



(12)

Veröffentlichung

der internationalen Anmeldung mit der
(87) Veröffentlichungs-Nr.: **WO 2019/021114**
in der deutschen Übersetzung (Art. III § 8 Abs. 2
IntPatÜG)
(21) Deutsches Aktenzeichen: **11 2018 002 955.7**
(86) PCT-Aktenzeichen: **PCT/IB2018/055296**
(86) PCT-Anmeldetag: **17.07.2018**
(87) PCT-Veröffentlichungstag: **31.01.2019**
(43) Veröffentlichungstag der PCT Anmeldung
in deutscher Übersetzung: **07.05.2020**

(51) Int Cl.: **G06F 16/10 (2019.01)**

(30) Unionspriorität:
15/660,715 26.07.2017 US
(71) Anmelder:
**International Business Machines Corporation,
Armonk, N.Y., US**

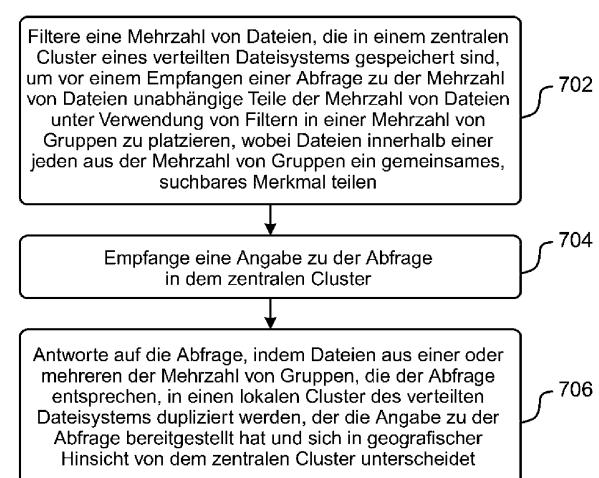
(74) Vertreter:
**Richardt Patentanwälte PartG mbB, 65185
Wiesbaden, DE**
(72) Erfinder:
**Balachandran, Subashini, San Jose, CA, US;
Zhang, Rui, San Francisco, CA, US**

Prüfungsantrag gemäß § 44 PatG ist gestellt.

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen.

(54) Bezeichnung: **KOGNITIVE DATEI- UND OBJEKTVERWALTUNG FÜR VERTEILTE SPEICHERUMGEBUNGEN**

(57) Zusammenfassung: Ein Verfahren beinhaltet ein Filtern einer Mehrzahl von Dateien, die in einem zentralen Cluster eines verteilten Dateisystems gespeichert sind, um vor einem Empfangen einer Abfrage zu der Mehrzahl von Dateien unabhängige Teile der Mehrzahl von Dateien unter Verwendung von Filtern in einer Mehrzahl von Gruppen zu platzieren. Dateien innerhalb einer jeden aus der Mehrzahl von Gruppen teilen ein gemeinsames, suchbares Merkmal. Das Verfahren beinhaltet außerdem eine Angabe zu der Abfrage in dem zentralen Cluster. Darüber hinaus beinhaltet das Verfahren ein Antworten auf die Abfrage durch ein Duplizieren von Dateien aus einer oder mehreren der Mehrzahl von Gruppen, die der Abfrage entsprechen, in einen lokalen Cluster des verteilten Dateisystems, der die Angabe zu der Abfrage bereitgestellt hat und sich in geografischer Hinsicht von dem zentralen Cluster unterscheidet.



Beschreibung**HINTERGRUND**

[0001] Die vorliegende Erfindung bezieht sich auf verteilte Speicherumgebungen und im Besonderen auf eine Datei- und Objektverwaltung in verteilten Speicherumgebungen unter Verwendung von kognitiven Methoden.

