



(12)

Veröffentlichung

der internationalen Anmeldung mit der

(87) Veröffentlichungs-Nr.: **WO 2011/110571**

in deutscher Übersetzung (Art. III § 8 Abs. 2 IntPatÜG)

(21) Deutsches Aktenzeichen: **11 2011 100 166.5**

(86) PCT-Aktenzeichen: **PCT/EP2011/053489**

(86) PCT-Anmeldetag: **08.03.2011**

(87) PCT-Veröffentlichungstag: **15.09.2011**

(43) Veröffentlichungstag der PCT Anmeldung
in deutscher Übersetzung: **31.10.2012**

(51) Int Cl.: **G06F 11/14 (2012.01)**

(30) Unionspriorität:

12/722,809 12.03.2010 US

(72) Erfinder:

Betzler, Boas, 71032, Böblingen, DE; Pierre-Louis, Marc-Arthur, Charlotte, N.C., US; Etkins, Robert Jonathan, Austin, Tex., US; Macho, Holger Jürgen, 71032, Böblingen, DE

(71) Anmelder:

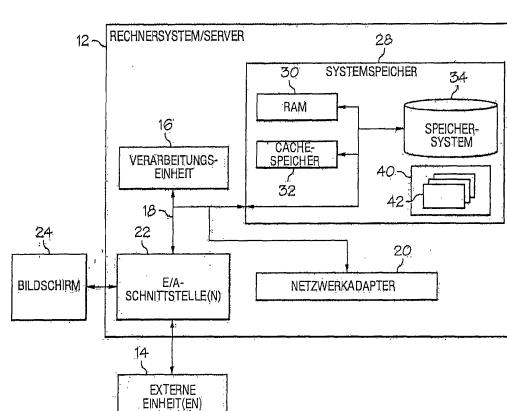
**International Business Machines Corporation,
Armonk, N.Y., US**

(74) Vertreter:

**RICHARDT PATENTANWÄLTE GbR, 65185,
Wiesbaden, DE**

(54) Bezeichnung: **Starten von virtuellen Instanzen in einer Cloud-Computing-Umgebung**

(57) Zusammenfassung: Ein System, welches das beziehungsweise die Operational Support System(s) (OSS) und das beziehungsweise die Business Support System(s) einer (z. B. öffentlichen) Rechnerwolke mit einem Dienst zur Automatisierung von Neustarts von virtuellen Instanzen vorteilhaft nutzt. Eine ausgefallene virtuelle Instanz wird in der Cloud-Computing-Umgebung erkannt und als Reaktion darauf wird eine Anforderung für eine neue virtuelle Instanz empfangen. Nach dem Empfang der Anforderung wird eine Berechtigung eines Benutzers, der der ausgefallenen virtuellen Instanz zugeordnet ist, geprüft. Ein Satz von Identitätsnachweis-Forderungen und -prüfungen wird verwendet, um die Unverfälschtheit der Anforderungen sowie die Berechtigung des Anforderers zur Nutzung der Ressourcen sicherzustellen. Unter der Annahme, dass die Prüfung bestanden wird, wird ein zu der ausgefallenen virtuellen Instanz gehörender Abwärtszähler verringert. Wenn der Abwärtszähler einen vorher festgelegten Schwellwert (z. B. null) erreicht, wird die neue virtuelle Instanz gestartet, die ausgefallene virtuelle Instanz wird beendet und der Abwärtszähler wird von der Instanz zurückgesetzt/neu gestartet.



Beschreibung**TECHNISCHES GEBIET**

[0001] Die vorliegende Erfindung betrifft allgemein die Datenverarbeitung in einer Rechnerwolke (Cloud-Computing). Insbesondere betrifft die vorliegende Erfindung das Starten von virtuellen Instanzen in einer Cloud-Computing-Umgebung.

HINTERGRUND

[0002] Die Cloud-Computing-Umgebung ist eine Verbesserung der dieser Umgebung vorausgehenden Grid-Umgebung, wobei mehrere Grids und andere Rechenressourcen von einer Cloud-Ebene noch weiter zusammengefasst werden können, so dass ganz verschiedene Einheiten einem Endbenutzer als ein einziges Reservoir von nahtlos ineinander übergehender Ressourcen erscheinen. Zu diesen Ressourcen können solche Dinge wie physische oder logische Compute Engines, Server und Einheiten, der Hauptspeicher von Einheiten, Speichereinheiten, Netzwerke, Unternehmensanwendungen und andere Software und dergleichen gehören.

[0003] In einer typischen Cloud-Umgebung haben Kunden entsprechende virtuelle "Kunden"-Instanzen, über die die gewünschte Funktionalität erreicht wird. Gelegentlich fallen diese Instanzen aus (z. B. aufgrund von Beschädigungen, Angriffen usw.) Bei den vorhandenen Lösungen ist es erforderlich, dass mehrere Systeminstanzen einander überwachen. Wenn sich die mehreren Instanzen auf demselben physischen Knoten befinden, würden sie im Falle einer Funktionsstörung des physischen Knotens alle unverfügbar werden. Außerdem macht die virtualisierte Ansicht bei vielen Cloud-Systemen die physische Zuordnung nicht sichtbar. Überdies gehen vorhandene Lösungen von Natur aus von einem gewissen Grad an Vertrauen zwischen den Komponenten aus, was in einer öffentlichen Cloud-Umgebung nicht vorausgesetzt werden kann. Zu den Nutzungsmöglichkeiten gehören die unbefugte Inanspruchnahme des zur Verfügung stehenden Dienstes, um: Sabotageangriffe zu starten; auf vertrauliche Daten eines anderen Cloud-Benutzers zuzugreifen; und/oder nicht genehmigte Kosten zu verursachen.

KURZDARSTELLUNG

[0004] Die vorliegende Erfindung stellt in einer ersten Erscheinungsform folglich ein Verfahren bereit, um eine virtuelle Instanz in einer Cloud-Computing-Umgebung zu starten, wobei das Verfahren Folgendes umfasst: Erkennen einer ausgefallenen virtuellen Instanz in der Cloud-Computing-Umgebung; als Reaktion auf die ausgefallene virtuelle Instanz Empfangen einer Anforderung für eine neue virtuelle Instanz; Prüfen einer Berechtigung eines Benutzers, welcher

der ausgefallenen virtuellen Instanz zugeordnet ist; Verringern eines Abwärtszählers, der zu der ausgefallenen virtuellen Instanz gehört; Starten der neuen virtuellen Instanz, wenn der Abwärtszähler einen vorher festgelegten Schwellwert erreicht; und Beenden der ausgefallenen virtuellen Instanz.

[0005] Vorzugsweise umfasst die neue virtuelle Instanz eine virtuelle Maschine. Vorzugsweise umfasst der Prüfvorgang die Feststellung, ob der Benutzer ein Eigentümer der ausgefallenen virtuellen Instanz ist. Vorzugsweise umfasst der Prüfvorgang die Feststellung, ob der Benutzer ein Eigentümer eines zu der ausgefallenen virtuellen Instanz gehörenden Urlade-(Boot-)Abbilds ist. Vorzugsweise umfasst der Prüfvorgang die Feststellung, ob der Benutzer ein Eigentümer einer reservierten Internet-Protocol-(IP-)Adresse ist, die zu der ausgefallenen virtuellen Instanz gehört. Vorzugsweise umfasst der Prüfvorgang die Feststellung, ob der Benutzer ein Eigentümer des zu der ausgefallenen virtuellen Instanz gehörenden permanenten Speichers ist. Vorzugsweise umfasst der vorher festgelegte Schwellwert null. Das Verfahren kann darüber hinaus das Eintragen eines Urlade-Abbilds, einer reservierten IP-Adresse und eines permanenten Speichers der neuen virtuellen Instanz für den Benutzer umfassen. Vorzugsweise umfasst die neue virtuelle Instanz einen Neustart der ausgefallenen virtuellen Instanz.

[0006] In einer zweiten Erscheinungsform wird ein System bereitgestellt, um eine virtuelle Instanz in einer Cloud-Computing-Umgebung zu starten, wobei das System Folgendes umfasst: ein Speichermedium, das Befehle umfasst; einen mit dem Speichermedium verbundenen Bus; und einen mit dem Bus verbundenen Prozessor, der, wenn er die Befehle ausführt, bewirkt, dass das System: eine ausgefallene virtuelle Instanz in der Cloud-Computing-Umgebung erkennt; als Reaktion auf die ausgefallene virtuelle Instanz eine Anforderung für eine neue virtuelle Instanz empfängt; eine Berechtigung eines Benutzers, welcher der ausgefallenen virtuellen Instanz zugeordnet ist, prüft; einen Abwärtszähler, der zu der ausgefallenen virtuellen Instanz gehört, verringert; die neue virtuelle Instanz startet, wenn der Abwärtszähler einen vorher festgelegten Schwellwert erreicht; und die ausgefallene virtuelle Instanz beendet.

[0007] Vorzugsweise umfasst die neue virtuelle Instanz eine virtuelle Maschine. Vorzugsweise umfasst der Prüfvorgang die Feststellung, ob der Benutzer ein Eigentümer der ausgefallenen virtuellen Instanz ist. Vorzugsweise umfasst der Prüfvorgang die Feststellung, ob der Benutzer ein Eigentümer eines zu der ausgefallenen virtuellen Instanz gehörenden Urlade-Abbilds ist. Vorzugsweise umfasst der Prüfvorgang die Feststellung, ob der Benutzer ein Eigentümer einer reservierten Internet-Protocol-(IP-)Adresse ist, die zu der ausgefallenen virtuellen Instanz gehört.

Vorzugsweise umfasst der Prüfvorgang die Feststellung, ob der Benutzer ein Eigentümer des zu der aus gefallenen virtuellen Instanz gehörenden permanenten Speichers ist. Vorzugsweise umfasst der vorher festgelegte Schwellwert null. Das System kann darüber hinaus das Eintragen eines Urlade-Abbilds, einer reservierten IP-Adresse und eines permanenten Speichers der neuen virtuellen Instanz für den Benutzer umfassen. Vorzugsweise umfasst die neue virtuelle Instanz einen Neustart der ausgefallenen virtuellen Instanz.

[0008] In einer dritten Erscheinungsform wird ein Rechnerprogramm bereitgestellt, das einen auf einem rechnerlesbaren Datenträger gespeicherten Rechnerprogrammcode umfasst, um, wenn es in das Rechnersystem geladen und darauf ausgeführt wird, das Rechnersystem zu veranlassen, alle Schritte eines Verfahrens gemäß der ersten Erscheinungsform durchzuführen.

