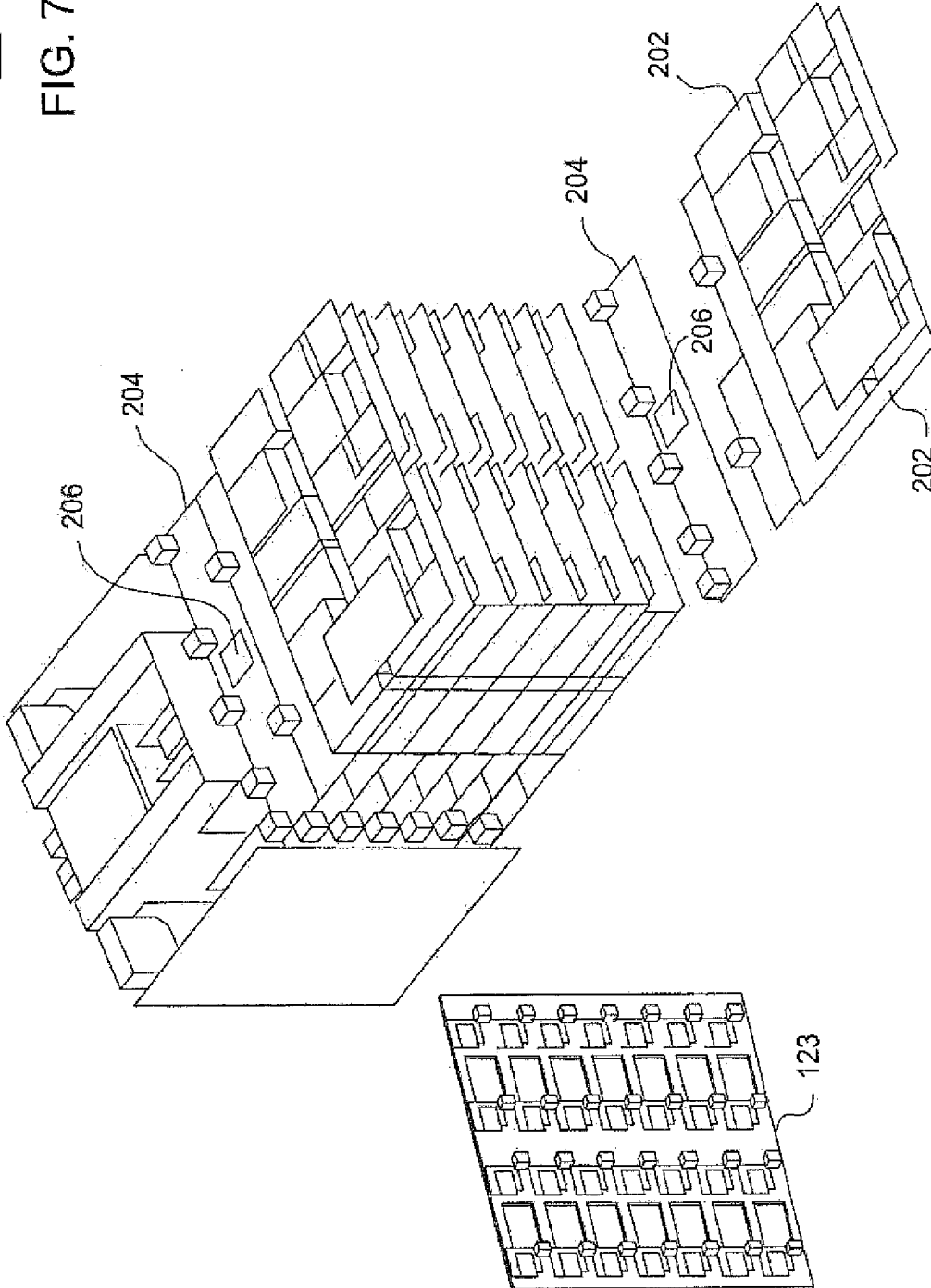


600
FIG. 6



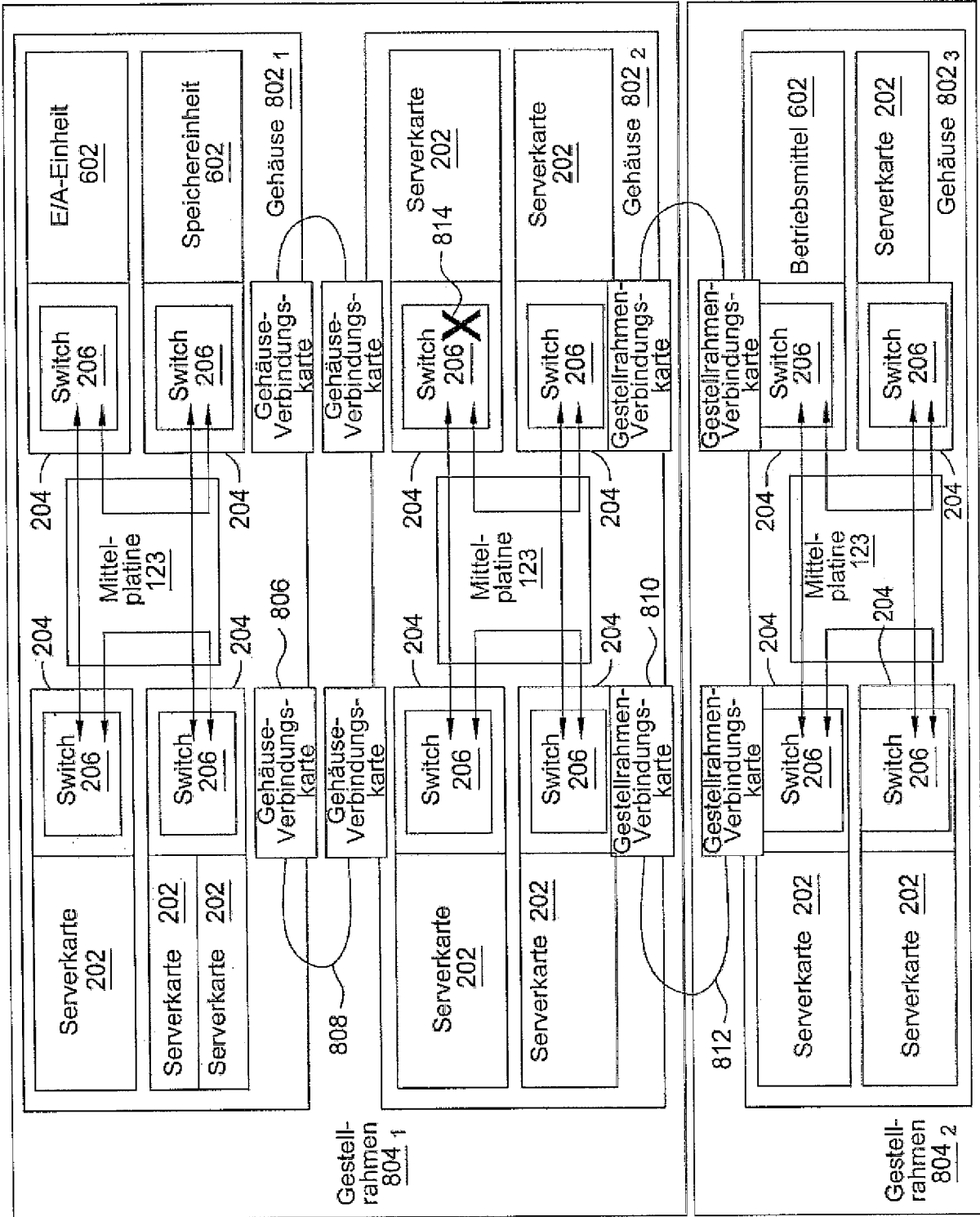