[0002] Ein zentrales, verteiltes, geclustertes Dateisystem ist in der Lage, für alle Dateien und/oder Objekte, die innerhalb des verteilten Dateisystems gespeichert sind, einen globalen Namensbereich bereitzustellen. Ein solches System fungiert somit als ein globales Daten-Repository und kann gewaltige Mengen an Datenspeicherkapazität enthalten sowie über verschiedene Speicherklassen hinweg abgestuft sein. Ein solches System beinhaltet einen zentralen Daten-Repository-Cluster und eine Mehrzahl von lokalen Clustern, die kleiner als das zentrale Daten-Repository sind. Jeder - auch als ein Cache-Cluster bezeichnete - lokale Cluster beinhaltet eine Mehrzahl von Speichereinheiten einer oder mehrerer Speicherklassen und kann WAN-Caching-Methoden (Wide Area Network, Weitverkehrsnetzwerk) verwenden, um Teile der in dem verteilten Dateisystem gespeicherten Daten innerhalb eines Caches des lokalen Clusters zwischenspeichern.

[0003] WAN-Caching dient üblicherweise dazu, Daten effizienter bereitzustellen, auf die ein Benutzer in dem lokalen Cluster häufig zugreift. Auf diese Weise werden in dem lokalen Cluster ausgeführte Abfragen und/oder Analyseoperationen über den lokalen Cluster durchgeführt. Ein Speichern aller Daten in jedem lokalen Cluster und in dem zentralen Cluster ist aufgrund der hohen Datenspeicheranforderungen und des hohen Datensynchronisierungsaufwands unzweckmäßig. Aus diesem Grund wird zum Speichern einer globalen Kopie der Daten häufig der zentrale Cluster verwendet, aus dem die verschiedenen lokalen Cluster die Daten für Abfragen und/oder Analyseoperationen abrufen können. Bei diesem typischen Nutzungsmodell werden Daten in den verschiedenen lokalen Clustern empfangen und in diese eingespeist, wobei die Cluster in geografischer Hinsicht über die physische Fläche des verteilten Dateisystems verteilt sein können. Die eingespeisten Daten werden daraufhin aus den verschiedenen lokalen Clustern in den zentralen Cluster dupliziert, um als eine globale Kopie gespeichert zu werden. Bei einem alternativen Nutzungsmodell kann der zentrale Cluster als ein Daten-Repository dienen, und mehrere lokale Cluster können in dem zentralen Repository gespeicherte Daten als Nur-Lese-Instanzen verwenden, um schneller darauf zugreifen zu können.

[0004] Bei beiden Nutzungsmodellen ist eine Latenz für einen Fernzugriff auf Daten, die in dem zentralen

Cluster gespeichert sind, für Benutzer des verteilten Dateisystems höher als erwünscht, sofern die Daten nicht auch in dem lokalen Cluster gespeichert sind, von dem aus der Benutzer zuzugreifen versucht. Dies verursacht Schwankungen bei der Zugriffsleistung für das verteilte Dateisystem.

[0005] Darüber hinaus können Big-Data-Analysen ein Untersuchen sehr großer Teile (oder aller) Daten erfordern, die in dem globalen Daten-Repository gespeichert sind. Unter Umständen kann es zu vielen Ressourcen verbrauchen, alle Abfragen direkt in dem globalen Daten-Repository auszuführen, weshalb Abfragen möglicherweise in einer von dem Controller abweichenden Einheit ausgeführt werden, die Zugriff auf das globale Daten-Repository hat, indem alle Daten aus dem globalen Daten-Repository in die andere Einheit kopiert werden. Nach dem Ausführen der Analyseoperation(en) werden in manchen Fällen die kopierten Daten aus der anderen Einheit gelöscht, was ein erneutes Kopieren aller Daten aus dem globalen Daten-Repository erforderlich macht, um eine nachfolgende Analyseoperation auszuführen. In anderen Fällen können die kopierten Daten zwischengespeichert werden, was eine gewaltige Menge an lokalem Speicher in der anderen Einheit erfordert. In beiden Fällen beinhalten die kopierten Daten alle Daten in dem globalen Daten-Repository, da es keine intelligente Instanz gibt, die ermittelt, welche Daten bei der Analyseoperation verwendet werden sollen und was unkopiert in dem globalen Daten-Repository verbleiben kann. Darüber hinaus werden womöglich nicht alle kopierten Daten bei der Analyseoperation verwendet und bei der Verarbeitung der Abfrage daher einfach verworfen, wodurch beträchtliche Ressourcen wie z.B. Netzwerkbandbreite, Verarbeitungsbandbreite, Arbeitsspeicherkapazität, Zeit usw. verschwendet werden.