[0009] Das Rechnerprogramm kann in einem rechnerlesbaren Datenträger enthalten sein, der ein Programmprodukt enthält, um eine virtuelle Instanz in einer Cloud-Computing-Umgebung zu starten, wobei der rechnerlesbare Datenträger Programmcode umfasst, um einen Rechner zu Folgendem zu veranlassen: Erkennen einer ausgefallenen virtuellen Instanz in der Cloud-Computing-Umgebung; als Reaktion auf die ausgefallene virtuelle Instanz Empfangen einer Anforderung für eine neue virtuelle Instanz; Prüfen einer Berechtigung eines Benutzers, welcher der ausgefallenen virtuellen Instanz zugeordnet ist; Verringern eines Abwärtszählers, der zu der ausgefallenen virtuellen Instanz gehört; Starten der neuen virtuellen Instanz, wenn der Abwärtszähler einen vorher festgelegten Schwellwert erreicht; und Beenden der ausgefallenen virtuellen Instanz.

[0010] In einer vierten Erscheinungsform wird ein Verfahren zum Einsatz eines Systems bereitgestellt, um eine virtuelle Instanz in einer Cloud-Computing-Umgebung zu starten, wobei das Verfahren Folgendes umfasst: Bereitstellen einer Rechner-Infrastruktur, die zu Folgendem dient: Erkennen einer ausgefallenen virtuellen Instanz in der Cloud-Computing-Umgebung; als Reaktion auf die ausgefallene virtuelle Instanz Empfangen einer Anforderung für eine neue virtuelle Instanz; Prüfen einer Berechtigung eines Benutzers, welcher der ausgefallenen virtuellen Instanz zugeordnet ist; Verringern eines Abwärtszählers, der zu der ausgefallenen virtuellen Instanz gehört; Starten der neuen virtuellen Instanz, wenn der Abwärtszähler einen vorher festgelegten Schwellwert erreicht; und Beenden der ausgefallenen virtuellen Instanz.

[0011] Ausführungsformen der vorliegenden Erfindung stellen folglich ein System bereit, welches das beziehungsweise die Operational Support System(s)

(OSS)(Betriebsunterstützungssysteme) und das beziehungsweise die Business Support System(s) (Geschäftsprozess-Unterstützungssysteme) einer (z. B. öffentlichen) Rechnerwolke mit einem Dienst zur Automatisierung von Neustarts von virtuellen Instanzen vorteilhaft nutzt. Im Einzelnen wird bei Ausführungsformen der vorliegenden Erfindung eine ausgefallene virtuelle Instanz (z. B. eine virtuelle Maschine) in der Cloud-Computing-Umgebung erkannt und als Reaktion darauf wird eine Anforderung für eine neue virtuelle Instanz empfangen. Nach dem Empfang der Anforderung wird eine Berechtigung eines Benutzers, der der ausgefallenen virtuellen Instanz zugeordnet ist, geprüft. Insbesondere wird ein Satz von Identitätsnachweis-Forderungen und -prüfungen verwendet, um die Unverfälschtheit der Anforderungen sowie die Berechtigung des Anforderers zur Nutzung der Ressourcen sicherzustellen. Unter der Annahme, dass die Prüfung bestanden wird, wird ein zu der ausgefallenen virtuellen Instanz gehörender Abwärtszähler verringert. Wenn der Abwärtszähler einen vorher festgelegten Schwellwert (z. B. null) erreicht, wird die neue virtuelle Instanz gestartet, die ausgefallene virtuelle Instanz wird beendet und der Abwärtszähler wird von der Instanz zurückgesetzt/ neu gestartet. Eine Einbindung in ein BSS ermöglicht darüber hinaus die Überwachung der Nutzung von Diensten sowie die Erhebung von Gebühren für die Nutzung der Dienste sowie die Eröffnung der Option für die Erhebung von Gebühren für vertragsungebundene, nur bei Bedarf abgerufene und bezahlte Dienstleistungen (Pay-as-you-go-Gebühren) für den Dienst des Neustarts selbst. Das Initiieren einer neuen virtuellen Instanz kann Folgendes umfassen: einen Neustart der alten virtuellen Instanz; und/oder ein Starten einer ganz bestimmten und gesonderten virtuellen Instanz.

[0012] Eine erste gegebene Art der Ausführung der vorliegenden Erfindung stellt ein Verfahren bereit, um eine virtuelle Instanz in einer Cloud-Computing-Umgebung zu starten, wobei das Verfahren Folgendes umfasst: Erkennen einer ausgefallenen virtuellen Instanz in der Cloud-Computing-Umgebung; als Reaktion auf die ausgefallene virtuelle Instanz Empfangen einer Anforderung für eine neue virtuelle Instanz; Prüfen einer Berechtigung eines Benutzers, welcher der ausgefallenen virtuellen Instanz zugeordnet ist; Verringern eines Abwärtszählers, der zu der ausgefallenen virtuellen Instanz gehört; Starten der neuen virtuellen Instanz, wenn der Abwärtszähler einen vorher festgelegten Schwellwert erreicht; und Beenden der ausgefallenen virtuellen Instanz.

[0013] Eine zweite gegebene Art der Ausführung der vorliegenden Erfindung stellt ein System bereit, um eine virtuelle Instanz in einer Cloud-Computing-Umgebung zu starten, wobei das System Folgendes umfasst: ein Speichermedium, das Befehle umfasst; einen mit dem Speichermedium verbundenen Bus;

und einen mit dem Bus verbundenen Prozessor, der, wenn er die Befehle ausführt, bewirkt, dass das System: eine ausgefallene virtuelle Instanz in der Cloud-Computing-Umgebung erkennt; als Reaktion auf die ausgefallene virtuelle Instanz eine Anforderung für eine neue virtuelle Instanz empfängt; eine Berechtigung eines Benutzers, welcher der ausgefallenen virtuellen Instanz zugeordnet ist, prüft; einen Abwärtszähler, der zu der ausgefallenen virtuellen Instanz gehört, verringert; die neue virtuelle Instanz startet, wenn der Abwärtszähler einen vorher festgelegten Schwellwert erreicht; und die ausgefallene virtuelle Instanz beendet.

[0014] Eine dritte gegebene Art der Ausführung der vorliegenden Erfindung stellt einen rechnerlesbaren Datenträger bereit, der ein Programmprodukt enthält, um eine virtuelle Instanz in einer Cloud-Computing-Umgebung zu starten, wobei der rechnerlesbare Datenträger Programmcode umfasst, um einen Rechner zu Folgendem zu veranlassen: Erkennen einer ausgefallenen virtuellen Instanz in der Cloud-Computing-Umgebung; als Reaktion auf die ausgefallene virtuelle Instanz Empfangen einer Anforderung für eine neue virtuelle Instanz; Prüfen einer Berechtigung eines Benutzers, welcher der ausgefallenen virtuellen Instanz zugeordnet ist; Verringern eines Abwärtszählers, der zu der ausgefallenen virtuellen Instanz gehört; Starten der neuen virtuellen Instanz, wenn der Abwärtszähler einen vorher festgelegten Schwellwert erreicht; und Beenden der ausgefallenen virtuellen Instanz.

[0015] Eine vierte gegebene Art der Ausführung der vorliegenden Erfindung stellt ein Verfahren zum Einsatz eines Systems bereit, um eine virtuelle Instanz in einer Cloud-Computing-Umgebung zu starten, wobei das Verfahren Folgendes umfasst: Bereitstellen einer Rechner-Infrastruktur, die zu Folgendem dient: Erkennen einer ausgefallenen virtuellen Instanz in der Cloud-Computing-Umgebung; als Reaktion auf die ausgefallene virtuelle Instanz Empfangen einer Anforderung für eine neue virtuelle Instanz; Prüfen einer Berechtigung eines Benutzers, welcher der ausgefallenen virtuellen Instanz zugeordnet ist; Verringern eines Abwärtszählers, der zu der ausgefallenen virtuellen Instanz gehört; Starten der neuen virtuellen Instanz, wenn der Abwärtszähler einen vorher festgelegten Schwellwert erreicht; und Beenden der ausgefallenen virtuellen Instanz.

KURZE BESCHREIBUNG DER ZEICHNUNGEN

[0016] Eine bevorzugte Ausführungsform der vorliegenden Erfindung wird nun lediglich anhand eines Beispiels und mit Bezug auf die beiliegenden Zeichnungen beschrieben, bei denen:

[0017] [Fig. 1](#) einen Cloud-Computing-Knoten gemäß einer Ausführungsform der vorliegenden Erfindung zeigt;

[0018] [Fig. 2](#) eine Cloud-Computing-Umgebung gemäß einer Ausführungsform der vorliegenden Erfindung zeigt;

[0019] [Fig. 3](#) Ebenen eines Cloud-Abstraktionsmodells gemäß einer Ausführungsform der vorliegenden Erfindung zeigt;

[0020] [Fig. 4](#) ein System zum automatisierten Starten einer virtuellen Instanz in einer Cloud-Computing-Umgebung gemäß einer Ausführungsform der vorliegenden Erfindung zeigt;

[0021] [Fig. 5](#) ein Flussdiagramm eines Verfahrens gemäß einer Ausführungsform der vorliegenden Erfindung zeigt.

[0022] Die Zeichnungen sind nicht unbedingt maßstabsgetreu. Bei den Zeichnungen handelt es sich lediglich um schematische Darstellungen, die keine bestimmten Parameter der Erfindung darstellen sollen. Die Zeichnungen sollen nur typische Ausführungsformen der Erfindung zeigen und sind daher nicht als Einschränkung des Umfangs der Erfindung zu verstehen. In den Zeichnungen stellen gleiche Bezugszahlen gleiche Elemente dar.

AUSFÜHRLICHE BESCHREIBUNG

[0023] Der Zweckmäßigkeit halber ist die ausführliche Beschreibung in die folgenden Abschnitte gegliedert:

- I. Begriffsbestimmungen des Cloud-Computing
- II. Ausführliche Beschreibung der Realisierung von Ausführungsformen der Erfindung
- I. Begriffsbestimmungen des Cloud-Computing

[0024] Die folgenden Begriffsbestimmungen wurden aus "Draft NIST Working Definition of Cloud Computing" von Peter Mell und Tim Grance vom 7. Oktober 2009 hergeleitet.

[0025] Cloud-Computing ist ein Modell, um einen mühelosen, bedarfsweisen Zugriff über das Netzwerk auf einen gemeinsam genutzten Bestand an konfigurierbaren Datenverarbeitungsressourcen (z. B. Netzwerken, Servern, Speichern, Anwendungen und Diensten) zu ermöglichen, die schnell bereitgestellt und mit äußerst geringem Verwaltungsaufwand oder Eingriff seitens des Dienstanbieters freigegeben werden können. Dieses Cloud-Modell unterstützt die Verfügbarkeit und umfasst mindestens fünf Merkmale, mindestens drei Dienst-Modelle und mindestens vier Einsatzmodelle.

[0026] Die Merkmale sind wie folgt:

Selbstbedienung nach Bedarf: Ein Kunde kann einseitig nach Bedarf Datenverarbeitungsfunktionen wie zum Beispiel Serverzeit und Netzwerkspeicher automatisch, ohne dass eine Person bei dem Anbieter eines jeden Dienstes eingreifen muss, bereitstellen.