700

FIG. 7

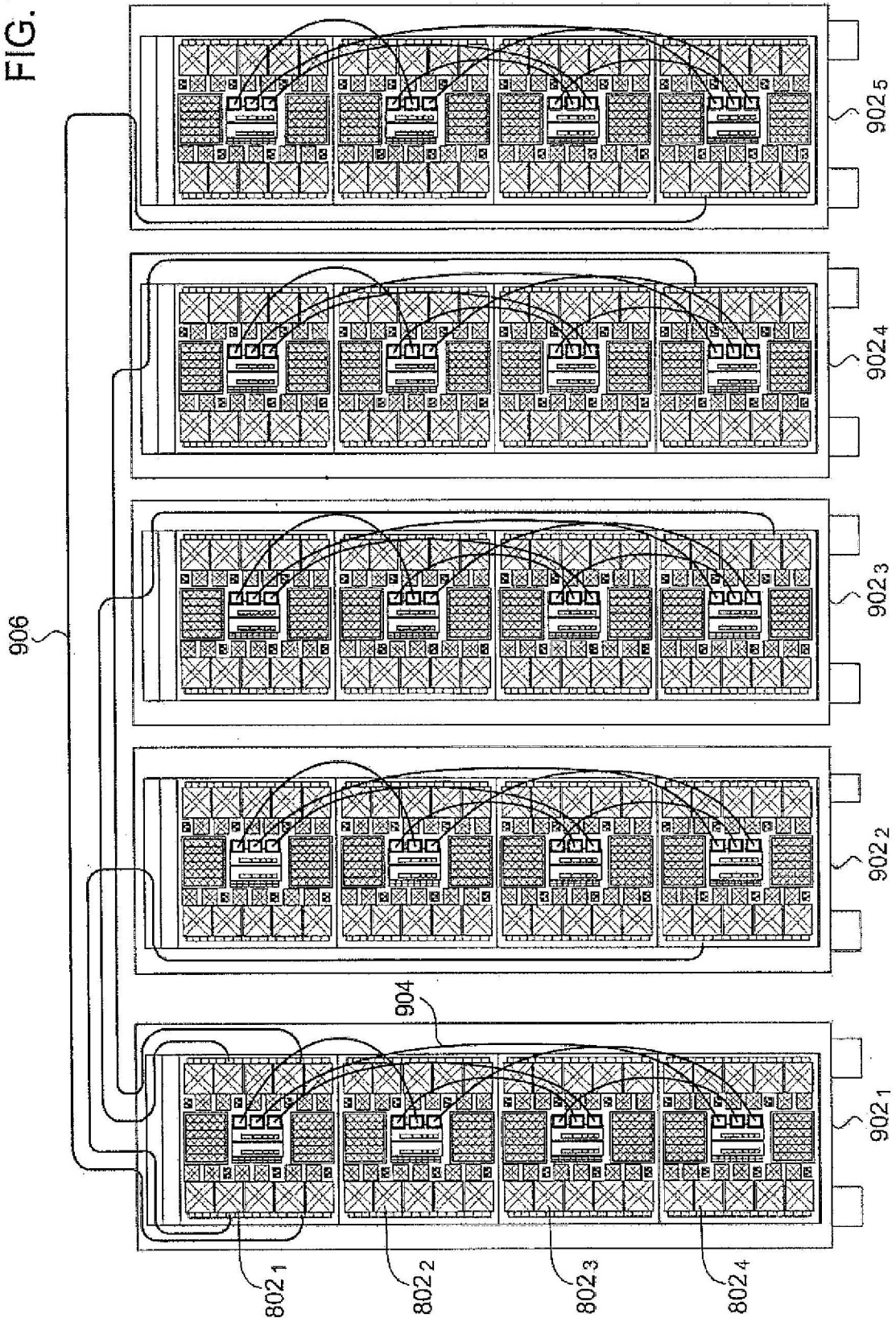


800