[0006] Wenn in dem globalen Daten-Repository ausreichend Verarbeitungskapazität vorhanden ist, um alle Abfragen zu verarbeiten, können die Daten alternativ dennoch zwischengespeichert (d.h. kopiert und in einem lokalen Arbeitsspeicher des globalen Daten-Repositorys beibehalten) werden, um die Abfrageverarbeitung zu beschleunigen. Wenn die Abfragen auf diese Weise ausgeführt werden, entsteht ein ähnlicher Ressourcenverbrauch, wie wenn die Daten kopiert werden, um die Abfragen in der anderen Einheit auszuführen.

KURZDARSTELLUNG

[0007] Bei einer Ausführungsform beinhaltet ein Verfahren ein Filtern einer Mehrzahl von Dateien, die in einem zentralen Cluster eines verteilten Dateisystems gespeichert sind, um vor einem Empfangen einer Abfrage zu der Mehrzahl von Dateien unabhängige Teile der Mehrzahl von Dateien unter Verwendung von Filtern in einer Mehrzahl von Gruppen zu plat-

zieren. Dateien innerhalb einer jeden aus der Mehrzahl von Gruppen teilen ein gemeinsames, suchbares Merkmal. Das Verfahren beinhaltet außerdem eine Angabe zu der Abfrage in dem zentralen Cluster. Darüber hinaus beinhaltet das Verfahren ein Antworten auf die Abfrage durch ein Duplizieren von Dateien aus einer oder mehreren der Mehrzahl von Gruppen, die der Abfrage entsprechen, in einen lokalen Cluster des verteilten Dateisystems, der die Angabe zu der Abfrage bereitgestellt hat und sich in geografischer Hinsicht von dem zentralen Cluster unterscheidet.

[0008] Bei einer weiteren Ausführungsform beinhaltet ein Computerprogrammprodukt ein durch einen Computer lesbaren Speichermedium mit darin enthaltenen Programmanweisungen. Das durch einen Computer lesbare Speichermedium ist kein flüchtiges Signal an sich, und die enthaltenen Programmanweisungen sind durch eine Verarbeitungsschaltung ausführbar, um die Verarbeitungsschaltung veranlassen, durch die Verarbeitungsschaltung eine Mehrzahl von Dateien zu filtern, die in einem zentralen Cluster eines verteilten Dateisystems gespeichert sind, um unabhängige Teile der Mehrzahl von Dateien vor einem Empfangen einer Abfrage zu der Mehrzahl von Dateien unter Verwendung von Filtern in einer Mehrzahl von Gruppen zu platzieren. Dateien innerhalb einer jeden aus der Mehrzahl von Gruppen teilen ein gemeinsames, suchbares Merkmal. Die enthaltenen Programmanweisungen sind außerdem durch die Verarbeitungsschaltung ausführbar, um die Verarbeitungsschaltung zu veranlassen, durch die Verarbeitungsschaltung in dem zentralen Cluster eine Angabe zu der Abfrage zu empfangen. Darüber hinaus sind die enthaltenen Programmanweisungen durch die Verarbeitungsschaltung ausführbar, um die Verarbeitungsschaltung zu veranlassen, durch die Verarbeitungsschaltung auf die Abfrage zu antworten, indem Dateien aus einer oder mehreren der Mehrzahl von Gruppen, die der Abfrage entsprechen, in einen lokalen Cluster des verteilten Dateisystems dupliziert werden, der die Angabe zu der Abfrage bereitgestellt hat.