[0027] Umfassender Netzwerkzugriff: Funktionen stehen über ein Netzwerk zur Verfügung und der Zugriff auf die Funktionen erfolgt über standardmäßige Mechanismen, welche die Nutzung durch heterogene Thin-Client oder Thick-Client-Plattformen (z. B. Mobiltelefone, Laptops und persönliche digitale Assistenten (PDAs) unterstützen.

[0028] Gemeinsame Nutzung physischer Ressourcen: Die Datenverarbeitungsressourcen des Diensteanbieters werden unter Verwendung eines Modells mit mehreren Mandanten (Multi-Tenant-Modell) zusammengeführt, um mehrere Kunden zu bedienen, wobei entsprechend dem Bedarf des Kunden verschiedene physische und virtuelle Ressourcen dynamisch zugewiesen und neu zugewiesen werden. Man hat ein Gefühl der Standort-Unabhängigkeit, da der Kunde im Allgemeinen keine Kontrolle über den genauen Standort der bereitgestellten Ressourcen hat oder den genauen Standort nicht kennt, diesen gegebenenfalls aber auf einer höheren Abstraktionsebene (z. B. Land, Staat oder Rechenzentrum) angeben kann. Zu Beispielen für Ressourcen gehören Plattspeicher, Rechenleistung, Hauptspeicher, Netzwerkbandbreite und virtuelle Maschinen.

[0029] Unverzügliche Anpassbarkeit an den Ressourcenbedarf:

Funktionen können schnell und elastisch bereitgestellt werden, in manchen Fällen automatisch, um eine schnelle horizontale Skalierbarkeit zu ermöglichen, und sie können schnell freigegeben werden, um eine schnelle vertikale Skalierbarkeit zu ermöglichen. Dem Kunden erscheinen die zur Versorgung vorhandenen Funktionen oftmals unerschöpfig und sie können jederzeit in jeder beliebigen Menge erworben werden.

[0030] Messung der Service-Nutzung: Cloud-Systeme steuern und optimieren die Nutzung der Ressourcen automatisch, indem sie auf einer bestimmten Abstraktionsebene, die für die Art des Dienstes (z. B. Speicherkapazität, Rechenleistung, Bandbreite und aktive Benutzerkonten) geeignet ist, eine Zählerfunktion vorteilhaft einsetzen. Die Nutzung der Ressourcen kann überwacht, gesteuert und gemeldet werden, was Transparenz sowohl für den Anbieter als auch den Kunden des in Anspruch genommenen Dienstes schafft.

[0031] Es gibt folgende Dienst-Modelle:

Cloud Software as a Service (SaaS): Die dem Kunden zur Verfügung gestellte Funktion besteht in der Nutzung der Anwendungen des Anbieters, die in ei-

ner Cloud-Infrastruktur ausgeführt werden. Auf die Anwendungen kann von verschiedenen Client-Einheiten über eine Thin-Client-Schnittstelle wie zum Beispiel einen Webbrower (z. B. webbasierte E-Mail) zugegriffen werden. Der Kunde verwaltet oder steuert die zugrunde liegende Cloud-Infrastruktur einschließlich des Netzwerks, der Server, der Betriebssysteme, des Speichers oder selbst einzelner Anwendungsfunktionen nicht, wobei es die mögliche Ausnahme in Form von begrenzten benutzerspezifischen Anwendungs-Konfigurationseinstellungen gibt.

[0032] Cloud Platform as a Service (PaaS): Die dem Kunden zur Verfügung gestellte Funktion besteht darin, auf der Cloud-Infrastruktur vom Kunden erzeugte oder erworbene Anwendungen aufzusetzen, die mittels Programmiersprachen und Werkzeugen, die von dem Anbieter unterstützt werden, erzeugt wurden. Der Kunde verwaltet oder steuert die zugrunde liegende Cloud-Infrastruktur einschließlich des Netzwerks, der Server, der Betriebssysteme oder des Speichers nicht, hat aber die Kontrolle über die eingesetzten Anwendungen und möglicherweise über Konfigurationen der Application-Hosting-Umgebung.

[0033] Cloud Infrastructure as a Service (IaaS): Die dem Kunden zur Verfügung gestellte Funktion besteht in der Bereitstellung von Verarbeitungs-, Speicherressourcen, Netzwerken und anderen grundlegenden Datenverarbeitungsressourcen, auf denen der Kunde beliebige Software-Programme, die Betriebssysteme und Anwendungen einschließen können, installieren und ausführen kann. Der Kunde verwaltet oder steuert die zugrunde liegende Cloud-Infrastruktur nicht, hat aber die Kontrolle über Betriebssysteme, Speicher, eingesetzte Anwendungen und möglicherweise die eingeschränkte Kontrolle über ausgewählte Komponenten für den Netzwerkbetrieb (z. B. Host-Zugangsschutzsysteme).

[0034] Die Einsatzmodelle (Deployment Models) sind wie folgt:

Private Cloud: Die Cloud-Infrastruktur wird ausschließlich für eine Organisation betrieben. Sie kann von der Organisation oder einem Dritten verwaltet werden und kann sich auf dem Betriebsgelände oder außerhalb des Betriebsgeländes befinden.

[0035] Gemeinschafts-Cloud: Die Cloud-Infrastruktur wird von mehreren Organisationen gemeinsam genutzt und unterstützt eine bestimmte Gemeinschaft, die gemeinsame Interessen (z. B.

[0036] Aufgaben, Sicherheitsanforderungen, Richtlinien und Anforderungen an die Einhaltung der Richtlinien (Compliance-Anforderungen)) hat. Sie kann von der Organisation oder einem Dritten verwaltet werden und kann sich auf dem Betriebsgelände oder außerhalb des Betriebsgeländes befinden.

[0037] Öffentliche Cloud: Die Cloud-Infrastruktur wird der allgemeinen Öffentlichkeit oder einer großen Industriegruppe zur Verfügung gestellt und gehört einer Organisation, die Cloud-Dienste verkauft.

[0038] Hybride Cloud: Die Cloud-Infrastruktur setzt sich aus zwei oder mehr Clouds (privaten, Gemeinschafts- oder öffentlichen Clouds) zusammen, die individuelle Entitäten bleiben, welche jedoch durch eine standardisierte oder firmeneigene Technologie miteinander verbunden sind, die die Portierbarkeit von Daten und Anwendungen ermöglicht (z. B. den kombinierten Einsatz von privaten und öffentlichen Clouds (Cloud-Bursting) für einen Lastausgleich zwischen den Clouds).

[0039] Eine Cloud-Computing-Umgebung ist dienstorientiert, wobei der Schwerpunkt auf Zustandslosigkeit, loser Kopplung, Modularität und semantischer Interoperabilität liegt.

II. Ausführliche Beschreibung der Realisierung von Ausführungsformen der Erfindung

[0040] Ausführungsformen der vorliegenden Erfindung stellen ein System bereit, welches das beziehungsweise die Operational Support System(s) (OSS) und das beziehungsweise die Business Support System(s) einer (z. B. öffentlichen) Rechnerwolke mit einem Dienst zur Automatisierung von Neustarts von virtuellen Instanzen vorteilhaft nutzt. Im Einzelnen wird bei Ausführungsformen der vorliegenden Erfindung eine ausgefallene virtuelle Instanz (z. B. eine virtuelle Maschine) in der Cloud-Computing-Umgebung erkannt und als Reaktion darauf wird eine Anforderung für eine neue virtuelle Instanz empfangen. Nach dem Empfang der Anforderung wird eine Berechtigung eines Benutzers, der der ausgefallenen virtuellen Instanz zugeordnet ist, geprüft. Insbesondere wird ein Satz von Identitätsnachweis-Forderungen und -prüfungen verwendet, um die Unverfälschtheit der Anforderungen sowie die Berechtigung des Anforderers zur Nutzung der Ressourcen sicherzustellen. Unter der Annahme, dass die Prüfung bestanden wird, wird ein zu der ausgefallenen virtuellen Instanz gehörender Abwärtszähler verringert. Wenn der Abwärtszähler einen vorher festgelegten Schwellwert (z. B. null) erreicht, wird die neue virtuelle Instanz gestartet, die ausgefallene virtuelle Instanz wird beendet und der Abwärtszähler wird von der Instanz zurückgesetzt/neu gestartet. Eine Einbindung in ein BSS ermöglicht darüber hinaus die Überwachung der Nutzung von Diensten sowie die Erhebung von Gebühren für die Nutzung der Dienste sowie die Eröffnung der Option für die Erhebung von Gebühren für vertragsungebundene, nur bei Bedarf abgerufene und bezahlte Dienstleistungen (Pay-as-you-go-Gebühren) für den Dienst des Neustarts selbst. Das Initiiieren einer neuen virtuellen Instanz kann Folgendes umfassen: einen Neustart der alten virtuel-

len Instanz; und/oder ein Starten einer ganz bestimmten und gesonderten virtuellen Instanz. Es versteht sich bei einer typischen Ausführungsform im Voraus, dass sich "virtuelle Instanz" auf eine virtuelle Maschine (Virtual Machine (VM)), die zu einem Cloud-Kunden gehört, bezieht.

[0041] Nun Bezug nehmend auf [Fig. 1](#), ist eine schematische Darstellung eines beispielhaften Cloud-Datenverarbeitungsknotens gezeigt. Der Cloud-Datenverarbeitungsknoten **10** stellt lediglich ein Beispiel für einen geeigneten Cloud-Datenverarbeitungsknoten dar und ist nicht als Einschränkung des Nutzungs umfangs oder der Funktionalität der hier beschriebenen Erfindung zu verstehen. Ungeachtet dessen kann der Cloud-Datenverarbeitungsknoten **10** realisiert werden und/oder ist in der Lage, beliebige der im vorstehenden Abschnitt I beschriebenen Funktionen ausführen.

[0042] Im Cloud-Datenverarbeitungsknoten **10** gibt es ein Rechnersystem/einen Server **12**, das beziehungsweise der mit zahlreichen anderen Umgebungen oder Konfigurationen eines Datenverarbeitungssystems für allgemeine Anwendungen oder für spezielle Anwendungen betrieben werden kann. Zu Beispielen für bekannte Datenverarbeitungssysteme, – umgebungen und/oder – konfigurationen, die für die Verwendung mit dem Rechnersystem/Server **12** geeignet sein können, gehören, ohne auf diese beschränkt zu sein, Personal-Computer-Systeme, Server-Rechner-Systeme, Thin-Clients, Thick-Clients, tragbare oder Laptop-Einheiten, Mehrprozessorsysteme, auf einem Mikroprozessor beruhende Systeme, Aufsatzgeräte (Set-Top-Boxen), programmierbare Kundenelektronik, Netzwerk-PCs, Minirechnersysteme, Großrechnersysteme und verteilte Cloud-Computing-Umgebungen, die beliebige der vorstehenden Systeme oder Einheiten enthalten, und dergleichen.