FIG. 8



900
FIG. 9



1000

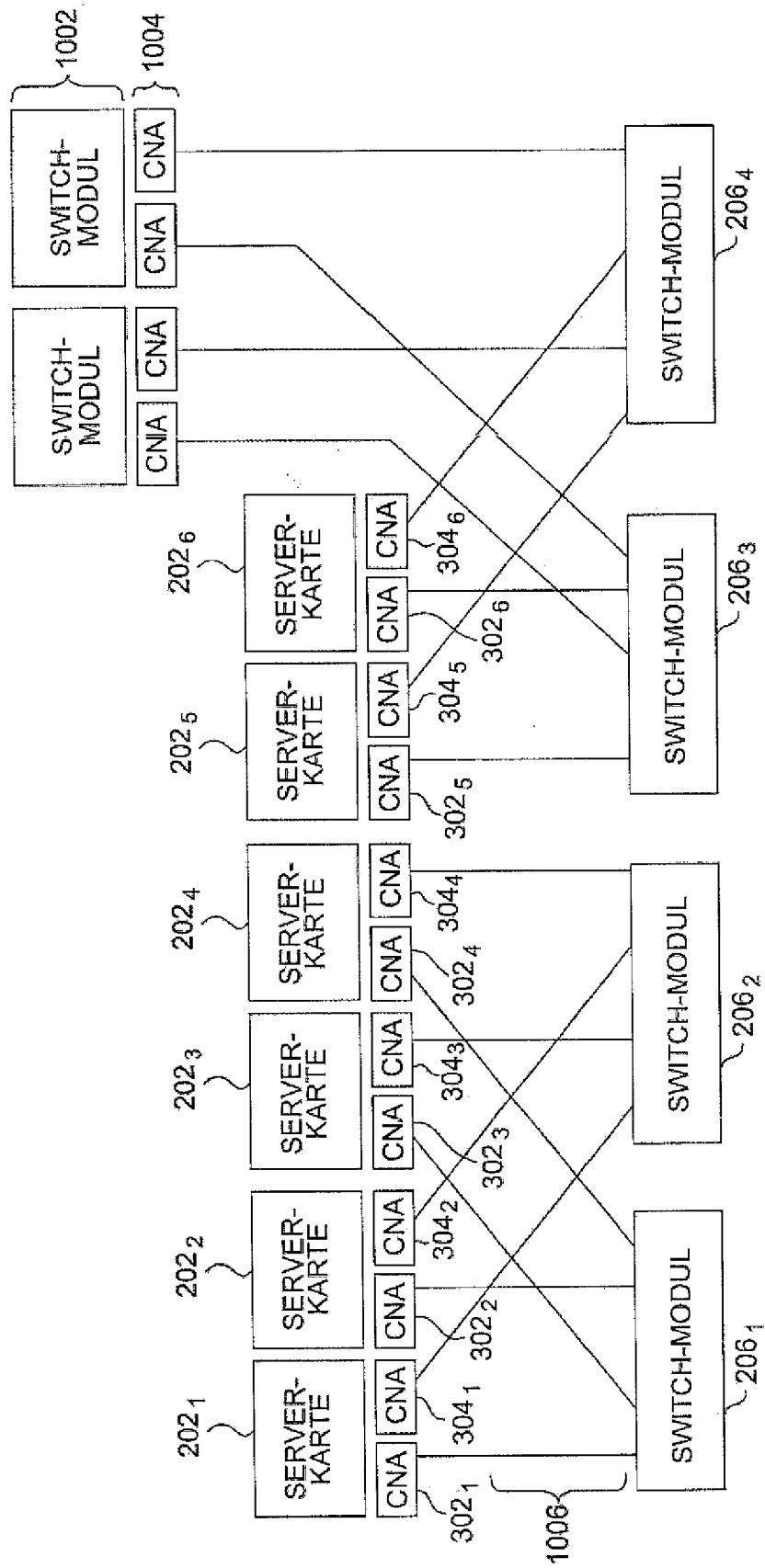


FIG. 10