[0009] Bei einer weiteren Ausführungsform beinhaltet ein System eine Verarbeitungsschaltung, einen Arbeitsspeicher und in dem Arbeitsspeicher gespeicherte Logik, die bei Ausführung durch die Verarbeitungsschaltung die Verarbeitungsschaltung veranlasst, eine Mehrzahl von Dateien zu filtern, die in einem zentralen Cluster eines verteilten Dateisystems gespeichert sind, um unabhängige Teile der Mehrzahl von Dateien vor einem Empfangen einer Abfrage zu der Mehrzahl von Dateien unter Verwendung von Filtern in einer Mehrzahl von Gruppen zu platzieren. Dateien innerhalb einer jeden aus der Mehrzahl von Gruppen teilen ein gemeinsames, suchbares Merkmal. Bei Ausführung durch die Verarbeitungsschaltung veranlasst die Logik die Verarbeitungsschaltung außerdem, in dem zentralen Cluster eine Angabe zu

der Abfrage zu empfangen. Bei Ausführung durch die Verarbeitungsschaltung veranlasst die Logik die Verarbeitungsschaltung darüber hinaus, auf die Abfrage zu antworten, indem Dateien aus einer oder mehreren der Mehrzahl von Gruppen, die der Abfrage entsprechen, in einen lokalen Cluster des verteilten Dateisystems dupliziert werden, der die Angabe zu der Abfrage bereitgestellt hat.

[0010] Gemäß einer weiteren Ausführungsform beinhaltet ein Verfahren ein Empfangen einer Mehrzahl von Dateien aus einer oder mehreren Quellen in einem zentralen Cluster eines verteilten Dateisystems, wobei die Mehrzahl von Dateien Text- und unstrukturierte Daten beinhaltet. Außerdem beinhaltet das Verfahren ein Speichern der Mehrzahl von Dateien in dem zentralen Cluster und ein Konvertieren der unstrukturierten Daten in Text in dem zentralen Cluster. Darüber hinaus beinhaltet das Verfahren ein Filtern der Mehrzahl von Dateien, um vor einem Empfangen einer Abfrage zu der Mehrzahl von Dateien unabhängige Teile der Mehrzahl von Dateien unter Verwendung von Filtern in einer Mehrzahl von Gruppen zu platzieren. Dateien innerhalb einer jeden aus der Mehrzahl von Gruppen teilen ein gemeinsames, suchbares Merkmal, und die Filter werden auf den Text der Dateien angewendet, nachdem dieser aus den unstrukturierten Daten konvertiert wurde.

[0011] Gemäß einer weiteren Ausführungsform beinhaltet ein Verfahren ein Durchsuchen eines lokalen Clusters eines verteilten Dateisystems nach Dateien, die für eine Abfrage relevant sind, bevor eine Angabe zu der Abfrage an einen zentralen Cluster des verteilten Dateisystems gesendet wird. Das Verfahren beinhaltet außerdem ein Senden der Angabe zu der Abfrage an den zentralen Cluster als Reaktion auf eine Ermittlung, dass die relevanten Dateien nicht in dem lokalen Cluster gespeichert sind. Das Verfahren beinhaltet außerdem ein Empfangen einer Gruppe von Dateien, die für die Abfrage relevant sind, in dem lokalen Cluster. Zusätzlich beinhaltet das Verfahren ein Ausführen der Abfrage zu der Gruppe von Dateien unter der Bedingung, dass die Gruppe von Dateien nicht alle Dateien enthält, die in dem zentralen Cluster gespeichert sind. Darüber hinaus beinhaltet das Verfahren ein Speichern der Gruppe von Dateien für eine vorbestimmte Zeitspanne in dem lokalen Cluster, die im Einklang mit einer Richtlinie für die Cache-Leerung mit einem letzten Zugriff auf die Gruppe von Dateien beginnt.

[0012] Andere Aspekte und Ausführungsformen der vorliegenden Erfindung ergeben sich aus der folgenden ausführlichen Beschreibung, die in Zusammenhang mit den Zeichnungen die Grundgedanken der Erfindung beispielhaft veranschaulicht.

Figurenliste

Fig. 1 stellt einen Cloud-Computing-Knoten gemäß einer Ausführungsform der vorliegenden Erfindung dar.

Fig. 2 stellt eine Cloud-Computing-Umgebung gemäß einer Ausführungsform der vorliegenden Erfindung dar.