[0043] Das Rechnersystem/der Server **12** kann in dem allgemeinen Kontext von Befehlen, die von einem Rechnersystem ausgeführt werden können, wie zum Beispiel Programmmodulen, die von einem Rechnersystem ausgeführt werden, beschrieben werden. Im Allgemeinen enthalten Programm module Routinen, Programme, Objekte, Komponenten, Logik, Datenstrukturen und so weiter, die bestimmte Aufgaben ausführen oder bestimmte abstrakte Datentypen realisieren. Das beispielhafte Rechnersystem/der beispielhafte Server **12** kann in verteilten Cloud-Computing-Umgebungen in die Praxis umgesetzt werden, in denen Aufgaben von fernen Verarbeitungseinheiten ausgeführt werden, welche über ein Datenübertragungsnetzwerk miteinander verbunden sind. In einer verteilten Cloud-Computing-Umgebung (CCU) können sich Programmmodul le sowohl in Speichermedien, einschließlich Haupt-

speicher-Einheiten, eines lokalen als auch eines fernen Rechnersystems befinden.

[0044] Wie in [Fig. 1](#) gezeigt ist, ist das Rechnersystem/der Server **12** in dem Cloud-Computing-Knoten **10** in Form von einer Datenverarbeitungseinheit für allgemeine Anwendungen gezeigt. Zu den Komponenten des Rechnersystems/des Servers **12** können, ohne auf diese beschränkt zu sein, ein oder mehrere Prozessoren oder Verarbeitungseinheiten **16**, ein Systemspeicher **28** und ein Bus **18** gehören, der verschiedene Systemkomponenten einschließlich des Systemspeichers **28** mit dem Prozessor **16** verbindet.

[0045] Der Bus **18** stellt eine oder mehrere von diversen beliebigen Arten von Busstrukturen einschließlich eines Hauptspeicher-Busses oder einer Hauptspeicher-Steuereinheit, eines peripheren Busses, eines Accelerated Graphics Port und eines Prozessor- oder lokalen Busses, der eine beliebige einer Vielzahl von Busarchitekturen nutzt, dar. Als Beispiel, aber nicht darauf beschränkt, gehören zu solchen Architekturen der Bus "Industry Standard Architecture (ISA)", der Bus "Micro Channel Architecture (MCA)", der Bus "Enhanced ISA (EISA)", der lokale Bus "Video Electronics Standards Association (VESA)" und der Bus "Peripheral Component Interconnects (PCI)".

[0046] Das Rechnersystem/der Server **12** enthält üblicherweise eine Vielzahl von Datenträgern, die von einem Rechnersystem gelesen werden können. Solche Datenträger können jedwede verfügbaren Datenträger sein, auf die von einem Rechnersystem/Server **12** zugegriffen werden kann, und sie beinhalten sowohl flüchtige als auch nicht flüchtige Datenträger sowie auswechselbare und nicht auswechselbare Datenträger.

[0047] Der Systemspeicher **28** kann von einem Rechnersystem lesbare Datenträger in Form von flüchtigem Speicher, wie beispielsweise einem Direktzugriffspeicher (RAM) **30** und/oder einem Cachespeicher **32**, beinhalten. Das Rechnersystem/der Server **12** kann darüber hinaus weitere auswechselbare/nicht auswechselbare und flüchtige/nicht flüchtige Speichermedien eines Rechnersystems enthalten. Lediglich als Beispiel kann das Speichersystem **34** für Leseoperationen von und für Schreiboperationen auf einen nicht auswechselbaren, nicht flüchtigen Magnetdatenträger (nicht gezeigt und üblicherweise als "Festplattenlaufwerk" bezeichnet) bereitgestellt werden. Obgleich nicht gezeigt, können ein Magnetplattenlaufwerk für Leseoperationen von und Schreiboperationen auf eine auswechselbare, nicht flüchtige Magnetplatte (z. B. eine "Diskette") und ein optisches Plattenlaufwerk für Leseoperationen von oder Schreiboperationen auf eine auswechselbare, nicht flüchtige optische Platte, wie zum Beispiel eine CD-ROM, eine DVD-ROM, oder andere optische Datenträger bereitgestellt werden. In diesen Fällen kann

jedes Plattenlaufwerk über eine oder mehrere Datenträger-Schnittstellen mit dem Bus **18** verbunden werden. Wie weiter gezeigt und nachstehend beschrieben wird, kann der Systemspeicher **28** mindestens ein Programmprodukt enthalten, das über einen Satz (z. B. mindestens einen Satz) von Programmmodulen verfügt, die so konfiguriert sind, dass sie die Funktionen der Erfindung ausführen.

[0048] Das Programm/Dienstprogramm **40**, das über einen Satz (mindestens einen Satz) von Programmmodulen **42** verfügt, kann als Beispiel, ohne darauf beschränkt zu sein, im Systemspeicher **28** abgelegt werden, ebenso ein Betriebssystem, ein oder mehrere Anwendungsprogramme, andere Programmmoduln und Programmdaten. Das Betriebssystem, ein oder mehrere Anwendungsprogramme, andere Programmmoduln und Programmdaten oder eine Kombination daraus können jeweils eine Ausführungsart einer Netzwerkumgebung haben. Die Programmmoduln **42** führen im Allgemeinen die Funktionen und/oder die methodischen Vorgehensweisen der hier beschriebenen Erfindung aus.

[0049] Das Rechnersystem/der Server **12** kann auch mit einer oder mehreren externen Einheiten **14** wie zum Beispiel einer Tastatur, einem Zeigegerät, einem Bildschirm **24** usw., mit einer oder mehreren Einheiten, die einem Benutzer den Dialogverkehr mit dem Rechnersystem/Server **12** ermöglichen, und/oder beliebigen Einheiten (z. B. Netzwerkarte, Modem usw.), die dem Rechnersystem/Server **12** den Datenaustausch mit einer oder mehreren anderen Datenverarbeitungseinheiten ermöglichen, Daten austauschen. Ein solcher Datenaustausch kann über die E/A-Schnittstellen **22** stattfinden. Weiterhin kann das Rechnersystem/der Server **12** mit einem oder mehreren Netzwerken wie zum Beispiel einem lokalen Netzwerk (LAN), einem allgemeinen Weitverkehrsnetz (WAN) und/oder einem öffentlichen Netzwerk (z. B. dem Internet) über den Netzwerkadapter **20** Daten austauschen. Wie gezeigt ist, tauscht der Netzwerkadapter **20** mit den anderen Komponenten des Rechnersystems/Servers **12** über den Bus **18** Daten aus. Es sollte sich verstehen, dass auch andere Hardware- und/oder Software-Komponenten in Verbindung mit dem Rechnersystem/Server **12** verwendet werden könnten, obgleich diese nicht gezeigt sind. Zu Beispielen hierzu gehören, ohne auf diese beschränkt zu sein: Mikrocode, Einheitentreiber, redundante Verarbeitungseinheiten, externe Anordnungen von Festplattenlaufwerken, RAID-Systeme, Bandlaufwerke und Speichersysteme zur Datenarchivierung usw.

[0050] Nun Bezug nehmend auf [Fig. 2](#), ist eine der Veranschaulichung dienende Cloud-Computing-Umgebung **50** dargestellt. Wie gezeigt ist, umfasst die Cloud-Computing-Umgebung **50** einen oder mehrere Cloud-Computing-Knoten **10**, mit dem beziehungs-

weise denen die Datenverarbeitungseinheiten wie zum Beispiel der persönliche digitale Assistent (PDA) oder das Mobiltelefon **54A**, der Tischrechner **54B**, der Laptop-Rechner **54C** und/oder das Fahrzeug-Rechnersystem **54N** Daten austauschen. Dies macht es möglich, aus der Cloud-Computing-Umgebung **50** eine Infrastruktur, Plattformen und/oder Software als Dienste anzubieten (wie vorstehend im Abschnitt I beschrieben wurde), damit nicht jeder Client solche Ressourcen gesondert unterhalten muss. Es versteht sich, dass die Arten der Datenverarbeitungseinheiten **54A** bis **54N**, die in [Fig. 2](#) gezeigt sind, lediglich dem Zweck der Veranschaulichung dienen und dass die Cloud-Computing-Umgebung **50** mit jeder beliebigen Art einer rechnergestützten Einheit über jede beliebige Art eines Netzwerks und/oder einer Netzwerk-/adressierbaren Verbindung (z. B. mittels eines Web-browsers) Daten austauschen kann.

[0051] Nun Bezug nehmend auf [Fig. 3](#) ist ein Satz von funktionalen Abstraktionsebenen gezeigt, die von der Cloud-Computing-Umgebung **50** ([Fig. 2](#)) bereitgestellt werden. Es sollte sich im Voraus verstehen, dass die Komponenten, Ebenen und Funktionen, die in [Fig. 3](#) gezeigt sind, lediglich dem Zweck der Veranschaulichung dienen und die Erfindung nicht darauf beschränkt ist. Wie dargestellt ist, werden die folgenden Ebenen und entsprechenden Funktionen bereitgestellt: Die Hardware- und Software-Ebene **60** enthält Hardware- und Software-Komponenten. Zu Beispielen für Hardware-Komponenten gehören Großrechner, in einem Beispiel Systeme der zSeries® von IBM®; Server, die auf der RISC-Architektur (RISC = Reduced Instruction Set Computer) beruhen, in einem Beispiel Systeme der pSeries® von IBM; Systeme der xSeries® von IBM; BladeCenter®-Systeme von IBM; Speichereinheiten; Netzwerke und Komponenten für den Netzwerkbetrieb. Zu Beispielen für Software-Komponenten gehören Server-Software für Netzwerk-Anwendungen. In einem Beispiel Software für WebSphere®-Anwendungsserver von IBM sowie Datenbank-Software. In einem Beispiel DB2®-Datenbank-Software von IBM. (IBM, zSeries, pSeries, xSeries, BladeCenter, WebSphere und DB2 sind Warenzeichen der International Business Machines Corporation in den Vereinigten Staaten von Amerika, anderen Ländern oder sowohl in den Vereinigten Staaten von Amerika als auch in anderen Ländern.)

[0052] Die Virtualisierungsebene **62** stellt eine Abstraktionsebene bereit, von der die folgenden beispielhaften virtuellen Entitäten bereitgestellt werden können: virtuelle Server; virtueller Speicher; virtuelle Netzwerke einschließlich virtueller privater Netzwerke; virtuelle Anwendungen; und virtuelle Clients.

[0053] Die Verwaltungsebene **64** stellt die nachstehend beschriebenen beispielhaften Funktionen zur Verfügung. Die Ressourcenbereitstellung ermög-

licht die dynamische Beschaffung von Datenverarbeitungsressourcen und anderen Ressourcen, die verwendet werden, um Aufgaben in der Cloud-Computing-Umgebung durchzuführen. Die Erfassung von Verbrauchswerten und die Preisermittlung ermöglichen die Überwachung von Kosten bei der Nutzung von Ressourcen in der Cloud-Computing-Umgebung sowie die Abrechnung oder Inrechnungstellung für die Inanspruchnahme dieser Ressourcen. In einem Beispiel können diese Ressourcen Lizenzen für Anwendungs-Software umfassen. Die Funktion "Sicherheit" ermöglicht die Überprüfung der Identität für Benutzer und Aufgaben sowie den Schutz für Daten und andere Ressourcen. Das Benutzerportal ermöglicht sowohl Benutzern als auch Systemadministratoren den Zugriff auf die Cloud-Computing-Umgebung. Die Verwaltung der Dienstgüte ermöglicht die Zuweisung und die Verwaltung von Ressourcen für das Cloud-Computing in der Weise, dass der erforderlichen Dienstgüte entsprochen wird. Die Planung und Erfüllung der Dienstgütevereinbarung (Service Level Agreement (SLA)) ermöglicht eine Vorab-Festlegung für und die Beschaffung von Ressourcen für das Cloud-Computing, bei denen ein zukünftiges Erfordernis gemäß einer SLA vorweggenommen wird.