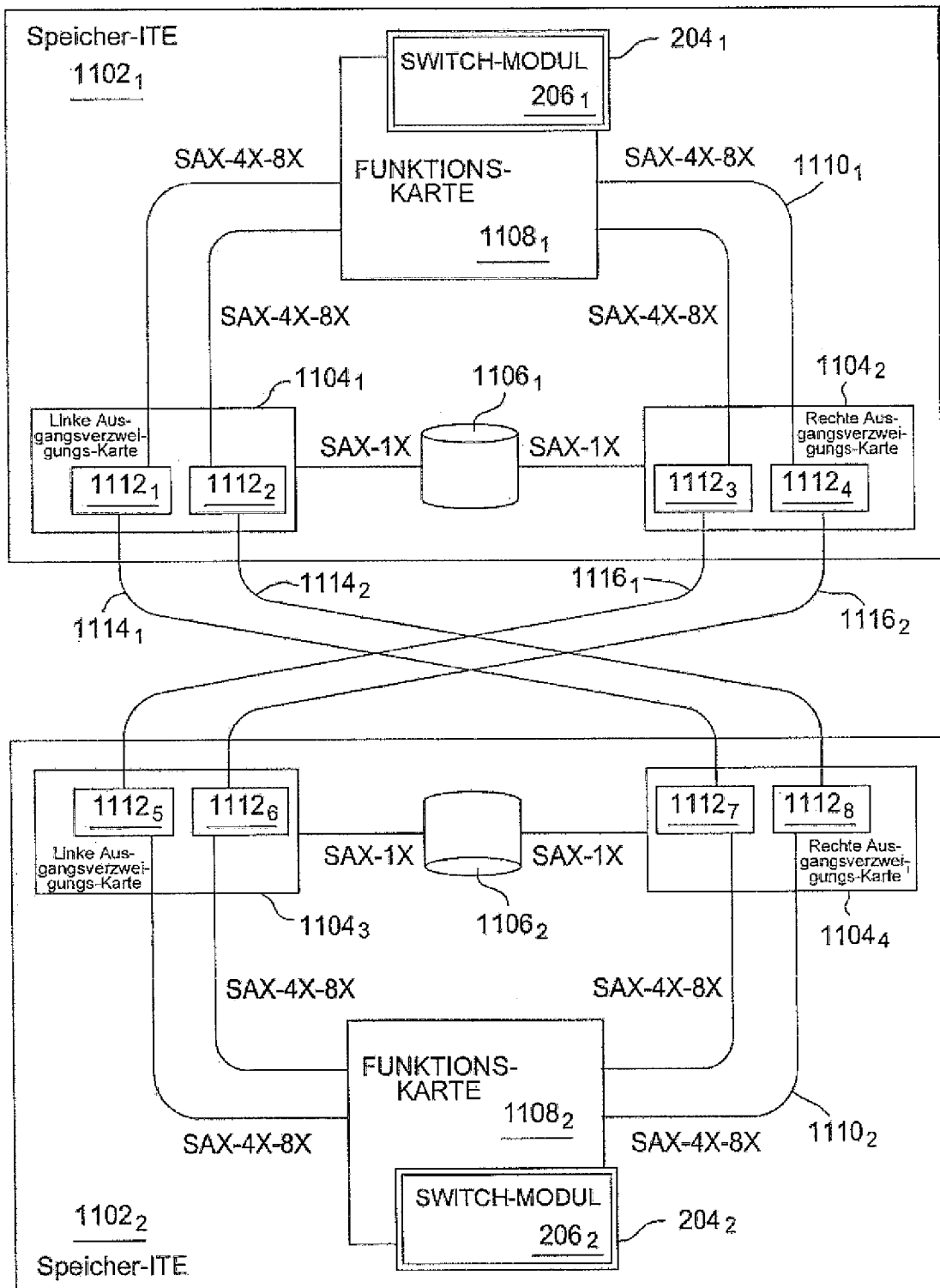
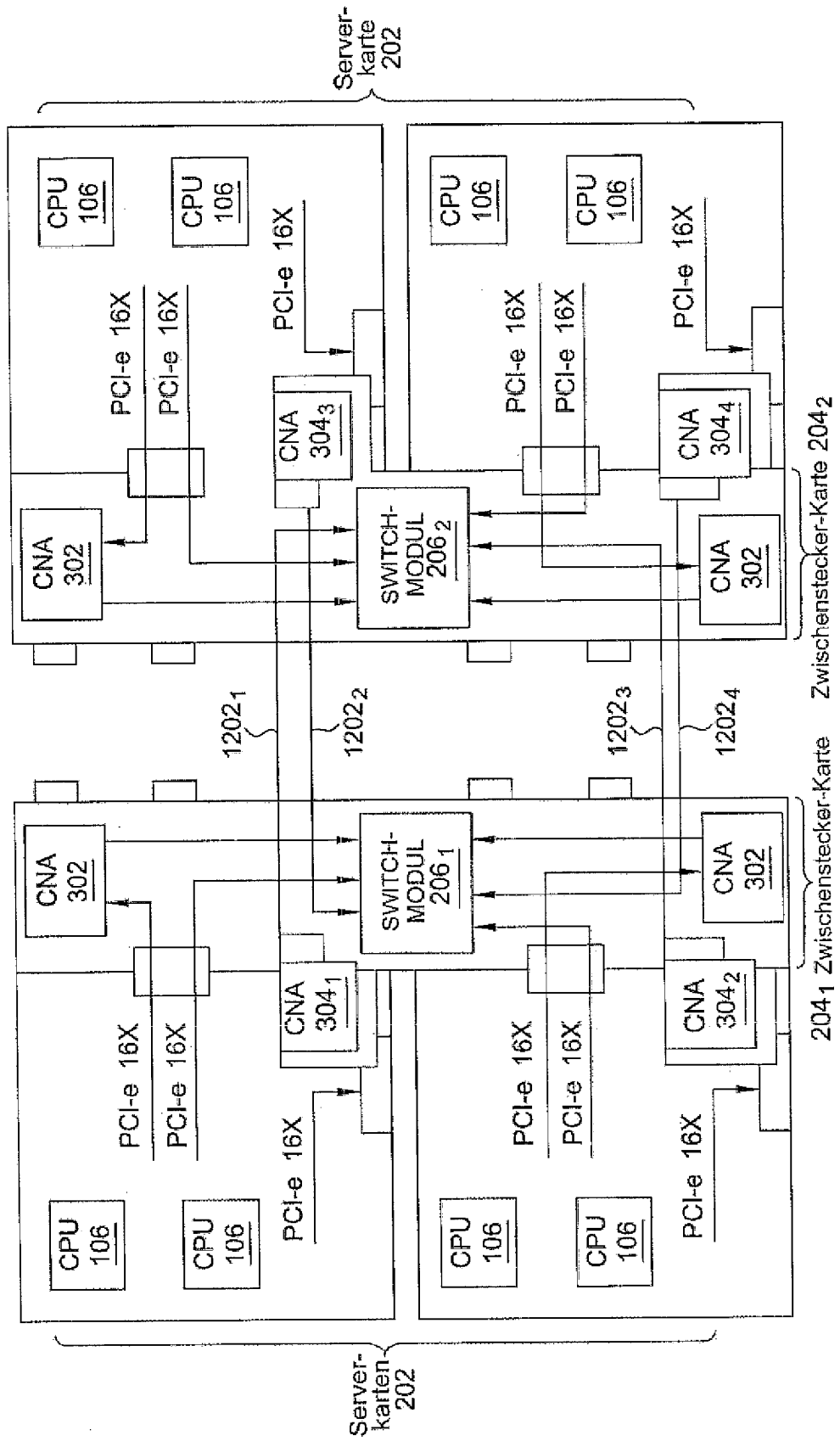