Fig. 3 stellt Abstraktionsmodellschichten gemäß einer Ausführungsform der vorliegenden Erfindung dar.

Fig. 4 veranschaulicht ein mehrstufiges Datenspeichersystem gemäß einer Ausführungsform.

Die **Fig. 5A** bis **Fig. 5C** zeigen ein verteiltes System während mehrerer Phasen eines Datenfilterns und -gruppierens für ein effizientes Verarbeiten einer Abfrage gemäß einer Ausführungsform.

Fig. 6 zeigt ein Datenfiltern und -gruppieren für ein effizientes Verarbeiten einer Abfrage in einem beispielhaften verteilten System.

Fig. 7 zeigt einen Ablaufplan eines Verfahrens gemäß einer Ausführungsform.

Fig. 8 zeigt einen Ablaufplan eines Verfahrens gemäß einer Ausführungsform.

AUSFÜHRLICHE BESCHREIBUNG

[0013] Die folgende Beschreibung dient zur Veranschaulichung der allgemeinen Grundgedanken der vorliegenden Erfindung und ist nicht als Einschränkung der hier beanspruchten erfinderischen Konzepte gedacht. Darüber hinaus können bestimmte hier beschriebene Merkmale in Kombination mit anderen beschriebenen Merkmalen in jeder der verschiedenen möglichen Kombinationen und Umsetzungen verwendet werden.

[0014] Sofern hier nicht anderweitig konkret genannt, sind sämtliche Begriffe in ihrer größtmöglichen Auslegung zu verstehen, einschließlich aus der Beschreibung hervorgehende Bedeutungen sowie durch einen Fachmann verstandene und/oder in Wörterbüchern, Abhandlungen usw. definierte Bedeutungen.

[0015] Darüber hinaus ist zu beachten, dass die in der Beschreibung und den beigefügten Ansprüchen verwendeten Singularformen „ein/eine/eines“ und „der/die/das“ auch die Pluralformen einschließt, soweit dies nicht anderweitig angegeben wird. Des Weiteren wird darauf verwiesen, dass die Begriffe „weist auf“ und/oder „aufweisend“ in dieser Beschreibung das Vorhandensein der genannten Merkmale, Ganzzahlen, Schritte, Operationen, Elementen und/oder Komponenten angeben, ohne jedoch das Vorhandensein oder die Hinzufügung von einem/ei-

ner oder mehreren anderen Merkmalen, Ganzzahlen, Schritten, Operationen, Elementen, Komponenten und/oder Gruppen derselben auszuschließen. Wie für den Fachmann offensichtlich sein sollte, gibt der Begriff „ungefähr“, wie hierin verwendet, den Wert an, der dem Begriff „ungefähr“ vorangeht, sowie jegliche Werte, die angemessen nahe dem Wert sind, der dem Begriff „ungefähr“ vorangeht. Sofern nicht anderweitig angegeben, bezeichnet der Begriff „ungefähr“ den Wert, der dem Begriff „ungefähr“ vorangeht, $\pm 10\%$ des Werts. Zum Beispiel gibt „ungefähr 10“ alle Werte von einschließlich 9,0 bis einschließlich 11,0 an.

[0016] Die folgende Beschreibung offenbart mehrere bevorzugte Ausführungsformen von Systemen, Verfahren und Computerprogrammprodukten für eine kognitive Verwaltung von Dateien und Objekten in einer verteilten Speicherumgebung.