[0054] Die Arbeitslasten-Ebene **66** stellt Funktionalität bereit, für die die Cloud-Computing-Umgebung genutzt wird. Zu Beispielen für Arbeitslasten und Funktionen, die von dieser Ebene bereitgestellt werden können, gehören: Abbildung und Navigation; Software-Entwicklung und Verwaltung des Lebenszyklus; Bereitstellung des Unterrichts im virtuellen Klassenzimmer; Verarbeitung zu datenanalytischen Zwecken, Transaktionsverarbeitung; und Starten von Instanzen. Wie vorstehend erwähnt wurde, dienen alle vorstehend mit Bezug auf [Fig. 3](#) beschriebenen Beispiele lediglich der Veranschaulichung, und die Erfindung ist nicht auf diese Beispiele beschränkt.

[0055] Es versteht sich, dass alle Funktionen der vorliegenden Erfindung, die hier beschrieben werden, üblicherweise durch das Starten von Instanzen ausgeführt werden, welche als Module des Programmcodes **42** des Programms/Dienstprogramms **40** ([Fig. 1](#)) physisch greifbar realisiert werden können.

[0056] Nun Bezug nehmend auf [Fig. 4](#) ist ein Diagramm der Architektur gemäß der vorliegenden Erfindung gezeigt. Es versteht sich, dass ein Teil der in den [Fig. 4](#) bis [Fig. 5](#) gezeigten Komponenten den Komponenten in den [Fig. 1](#) bis [Fig. 3](#) entsprechen. Die Cloud-Knoten **200** entsprechen beispielsweise den Cloud-Knoten **10**. In den [Fig. 4](#) bis [Fig. 5](#) wurden lediglich aus Gründen der Übersichtlichkeit der Darstellung und der leichteren Lesbarkeit gesonderte Bezugszahlen verwendet. Es versteht sich weiter, dass Ausführungsformen der vorliegenden Erfindung das beziehungsweise die Operational Support Sys-

tem(s) (OSS) und das beziehungsweise die Business Support System(s) (BSS) für die Cloud-Umgebung **50** ([Fig. 2](#)) vorteilhaft nutzen. Diese Systeme sind üblicherweise wie folgt definiert: Operational Support System: Das OSS befasst sich mit Datenverarbeitungs-, Netzwerk-, Speicher- und anderen Ressourcen und beinhaltet die Inventar-Verwaltung, die Bereitstellungs-Verwaltung, die Änderungs-Verwaltung und die Störungs-Verwaltung (unter anderem). Das OSS kann eine Vielzahl von Instanzen von virtuellen Maschinen auf einer Vielzahl von physischen Serversystemen, die einen Virtualisierungs-Hypervisor ausführen, bereitstellen. Die Bereitstellung kann des Weiteren die Auswahl eines Abbilds des anfänglichen Urladens für die Instanz der virtuellen Maschine und eine reservierte IP-Adresse, die der angeforderten Instanz zugeordnet werden soll, sowie die Zuweisung von permanenten Speichereinheiten zu bereitgestellten Instanzen beinhalten.

[0057] Business Support System: Das BSS, das mit dem OSS zusammenarbeitet, befasst sich mit Abonnements von Endbenutzern, der Auftragsabwicklung, der Erfassung von Verbrauchswerten und der Tarifierung sowie der Rückbuchung oder Bezahlung. Typische Arten von Aktivitäten, die von dem BSS durchgeführt werden können, sind (unter anderem) das Aufnehmen der Bestellung eines Kunden, die Konfiguration von Netzwerkkomponenten, die Rechnungslegung und das Fehlermanagement.

[0058] In jedem Fall tritt, wie in [Fig. 4](#) gezeigt ist, ein berechtigter Benutzer/Kunde **220** entweder mit einer Web-Benutzerschnittstelle (user interface ((UI)) **101** oder über einen Software-Vermittler unter Verwendung einer API **102** in Dialogverkehr, um eine Instanz einer virtuellen Maschine **204** auf dem Cloud-Knoten **200** (der in Verbindung mit der Hardware **202**, dem Hostrechner-System **202**, dem Hypervisor **203** und der temporären Platte **207** gezeigt ist) zu erzeugen. Diese Instanz hat die folgenden weiteren Merkmale (unter anderem): Sie wird von einem Hauptabbild **213** erzeugt, das dem berechtigten Benutzer **220** gehört und von diesem erstellt wird. Dadurch soll das unbefugte Starten von Abbildern mit möglicherweise persönlichen Daten vermieden werden. Der Instanz **204** ist eine reservierte Internet-Protocol-(IP-)Adresse **208** zugeordnet, die dem Benutzer **220** gehört. Optional wird in der Instanz eine permanente Platte als Dateisystem **214** eingesetzt. Die Instanz enthält auch eine Anwendung des Software-Stapels **210**, die so konfiguriert ist, dass sie automatisch startet und in Betrieb geht, wenn die Urlade-Festplattenabbilder **213** in einer Instanz **204** bereitgestellt werden.

[0059] Es gibt zwei Möglichkeiten, den automatisierten Neustart-Dienst **110** aufzurufen: Entweder trägt sich der berechtigte Benutzer **220**, nachdem er seine Identität bei der Web-UI **101** nachgewiesen hat, oder die API **102** bei dem automatisierten Neustart-Dienst

111 ein, der von jedem beliebigen Neustart-Initialisierungsdienst **221** im Netzwerk aufgerufen wird, und initialisiert den Rücklaufzeitgeber **222**. In diesem Fall stellt die Prüfung **114** zur Vermeidung von Angriffen sicher, dass der anfordernde Benutzer **220** identisch mit dem Benutzer ist, dem die Instanz **204** der virtuellen Maschine VM gehört, und dass er der Benutzer ist, der die reservierte IP-Adresse **208** angefordert hat.

[0060] Alternativ beinhaltet eine Initialisierung **211** des Neustarts in der Instanz **204** der virtuellen Maschine das Aufrufen der Registrierung **111**. In diesem Fall stellt die Prüfung **114** zur Vermeidung von Angriffen sicher, dass die IP-Ursprungsadresse des Aufrufers die reservierte IP-Adresse **208** ist und dass ein Zertifikat, das die Identität des Erstellers **220** der VM-Instanz **204** nachweist, gültig ist.

[0061] In beiden Fällen wird bei der Prüfung **114** zur Vermeidung von Angriffen getestet, ob die Anforderung für einen Neustart Metadaten mit den folgenden Merkmalen enthält: der Benutzer **220** ist der Eigentümer des Festplattenabbiels **213** des anfänglichen Urladens, der Eigentümer der VM-Instanz **204**, der reservierten IP-Adresse **208** und der optionalen permanenten Platte **214**; und ob die Anforderung für einen Neustart dazu dient, dieselbe Konfiguration neu zu starten, die gerade ausgeführt wird, das heißt, dasselbe Urlade-Festplattenabbild **213** und die IP-Adresse **208** sowie andere Metadaten, die in der ursprünglichen Anforderung für das Erzeugen einer Instanz verwendet wurden. Bei der Prüfung wird im BSS **103** ferner getestet, ob der Benutzer zur Nutzung des Neustart-Dienstes für diese Kombination aus Instanz und Abbild berechtigt ist.

[0062] Während des Normalbetriebs setzt die VM-Instanz **204** durch den Rücklauf **212** des Abwärtszählers den Abwärtszähler **113** in regelmäßigen Abständen zurück/startet ihn neu. Alternativ kann ein gesondertes Modul **222** diese Operation durchführen. Für jede eingetragene Instanz zählt der Abwärtszähler in dem automatisierten Neustart-Dienst **110** in regelmäßigen Abständen abwärts. Falls der Rücklauf des Abwärtszählers **113** eine Zeitlang nicht angefordert wird, fällt der Abwärtszähler auf null, wobei zu diesem Zeitpunkt die folgenden Vorgänge stattfinden: Die Anforderungsausführung **115** prüft mit dem OSS **104**, ob es einen allgemeinen Ausfallzustand gibt; die Anforderungsausführung **115** tritt mit dem OSS **104** in Dialogverkehr, um die VM-Instanz **204** zu beenden, die Zuweisung der reservierten IP-Adresse **208** aufzuheben und die optionale permanente ferne Platte **205** zu entfernen; die Anforderungsausführung **115** tritt mit dem OSS **104** in Dialogverkehr, um eine neue VM-Instanz zu starten, das eingetragene Urlade-Festplattenabbild **213** einzusetzen, die reservierte IP-Adresse **208** zuzuweisen und das optionale permanente Dateisys-

tem **214** einzusetzen. Auf der Grundlage dieser Vorgänge ist der Software-Anwendungsstapel **210** jetzt auf einer neuen VM-Instanz mit derselben IP-Adresse und denselben permanenten Daten, die vor dem Neustart zur Verfügung standen, betriebsbereit.

[0063] Eine enge Verflechtung zwischen dem automatisierten Neustart-Dienst **110** und dem OSS **104** und dem BSS **103** stellt sicher, dass der Neustart mit einem Zähler gemessen und dem berechtigten Benutzer optional berechnet werden kann. Beispielsweise kann ein Stunden- oder Tagestarif für die Nutzung des Neustart-Dienstes zwischen der Registrierung und der Deregistrierung berechnet werden, oder andere Ereignisse wie ein tatsächlicher Neustart könnten Ereignisse auslösen, die mit einem Verbrauchszähler gemessen werden. Eine solche Verflechtung macht es auch möglich, dass dem Benutzer über die Web-UI **101** Status- und Neustart-Meldungen angezeigt werden.

[0064] Erzwungener Neustart: Ein erzwungener Neustart **112** einer Instanz kann von der VM-Instanz oder einem berechtigten Benutzer ausgelöst werden. In diesem Fall wird dieselbe Prüfung **114** zur Vermeidung von Angriffen ausgeführt, die auch im Falle des auf null fallenden Abwärtszählers verwendet wird.

[0065] Automatisierter Standort-Neustart: Im Falle eines vollständigen Systemausfalls, beispielsweise sowohl von **100** als auch **200**, einschließlich des OSS und des BSS, werden Eintragungen **111** von Neustarts festgehalten. Wenn das OSS und das BSS neu gestartet werden, durchläuft die Anforderungsausführung **115** alle Eintragungen und führt einen Neustart von einem Teil der eingetragenen Instanzen oder von allen eingetragenen Instanzen durch.