FIG. 11

1200

FIG. 12



1300

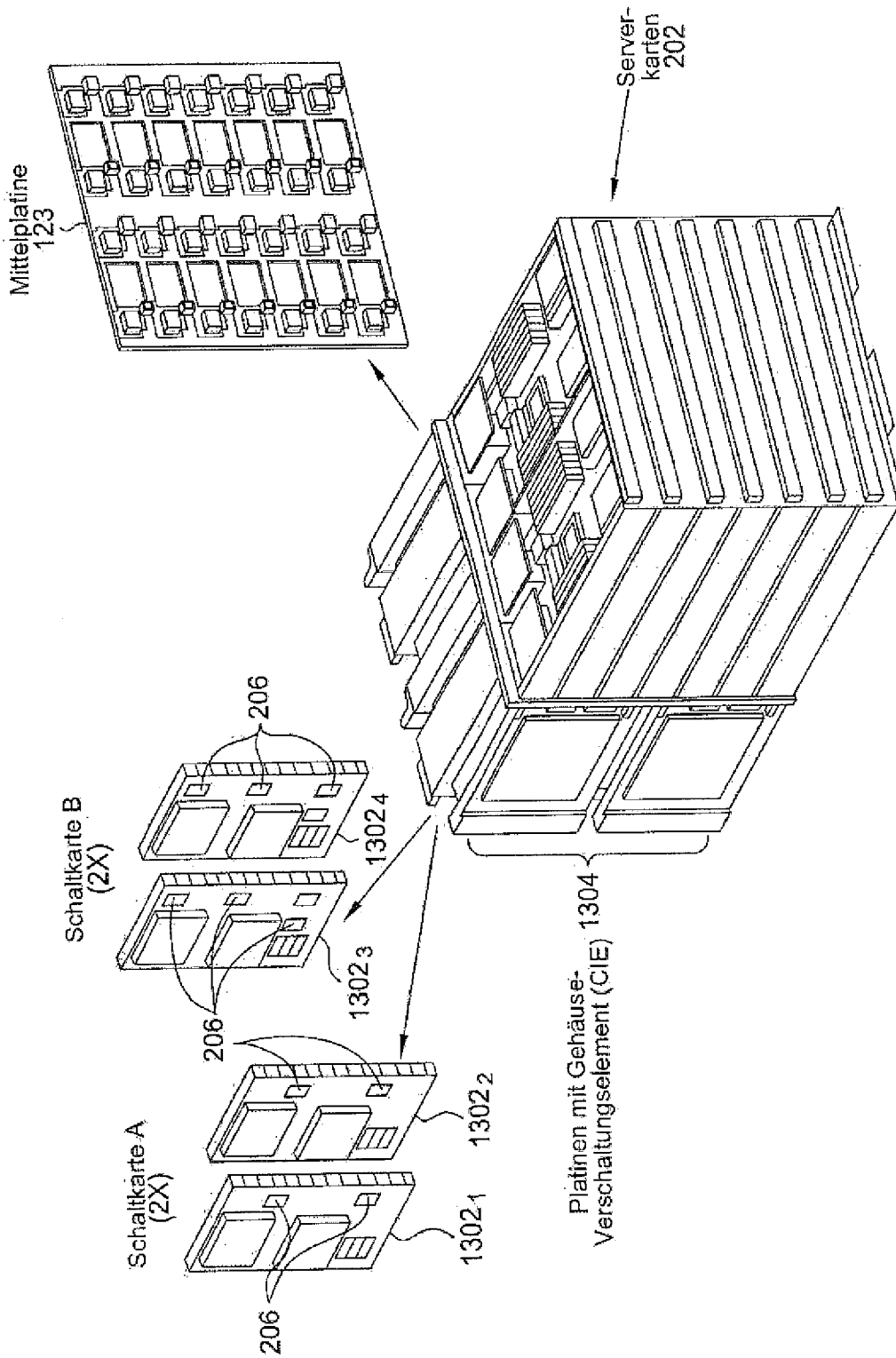


FIG. 13

1400