[0017] Bei einer allgemeinen Ausführungsform beinhaltet ein Verfahren ein Filtern einer Mehrzahl von Dateien, die in einem zentralen Cluster eines verteilten Dateisystems gespeichert sind, um vor einem Empfangen einer Abfrage zu der Mehrzahl von Dateien unabhängige Teile der Mehrzahl von Dateien unter Verwendung von Filtern in einer Mehrzahl von Gruppen zu platzieren. Dateien innerhalb einer jeden aus der Mehrzahl von Gruppen teilen ein gemeinsames, suchbares Merkmal. Das Verfahren beinhaltet außerdem eine Angabe zu der Abfrage in dem zentralen Cluster. Darüber hinaus beinhaltet das Verfahren ein Antworten auf die Abfrage durch ein Duplizieren von Dateien aus einer oder mehreren der Mehrzahl von Gruppen, die der Abfrage entsprechen, in einen lokalen Cluster des verteilten Dateisystems, der die Angabe zu der Abfrage bereitgestellt hat und sich in geografischer Hinsicht von dem zentralen Cluster unterscheidet.

tung zu veranlassen, durch die Verarbeitungsschaltung in dem zentralen Cluster eine Angabe zu der Abfrage zu empfangen. Darüber hinaus sind die enthaltenen Programmanweisungen durch die Verarbeitungsschaltung ausführbar, um die Verarbeitungsschaltung zu veranlassen, durch die Verarbeitungsschaltung auf die Abfrage zu antworten, indem Dateien aus einer oder mehreren der Mehrzahl von Gruppen, die der Abfrage entsprechen, in einen lokalen Cluster des verteilten Dateisystems dupliziert werden, der die Angabe zu der Abfrage bereitgestellt hat.

[0019] Bei einer weiteren allgemeinen Ausführungsform beinhaltet ein System eine Verarbeitungsschaltung, einen Arbeitsspeicher und in dem Arbeitsspeicher gespeicherte Logik, die bei Ausführung durch die Verarbeitungsschaltung die Verarbeitungsschaltung veranlasst, eine Mehrzahl von Dateien zu filtern, die in einem zentralen Cluster eines verteilten Dateisystems gespeichert sind, um unabhängige Teile der Mehrzahl von Dateien vor einem Empfangen einer Abfrage zu der Mehrzahl von Dateien unter Verwendung von Filtern in einer Mehrzahl von Gruppen zu platzieren. Dateien innerhalb einer jeden aus der Mehrzahl von Gruppen teilen ein gemeinsames, suchbares Merkmal. Bei Ausführung durch die Verarbeitungsschaltung veranlasst die Logik die Verarbeitungsschaltung außerdem, in dem zentralen Cluster eine Angabe zu der Abfrage zu empfangen. Bei Ausführung durch die Verarbeitungsschaltung veranlasst die Logik die Verarbeitungsschaltung darüber hinaus, auf die Abfrage zu antworten, indem Dateien aus einer oder mehreren der Mehrzahl von Gruppen, die der Abfrage entsprechen, in einen lokalen Cluster des verteilten Dateisystems dupliziert werden, der die Angabe zu der Abfrage bereitgestellt hat.

[0020] Gemäß einer weiteren allgemeinen Ausführungsform beinhaltet ein Verfahren ein Empfangen einer Mehrzahl von Dateien aus einer oder mehreren Quellen in einem zentralen Cluster eines verteilten Dateisystems, wobei die Mehrzahl von Dateien Text- und unstrukturierte Daten beinhaltet. Außerdem beinhaltet das Verfahren ein Speichern der Mehrzahl von Dateien in dem zentralen Cluster und ein Konvertieren der unstrukturierten Daten in Text in dem zentralen Cluster. Darüber hinaus beinhaltet das Verfahren ein Filtern der Mehrzahl von Dateien, um vor einem Empfangen einer Abfrage zu der Mehrzahl von Dateien unabhängige Teile der Mehrzahl von Dateien unter Verwendung von Filtern in einer Mehrzahl von Gruppen zu platzieren. Dateien innerhalb einer jeden aus der Mehrzahl von Gruppen teilen ein gemeinsames, suchbares Merkmal, und die Filter werden auf den Text der Dateien angewendet, nachdem dieser aus den unstrukturierten Daten konvertiert wurde.