[0066] Deregistrierung: Sowohl berechtigte Benutzer als auch die VM-Instanzen können die Eintragung eines automatisierten Neustarts jederzeit zurücknehmen. In diesem Fall testet die Prüfung zur Vermeidung von Angriffen erneut die Berechtigung und Stimmigkeit der Anforderung, um böswillige Anforderungen, die dazu dienen, Instanzen verwundbar zu machen, zu vermeiden.

[0067] Initialisierung **221** des Neustarts: Diese Komponente befindet sich in dem Abbild, das eine Vorlage von der in Ausführung befindlichen Instanz ist. Die Komponente wird während des anfänglichen Urladens der Instanz aufgerufen, so dass sie sich bei dem automatisierten Neustart-Dienst **110** einträgt **111**. Auf diese Weise kann die Instanz die Dienstgüte ohne Eingriff des Benutzers durch einen Neustart von dem Dienst herleiten. Auch können Hauptabbilder, die eine Unterstützung der Hochverfügbarkeit beinhalten, im Angebotskatalog in den BSS **103** zur Verfügung gestellt werden, welche keine zusätzliche Konfigura-

tion oder ein Eingreifen des Administrators erforderlich machen.

[0068] Nun Bezug nehmend auf [Fig. 5](#) ist ein Flussdiagramm gemäß einer Ausführungsform der vorliegenden Erfindung gezeigt. Wie im Schritt S1 gezeigt ist, wird eine ausgefallene virtuelle Instanz erkannt. Im Schritt S2 wird als Reaktion auf die ausgefallene virtuelle Instanz eine Anforderung für eine neue virtuelle Instanz empfangen. Im Schritt S3 wird eine Berechtigung eines Benutzers, der der ausgefallenen virtuellen Instanz zugeordnet ist, geprüft. Zu diesem Schritt gehört die Feststellung, ob der Benutzer ein Eigentümer von Folgendem ist: der ausgefallenen virtuellen Instanz; einem Urlade-Abbild, das zu der ausgefallenen virtuellen Instanz gehört; einem reservierten Internet Protocol (IP), das zu der ausgefallenen virtuellen Instanz gehört; und/oder einem permanenten Speicher, der zu der ausgefallenen virtuellen Instanz gehört. Im Schritt S4 wird ein Abwärtszähler, der zu der ausgefallenen virtuellen Instanz gehört, verringert. Im Schritt S5 wird die neue virtuelle Instanz gestartet, wenn der Abwärtszähler einen vorher festgelegten Schwellwert erreicht. Im Schritt S6 wird die ausgefallene virtuelle Instanz beendet. Im Schritt S7 wird die neue virtuelle Instanz eingetragen (z. B. ein Urlade-Abbild, ein reserviertes IP und ein permanenter Speicher der neuen virtuellen Instanz werden für den Benutzer eingetragen).

[0069] Zwar wird sie hier als Lösung zum automatisierten Starten von virtuellen Instanzen dargestellt und beschrieben, doch versteht es sich, dass die Erfindung des Weiteren verschiedene alternative Ausführungsformen bereitstellt. In einer Ausführungsform stellt die Erfindung zum Beispiel einen von einem Rechner lesbaren/nutzbaren Datenträger bereit, der Rechnerprogrammcode enthält, um es einer Rechnerinfrastruktur zu ermöglichen, eine Lösung zum Starten einer virtuellen Instanz zur Verfügung zu stellen. Insoweit enthält der von einem Rechner lesbare/nutzbare Datenträger Programmcode, der jeden einzelnen der verschiedenen Prozesse der Erfindung ausführt. Jede beliebige Kombination aus einem oder mehreren rechnerlesbaren Datenträgern kann verwendet werden. Der rechnerlesbare Datenträger kann ein rechnerlesbarer Signaldatenträger oder ein rechnerlesbares Speichermedium sein. Ein rechnerlesbares Speichermedium kann zum Beispiel, ohne auf diese beschränkt zu sein, ein(e) elektronische(s), magnetische(s), optische(s), elektromagnetische(s), Infrarot- oder Halbleitersystem, -vorrichtung, -einheit oder eine beliebige geeignete Kombination des Vorstehenden sein. Zu konkreteren Beispielen (wobei die Liste keinen Anspruch auf Vollständigkeit erhebt) für das rechnerlesbare Speichermedium würden folgende gehören: eine elektrische Verbindung mit einer oder mehreren Leitungen, eine tragbare Diskette eines Rechners, eine Festplatte, ein Direktzugriffsspeicher (RAM), ein

Nur-Lese-Speicher (ROM), ein löscherbarer programmierbarer Nur-Lese-Speicher (EPROM oder Flash-Speicher), ein Lichtwellenleiter, ein tragbarer Compact-Disk-Nur-Lese-Speicher (CD-ROM), eine optische Speichereinheit, eine magnetische Speichereinheit oder jede beliebige geeignete Kombination des Vorstehenden. Im Kontext dieses Schriftstücks kann ein rechnerlesbares Speichermedium jedes physisch greifbare Medium sein, das ein Programm zur Verwendung durch oder zur Verwendung in Verbindung mit einem Befehlsausführungssystem, einer Befehlsausführungsvorrichtung oder -einheit enthalten oder speichern kann.

[0070] In einer anderen Ausführungsform stellt die Erfindung ein Verfahren bereit, das den Prozess der Erfindung auf der Grundlage von Abonnements, Werbung und/oder Gebühren durchführt. Das heißt, ein Dienstanbieter wie zum Beispiel ein Lösungsintegrator, könnte anbieten, eine Lösung zum automatisierten Starten einer virtuellen Instanz zur Verfügung zu stellen. In diesem Fall kann der Dienstanbieter eine Rechner-Infrastruktur wie zum Beispiel die Rechner-Infrastruktur 102 (Fig. 4) erzeugen, verwalten, unterstützen usw., die den erfindungsgemäßen Prozess für einen oder mehrere Kunden durchführt. Als Leistung kann der Dienstanbieter im Rahmen eines Abonnements und/oder eines Gebührenvertrags Zahlungen von dem beziehungsweise den Kunden erhalten und/oder der Dienstanbieter kann Zahlungen aus dem Verkauf von Werbung erhalten.

[0071] In noch einer weiteren Ausführungsform stellt die Erfindung ein von einem Rechner ausgeführtes Verfahren zum automatisierten Starten einer virtuellen Instanz bereit. In diesem Fall kann eine Rechner-Infrastruktur wie zum Beispiel die Rechner-Infrastruktur 102 (Fig. 4) bereitgestellt werden, und ein oder mehrere Systeme zur Durchführung des erfindungsgemäßen Prozesses können beschafft (erzeugt, gekauft, verwendet, abgeändert usw.) und in der Rechner-Infrastruktur eingesetzt werden. Insofern kann der Einsatz eines Systems einen oder mehrere von folgenden Schritten umfassen: Installieren von Programmcode auf einer Datenverarbeitungseinheit wie zum Beispiel dem Rechnersystem 104 (Fig. 4) von einem rechnerlesbaren Datenträger; Hinzufügen von einer oder mehreren Datenverarbeitungseinheiten zu der Rechner-Infrastruktur; und Aufnehmen und/oder Abändern von einem oder mehreren vorhandenen Systemen der Rechner-Infrastruktur, um es der Rechner-Infrastruktur zu ermöglichen, den erfindungsgemäßen Prozess durchzuführen.

[0072] Ein rechnerlesbarer Signaldatenträger kann ein übertragenes Datensignal mit einem darin enthaltenen rechnerlesbaren Programmcode enthalten; zum Beispiel im Basisband oder als Teil einer Trägerwelle. Solch ein übertragenes Signal kann eine beliebige einer Vielzahl von Formen einschließlich elektro-

magnetischer, optischer Formen oder jede beliebige geeignete Kombination dieser Formen, ohne auf diese beschränkt zu sein, annehmen. Bei einem rechnerlesbaren Signaldatenträger kann es sich um jeden beliebigen rechnerlesbaren Datenträger handeln, der kein rechnerlesbares Speichermedium ist und der ein Programm zur Verwendung durch oder zur Verwendung in Verbindung mit einem Befehlsausführungssystem, einer Befehlsausführungsvorrichtung oder -einheit übertragen, weiterleiten oder transportieren kann.

[0073] Es versteht sich, dass die Begriffe "Programmcode" und "Rechnerprogrammcode" in der hier verwendeten Weise gleichbedeutend sind und jeden Ausdruck in jeder Sprache, in jedem Code oder in jeder Notation eines Satzes von Befehlen bedeuten, die eine Datenverarbeitungseinheit, welche über die Fähigkeit zur Verarbeitung von Daten verfügt, veranlassen sollen, eine bestimmte Funktion entweder direkt oder nach entweder einem oder beiden von Folgendem auszuführen: (a) Umwandlung in eine andere Sprache, in einen anderen Code oder in eine andere Notation; und/oder (b) Nachbildung in einer anderen materiellen Form. Insofern kann der Programmcode in Form von einem oder mehreren von Folgendem realisiert werden: einem Anwendungs-/Software-Programm, Komponenten-Software/einer Bibliothek mit Funktionen, einem Betriebssystem, einem grundlegenden Einheitensystem/-treiber für eine bestimmte Datenverarbeitungseinheit und dergleichen.

[0074] Ein Datenverarbeitungssystem, das zur Speicherung und/oder Ausführung von Programmcode geeignet ist, kann im Rahmen dieser Erfindung bereitgestellt werden und mindestens einen Prozessor enthalten, der über einen Systembus direkt oder indirekt mit einem oder mehreren Speicherelementen verbunden ist. Die Speicherelemente können, ohne auf diese beschränkt zu sein, einen lokalen Speicher, der während der tatsächlichen Ausführung des Programmcodes verwendet wird, einen Massenspeicher und Cachespeicher beinhalten, die eine vorübergehende Speicherung von mindestens einem Teil des Programmcodes ermöglichen, um die Häufigkeit zu verringern, mit der der Code während der Ausführung aus dem Massenspeicher abgerufen werden muss. Eingabe-/Ausgabe- und/oder andere externe Einheiten (einschließlich Tastaturen, Bildschirme, Zeigegeräte usw., ohne auf diese beschränkt zu sein) können entweder direkt oder über zwischengeschaltete Einheitensteuerungen mit dem System verbunden werden.

[0075] Netzwerkadapter können ebenfalls mit dem System verbunden werden, damit das Datenverarbeitungssystem mit anderen Datenverarbeitungssystemen, fernem Druckern, Speichereinheiten und/oder dergleichen über eine beliebige Kombination aus zwi-

schengeschalteten privaten oder öffentlichen Netzwerken verbunden werden kann. Zu der Veranschaulichung dienenden Netzwerkadapters gehören, ohne auf diese beschränkt zu sein, Modems, Kabelmodems und Ethernet-Karten.