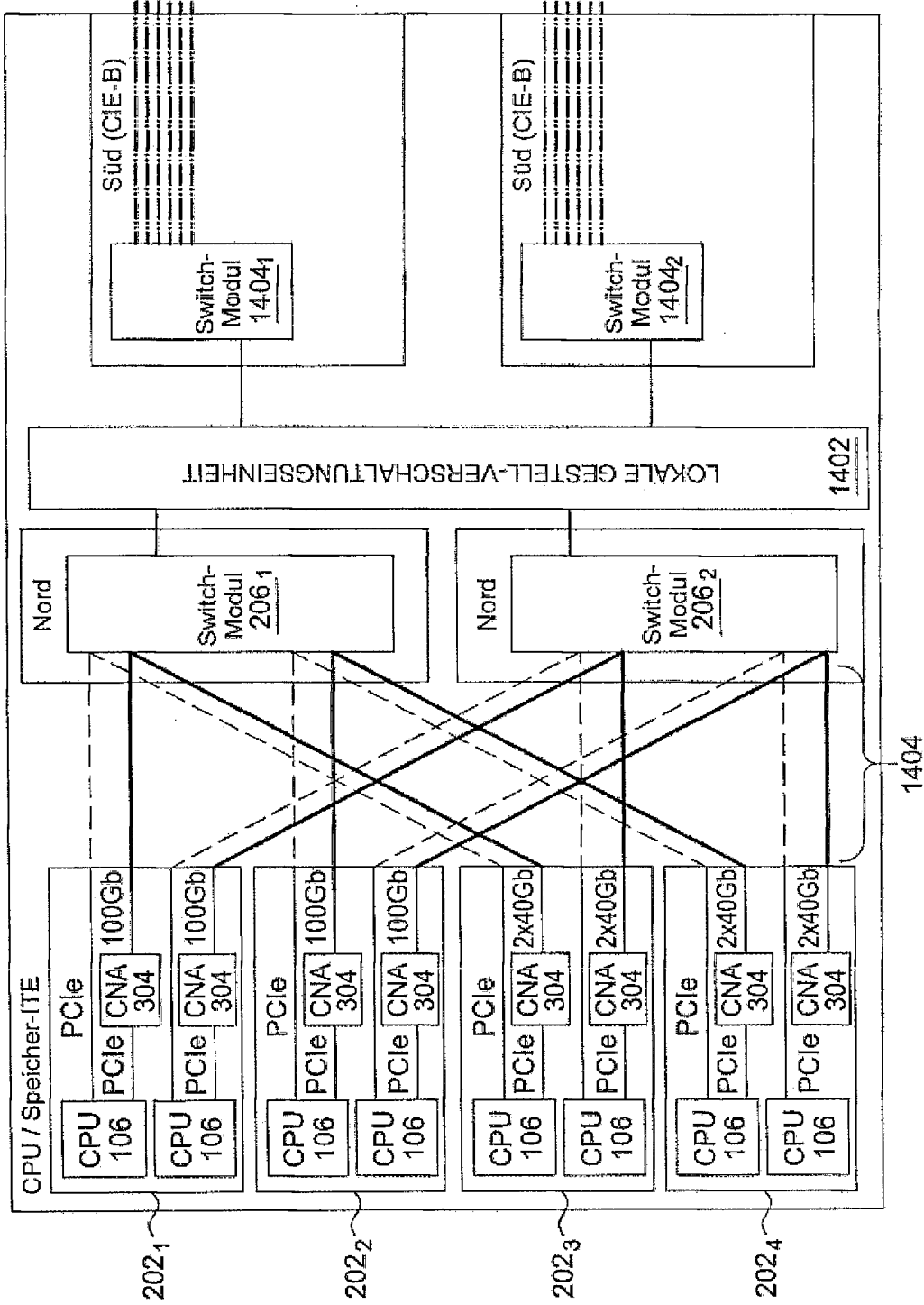


FIG. 14

1500

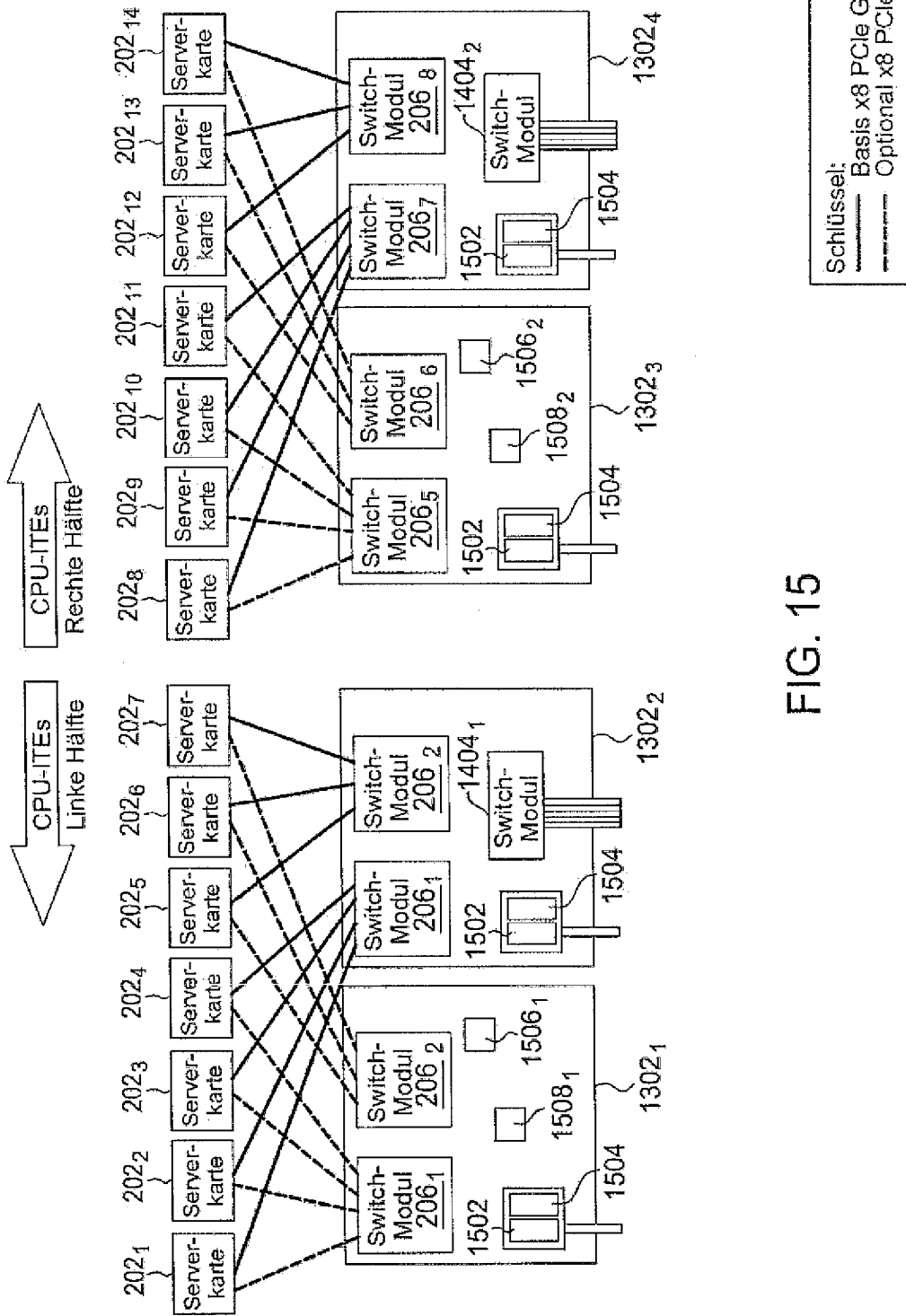