[0021] Gemäß einer weiteren allgemeinen Ausführungsform beinhaltet ein Verfahren ein Durchsuchen eines lokalen Clusters eines verteilten Dateisystems

nach Dateien, die für eine Abfrage relevant sind, bevor eine Angabe zu der Abfrage an einen zentralen Cluster des verteilten Dateisystems gesendet wird. Das Verfahren beinhaltet außerdem ein Senden der Angabe zu der Abfrage an den zentralen Cluster als Reaktion auf eine Ermittlung, dass die relevanten Dateien nicht in dem lokalen Cluster gespeichert sind. Das Verfahren beinhaltet außerdem ein Empfangen einer Gruppe von Dateien, die für die Abfrage relevant sind, in dem lokalen Cluster. Zusätzlich beinhaltet das Verfahren ein Ausführen der Abfrage zu der Gruppe von Dateien unter der Bedingung, dass die Gruppe von Dateien nicht alle Dateien enthält, die in dem zentralen Cluster gespeichert sind. Darüber hinaus beinhaltet das Verfahren ein Speichern der Gruppe von Dateien für eine vorbestimmte Zeitspanne in dem lokalen Cluster, die im Einklang mit einer Richtlinie für die Cache-Leerung mit einem letzten Zugriff auf die Gruppe von Dateien beginnt.

[0022] Es sei von vornherein klargestellt, dass das Umsetzen der hierin angeführten Lehren nicht auf eine Cloud-Computing-Umgebung beschränkt ist, obwohl diese Offenbarung eine ausführliche Beschreibung von Cloud-Computing enthält. Stattdessen können Ausführungsformen der vorliegenden Erfindung gemeinsam mit jeder beliebigen Art von jetzt bekannter oder später erfundener Datenverarbeitungsumgebung umgesetzt werden.

[0023] Cloud-Computing ist ein Servicebereitstellungsmodell zum ermöglichen eines problemlosen bedarfsgesteuerten Netzwerkzugriffs auf einen gemeinsam genutzten Pool von konfigurierbaren Datenverarbeitungsressourcen (z.B. Netzwerke, Netzwerkbandbreite, Server, Verarbeitung, Hauptspeicher, Speicher, Anwendungen, virtuelle Maschinen und Dienste), die mit minimalem Verwaltungsaufwand bzw. minimaler Interaktion mit einem Anbieter des Service schnell bereitgestellt und freigegeben werden können. Dieses Cloud-Modell kann mindestens fünf Eigenschaften enthalten, mindestens drei Dienstmodelle und mindestens vier Implementierungsmodelle.

[0024] Bei den Eigenschaften handelt es sich um die Folgenden:

On-Demand Self-Service: Ein Cloud-Nutzer kann einseitig automatisch nach Bedarf für Datenverarbeitungsfunktionen wie Serverzeit und Netzwerkspeicher sorgen, ohne dass eine menschliche Interaktion mit dem Anbieter der Dienste erforderlich ist.

[0025] Broad Network Access: Es sind Funktionen über ein Netzwerk verfügbar, auf die durch Standardmechanismen zugegriffen wird, welche die Verwendung durch heterogene Thin- oder Thick-Client-Plattformen (z.B. Mobiltelefone, Laptops und PDAs) unterstützen.

[0026] Resource-Pooling: Die Datenverarbeitungsressourcen des Anbieters werden zusammenge schlossen, um mehreren Nutzern unter Verwendung eines Multi-Tenant-Modells zu dienen, wobei verschiedene physische und virtuelle Ressourcen dynamisch nach Bedarf zugewiesen und neu zugewiesen werden. Es gibt eine gefühlte Standortabhängigkeit, da der Nutzer allgemein keine Kontrolle bzw. Kenntnis über den genauen Standort der bereitgestellten Ressourcen hat, aber in der Lage sein kann, einen Standort auf einer höheren Abstraktionsebene festzulegen (z.B. Land, Staat oder Rechenzentrum).

[0027] Rapid Elasticity: Funktionen können für eine schnelle horizontale Skalierung (Scale-out) schnell und elastisch bereitgestellt werden, in einigen Fällen auch automatisch, und für ein schnelles Scale-in schnell freigegeben werden. Für den Nutzer erscheinen die für das Bereitstellen verfügbaren Funktionen häufig unbegrenzt und sie können jederzeit in jeder beliebigen Menge gekauft werden.