ZITATE ENTHALTEN IN DER BESCHREIBUNG

Diese Liste der vom Anmelder aufgeführten Dokumente wurde automatisiert erzeugt und ist ausschließlich zur besseren Information des Lesers aufgenommen. Die Liste ist nicht Bestandteil der deutschen Patent- bzw. Gebrauchsmusteranmeldung. Das DPMA übernimmt keinerlei Haftung für etwaige Fehler oder Auslassungen.

Zitierte Nicht-Patentliteratur

- "Draft NIST Working Definition of Cloud Computing" von Peter Mell und Tim Grance vom 7. Oktober 2009 [\[0024\]](#)

Patentansprüche

1. Verfahren zum Starten einer virtuellen Instanz in einer Cloud-Computing-Umgebung, wobei das Verfahren Folgendes umfasst:
 - Erkennen einer ausgefallenen virtuellen Instanz in der Cloud-Computing-Umgebung;
 - als Reaktion auf die ausgefallene virtuelle Instanz Empfangen einer Anforderung für eine neue virtuelle Instanz;
 - Prüfen einer Berechtigung eines Benutzers, der der ausgefallenen virtuellen Instanz zugeordnet ist;
 - Verringern eines Abwärtszählers, der zu der ausgefallenen virtuellen Instanz gehört;
 - Starten der neuen virtuellen Instanz, wenn der Abwärtszähler einen vorher festgelegten Schwellwert erreicht; und
 - Beenden der ausgefallenen virtuellen Instanz.
2. Verfahren nach Anspruch 1, wobei der Prüfvorgang die Feststellung umfasst, ob der Benutzer ein Eigentümer der ausgefallenen virtuellen Instanz ist.
3. Verfahren nach Anspruch 1 oder Anspruch 2, wobei der Prüfvorgang die Feststellung umfasst, ob der Benutzer ein Eigentümer eines Urlade-Abbilds ist, das zu der ausgefallenen virtuellen Instanz gehört.
4. Verfahren nach einem der Ansprüche 1 bis 3, wobei der Prüfvorgang die Feststellung umfasst, ob der Benutzer ein Eigentümer einer reservierten Internet-Protocol-(IP-)Adresse ist, die zu der ausgefallenen virtuellen Instanz gehört.
5. Verfahren nach einem der vorhergehenden Ansprüche, wobei der Prüfvorgang die Feststellung umfasst, ob der Benutzer ein Eigentümer des permanenten Speichers ist, das zu der ausgefallenen virtuellen Instanz gehört.
6. Verfahren nach einem der vorhergehenden Ansprüche, wobei der vorher festgelegte Schwellwert null umfasst.
7. Verfahren nach einem der vorhergehenden Ansprüche, das des Weiteren das Eintragen eines Urlade-Abbilds, einer reservierten IP-Adresse und eines permanenten Speichers der neuen virtuellen Instanz für den Benutzer umfasst.
8. Verfahren nach einem der vorhergehenden Ansprüche, wobei die neue virtuelle Instanz einen Neustart der ausgefallenen virtuellen Instanz umfasst.
9. System zum Starten einer virtuellen Instanz in einer Cloud-Computing-Umgebung, wobei das System Folgendes umfasst:
 - ein Speichermedium, das Befehle umfasst;

1. einen Bus, der mit dem Speichermedium verbunden ist; und
1. einen mit dem Bus verbundenen Prozessor, der, wenn er die Befehle ausführt, bewirkt, dass das System:
 - 1. eine ausgefallene virtuelle Instanz in der Cloud-Computing-Umgebung erkennt;
 - 1. als Reaktion auf die ausgefallene virtuelle Instanz eine Anforderung für eine neue virtuelle Instanz empfängt;
 - 1. eine Berechtigung eines Benutzers prüft, der der ausgefallenen virtuellen Instanz zugeordnet ist;
 - 1. einen Abwärtszähler verringert, der zu der ausgefallenen virtuellen Instanz gehört;
 - 1. die neue virtuelle Instanz startet, wenn der Abwärtszähler einen vorher festgelegten Schwellwert erreicht; und
 - 1. die ausgefallene virtuelle Instanz beendet.
10. System nach Anspruch 9, wobei das System des Weiteren veranlasst wird, festzustellen, ob der Benutzer ein Eigentümer der ausgefallenen virtuellen Instanz ist.
11. System nach Anspruch 9 oder Anspruch 10, wobei das System des Weiteren veranlasst wird, festzustellen, ob der Benutzer ein Eigentümer eines Urlade-Abbilds ist, das zu der ausgefallenen virtuellen Instanz gehört.
12. System nach einem der Ansprüche 9 bis 11, wobei das System des Weiteren veranlasst wird, festzustellen, ob der Benutzer ein Eigentümer einer reservierten Internet-Protocol-(IP-)Adresse ist, die zu der ausgefallenen virtuellen Instanz gehört.
13. System nach einem der Ansprüche 9 bis 12, wobei das System des Weiteren veranlasst wird, festzustellen, ob der Benutzer ein Eigentümer des permanenten Speichers ist, der zu der ausgefallenen virtuellen Instanz gehört.
14. System nach einem der Ansprüche 9 bis 13, wobei das System des Weiteren veranlasst wird, ein Urlade-Abbild, eine reservierte IP-Adresse und einen permanenten Speicher der neuen virtuellen Instanz für den Benutzer einzutragen.
15. Rechnerprogramm, das Rechnerprogrammcode umfasst, der auf einem rechnerlesbaren Datenträger gespeichert wird, um, wenn es in ein Rechnersystem geladen und darauf ausgeführt wird, das Rechnersystem zu veranlassen, alle Schritte eines Verfahrens nach einem der Ansprüche 1 bis 8 durchzuführen.

Es folgen 5 Blatt Zeichnungen

Anhängende Zeichnungen

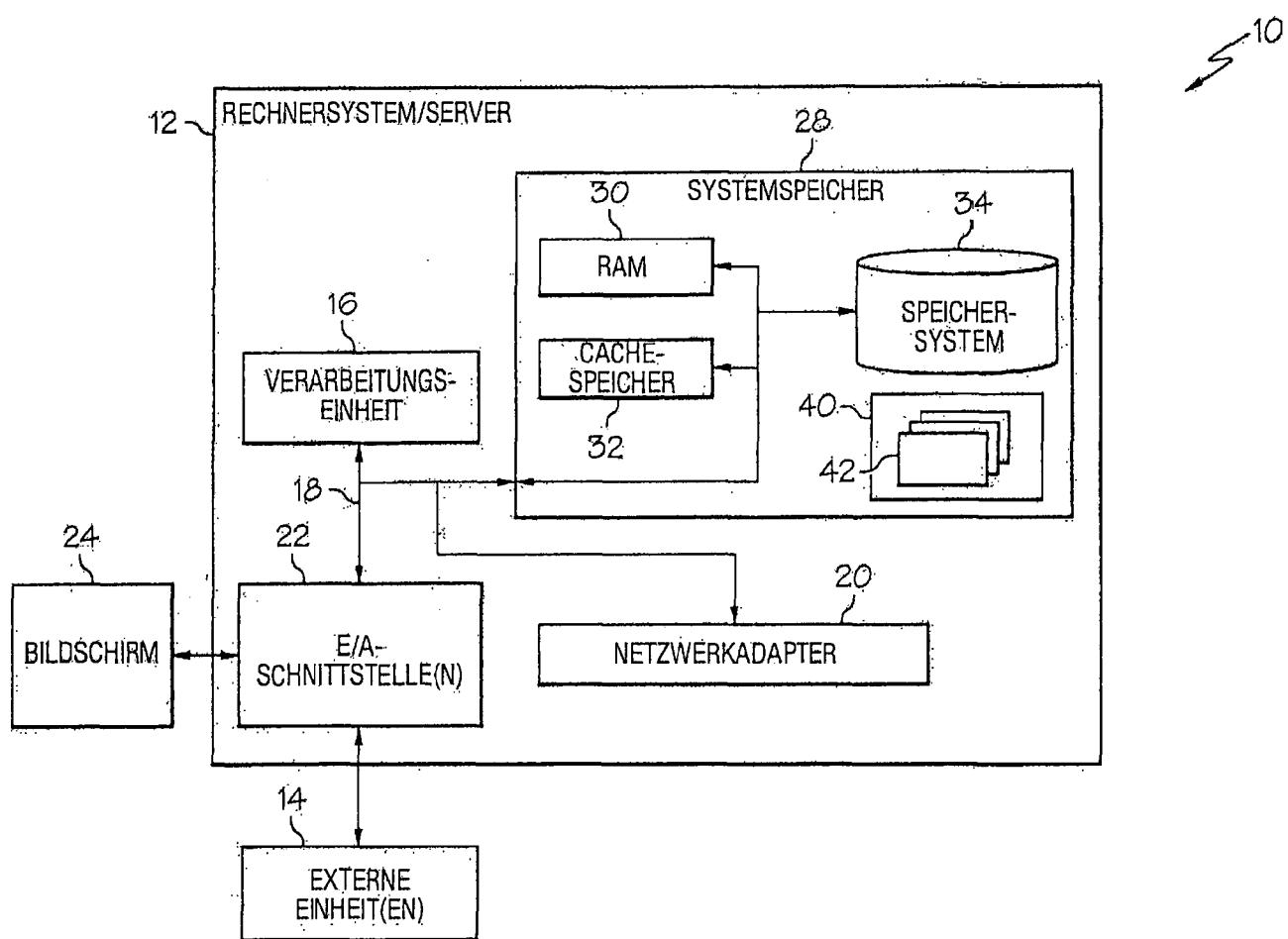
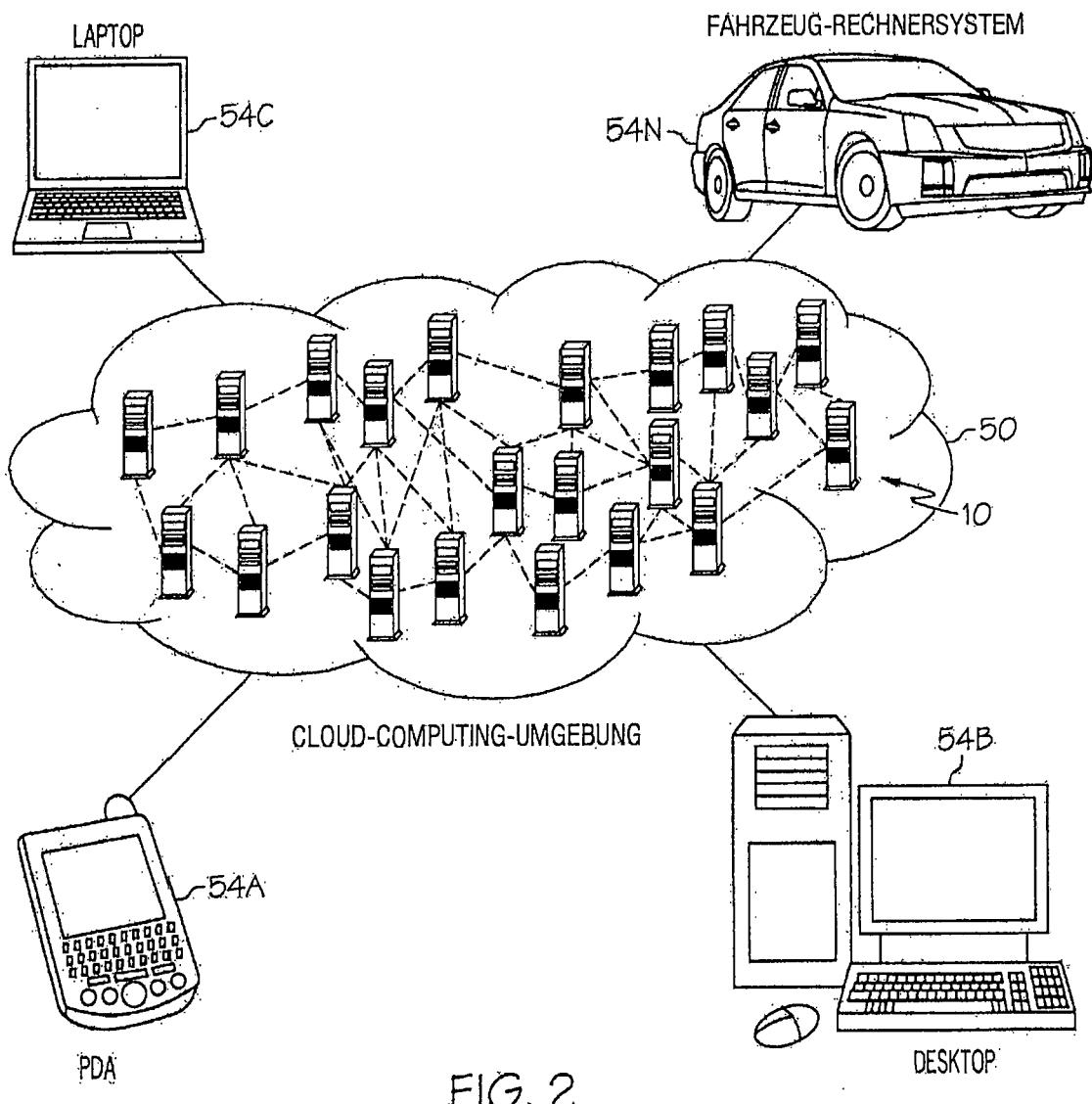