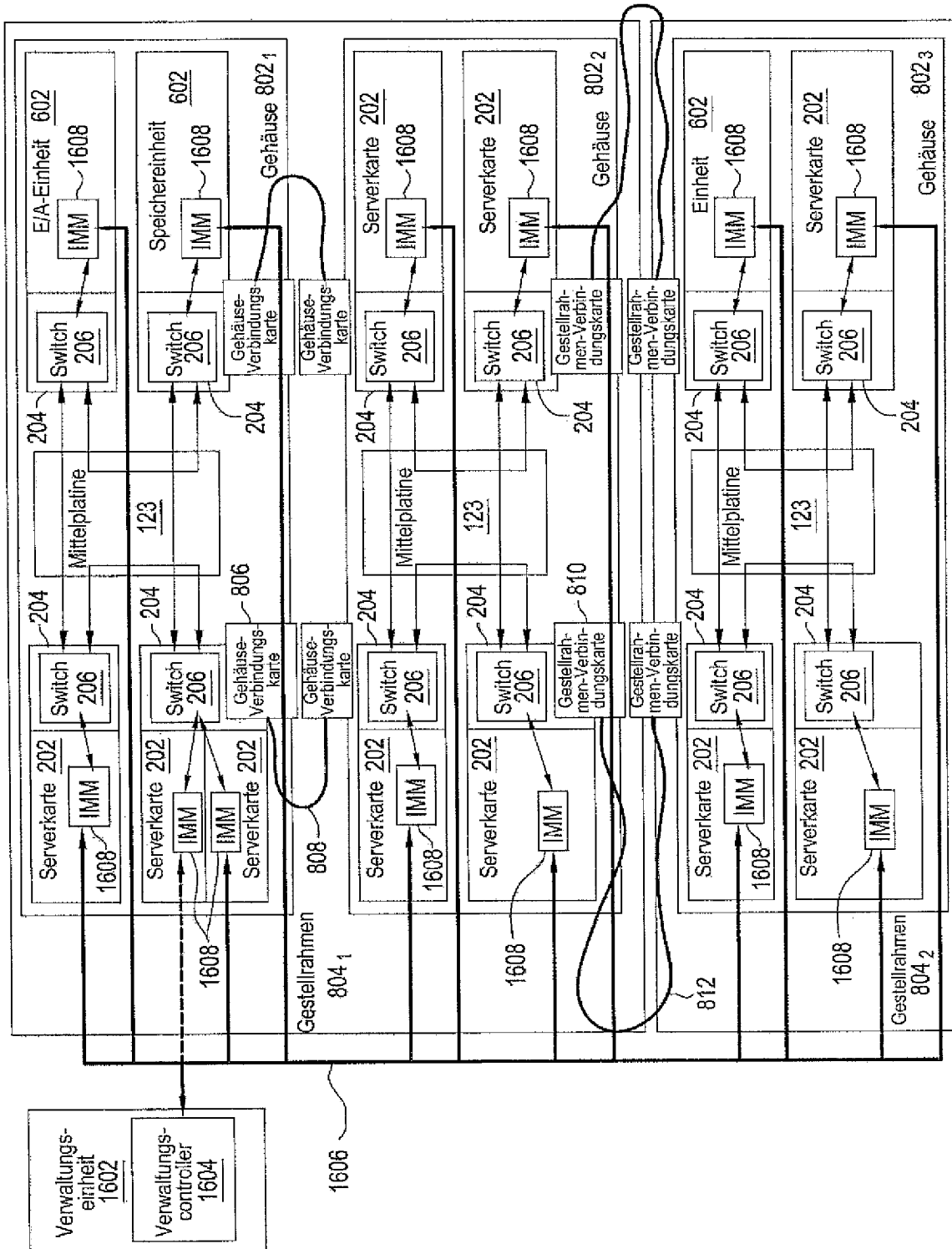
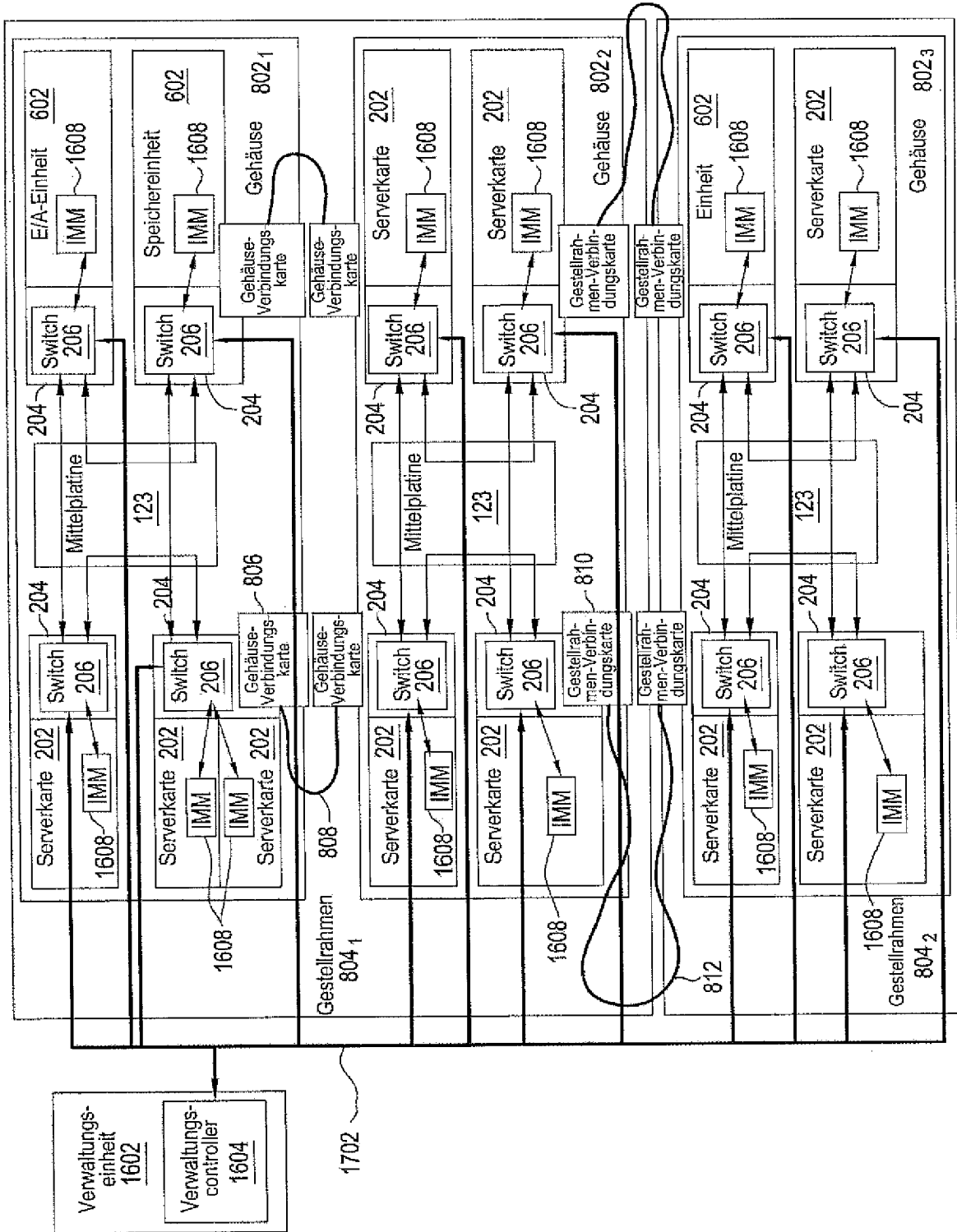