FIG. 1



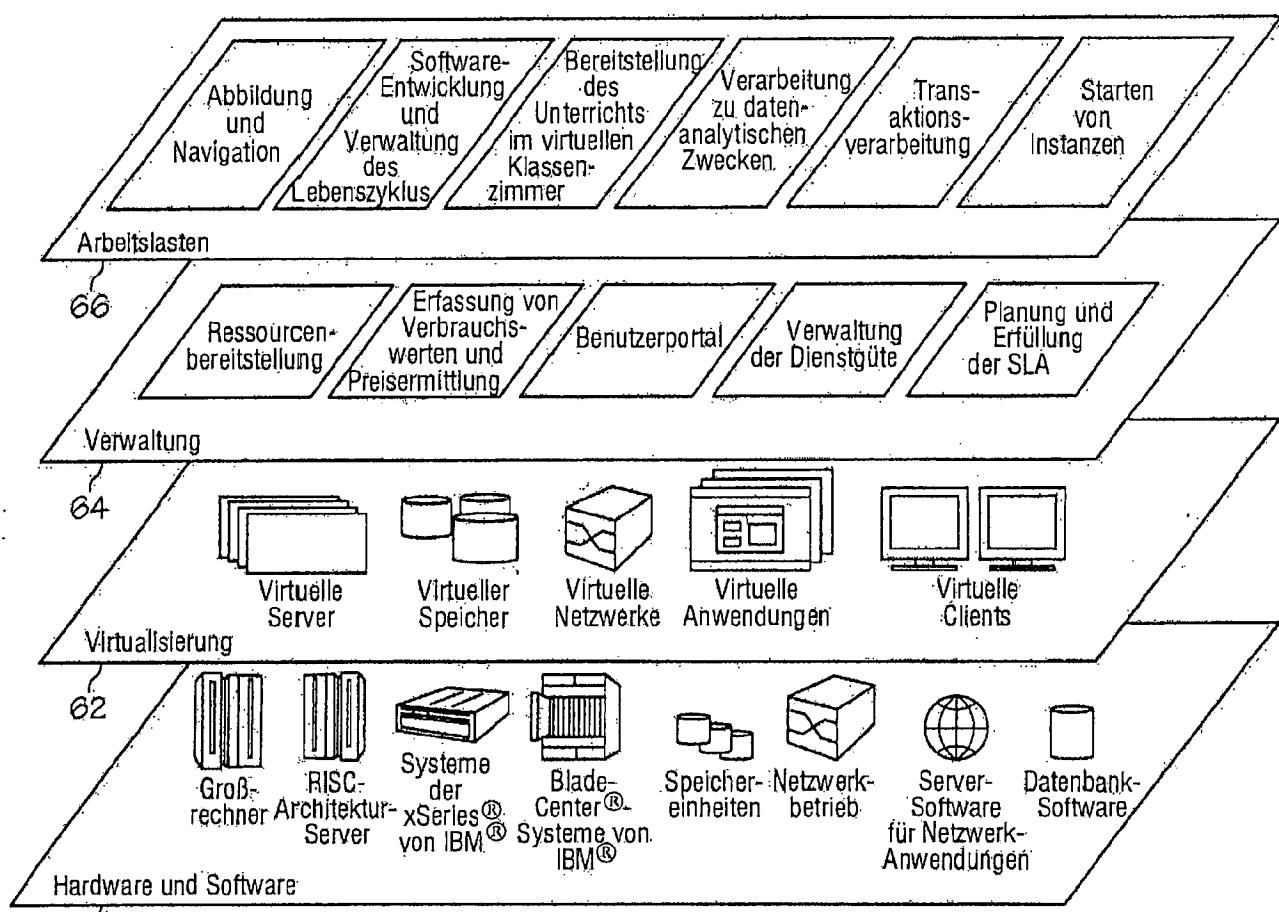
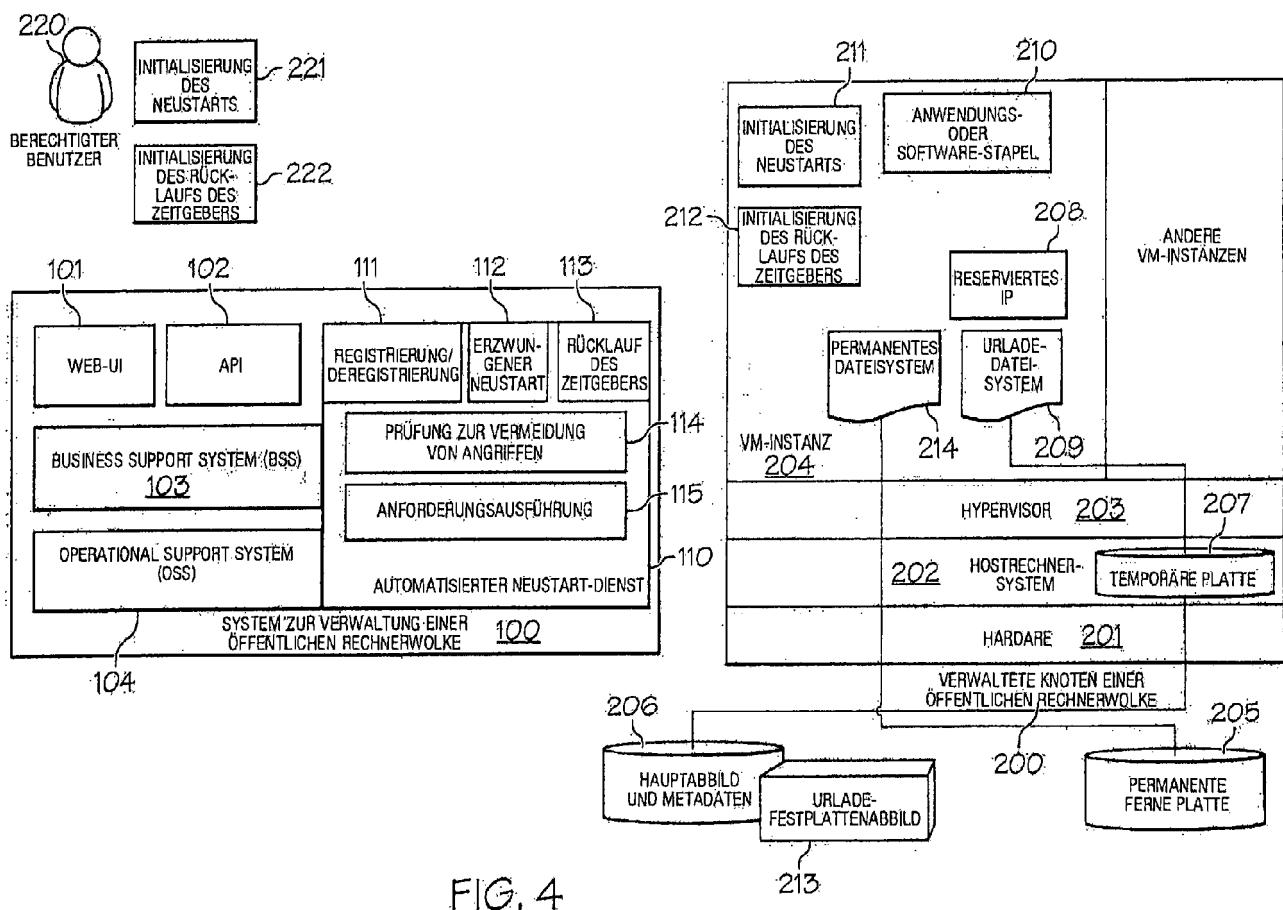


FIG. 3



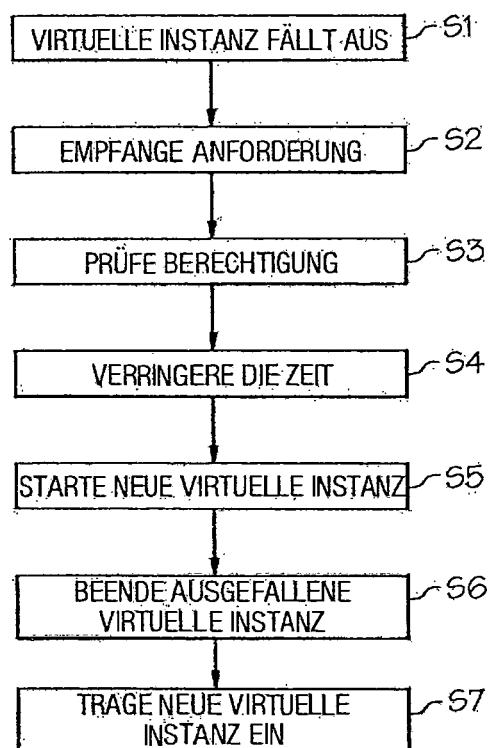


FIG. 5