FIG. 15

1600
FIG. 16



1700
FIG. 17



1800

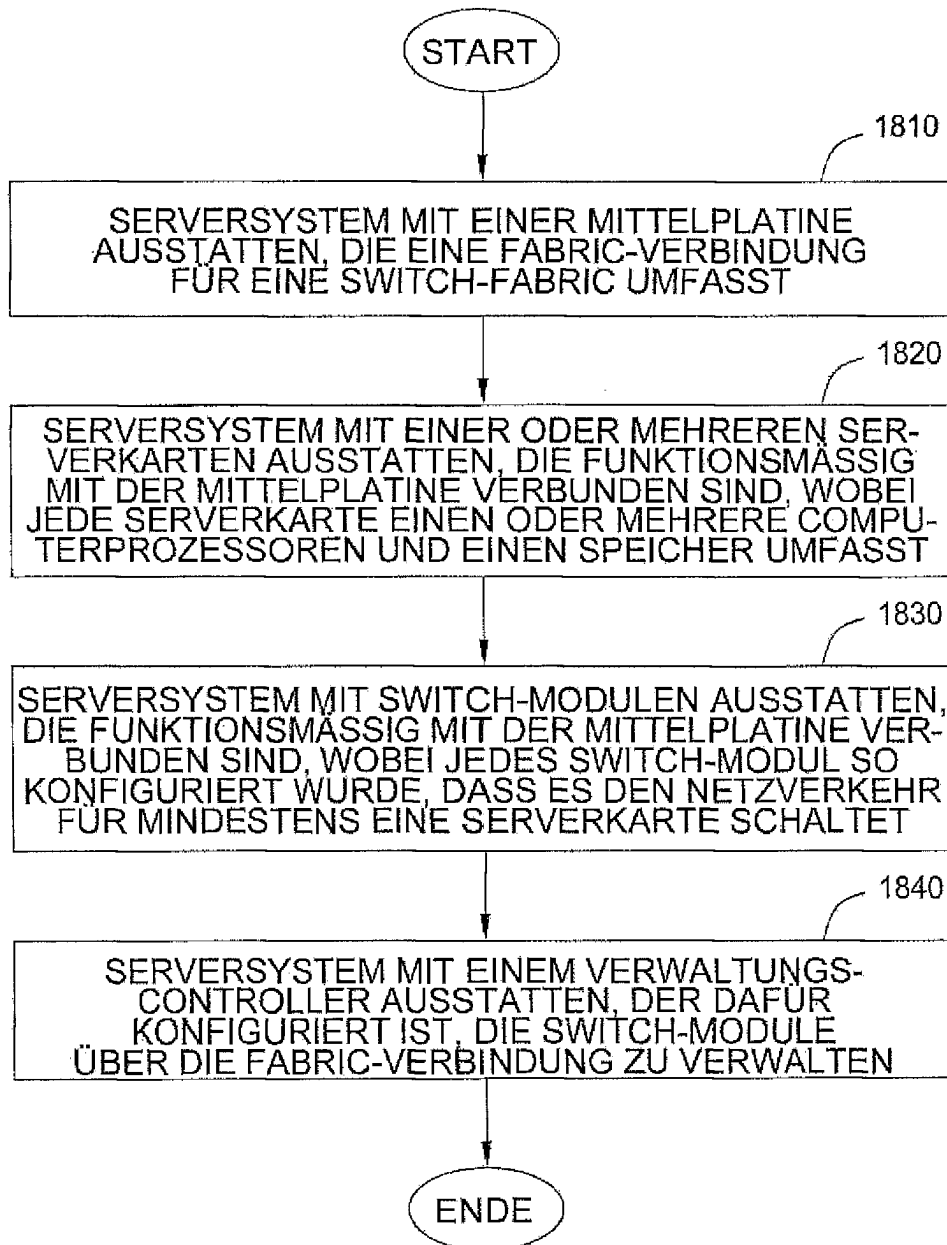


FIG. 18

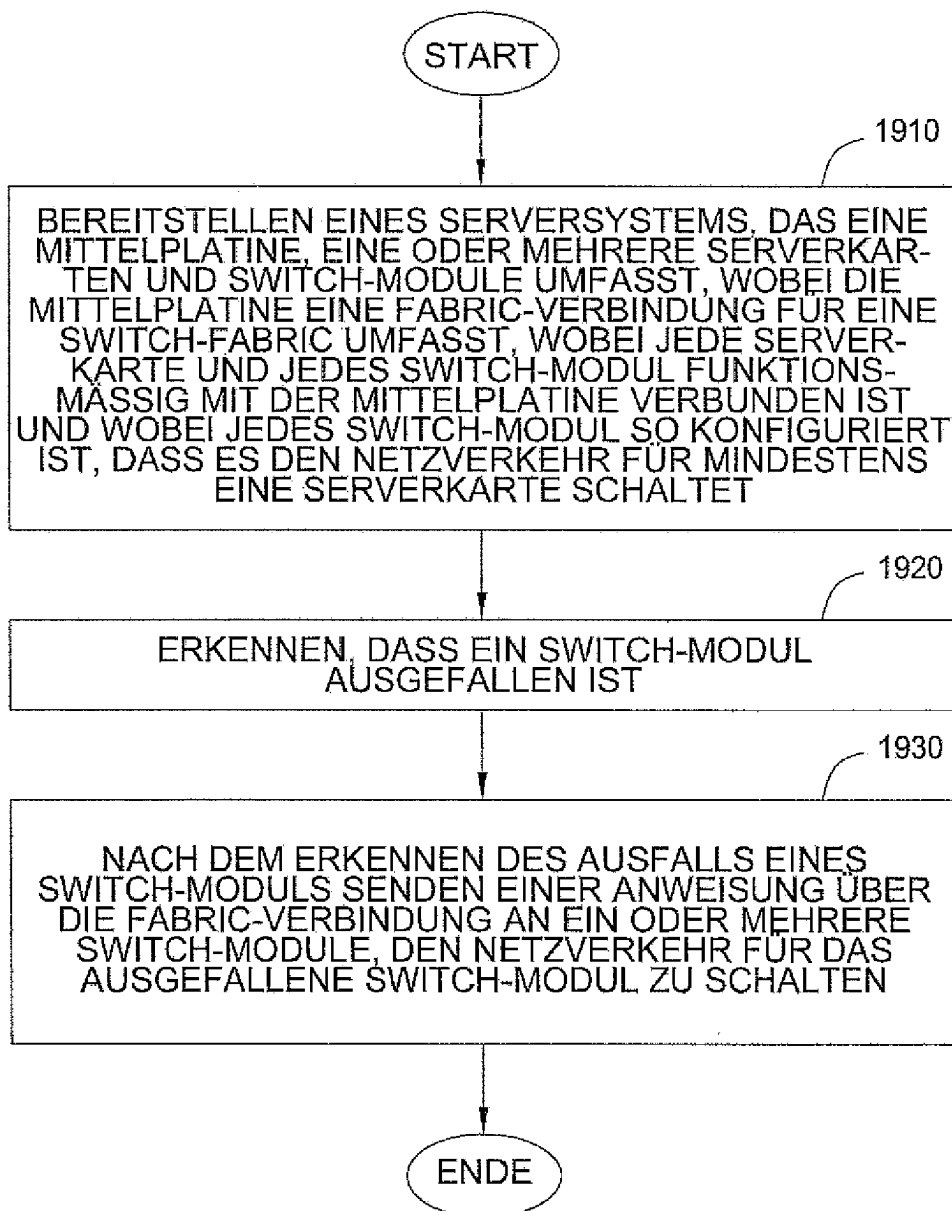


FIG. 19