[0028] Measured Service: Cloud-Systeme steuern und optimieren die Verwendung von Ressourcen automatisch, indem sie eine Messfunktion auf einer gewissen Abstraktionsebene nutzen, die für die Art von Dienst geeignet ist (z.B. Speicher, Verarbeitung, Bandbreite sowie aktive Benutzerkonten). Der Ressourcen-Verbrauch kann überwacht, gesteuert und gemeldet werden wodurch sowohl für den Anbieter als auch für den Nutzer des verwendeten Dienstes Transparenz geschaffen wird.

[0029] Bei den Dienstmodellen handelt es sich um die Folgenden:

Software as a Service (SaaS): Die dem Nutzer bereitgestellte Funktion besteht darin, die in einer Cloud-Infrastruktur laufenden Anwendungen des Anbieters zu verwenden. Die Anwendungen sind über eine Thin-Client-Schnittstelle wie einen Web-Browser (z.B. auf dem Web beruhende E-Mail) von verschiedenen Client-Einheiten her zugänglich. Der Nutzer verwaltet bzw. steuert die zugrunde liegende Cloud-Infrastruktur nicht, darunter das Netzwerk, Server, Betriebssysteme, Speicher bzw. sogar einzelne Anwendungsfunktionen, mit der möglichen Ausnahme von eingeschränkten benutzerspezifischen Anwendungskonfigurationseinstellungen.

[0030] Platform as a Service (PaaS): Die dem Nutzer bereitgestellte Funktion besteht darin, durch einen Nutzer erstellte bzw. erhaltene Anwendungen, die unter Verwendung von durch den Anbieter unterstützten Programmiersprachen und Tools erstellt wurden, in der Cloud-Infrastruktur einzusetzen. Der Nutzer verwaltet bzw. steuert die zugrunde liegende Cloud-Infrastruktur nicht, darunter Netzwerke, Server, Betriebssysteme bzw. Speicher, hat aber die Kontrolle über die eingesetzten Anwendungen und

möglicherweise über Konfigurationen des Application Hosting Environment.

[0031] Infrastructure as a Service (IaaS): Die dem Nutzer bereitgestellte Funktion besteht darin, das Verarbeiten, Speicher, Netzwerke und andere grundlegende Datenverarbeitungsressourcen bereitzustellen, wobei der Nutzer in der Lage ist, beliebige Software einzusetzen und auszuführen, zu der Betriebssysteme und Anwendungen gehören können. Der Nutzer verwaltet bzw. steuert die zugrunde liegende Cloud-Infrastruktur nicht, hat aber die Kontrolle über Betriebssysteme, Speicher, eingesetzte Anwendungen und möglicherweise eine eingeschränkte Kontrolle über ausgewählte Netzwerkkomponenten (z.B. Host-Firewalls).

[0032] Bei den Einsatzmodellen handelt es sich um die Folgenden:

Private Cloud: Die Cloud-Infrastruktur wird einzig und allein für eine Organisation betrieben. Sie kann durch die Organisation oder einen Dritten verwaltet werden und kann sich in den eigenen Räumen oder in fremden Räumen befinden.

[0033] Community Cloud: Die Cloud-Infrastruktur wird von mehreren Organisationen gemeinsam genutzt und unterstützt eine spezielle Benutzergemeinschaft, die gemeinsame Angelegenheiten hat (z.B. Mission, Sicherheitsanforderungen, Richtlinien sowie Überlegungen bezüglich der Einhaltung von Vorschriften). Sie kann durch die Organisationen oder einen Dritten verwaltet werden und kann in den eigenen Räumen oder fremden Räumen stehen.

[0034] Public Cloud: Die Cloud-Infrastruktur wird der allgemeinen Öffentlichkeit oder einer großen Industriegruppe zur Verfügung gestellt und sie gehört einer Cloud-Dienste verkaufenden Organisation.

[0035] Hybrid Cloud: Die Cloud-Infrastruktur ist eine Zusammensetzung aus zwei oder mehreren Clouds (privat, Benutzergemeinschaft oder öffentlich), die zwar einzelne Einheiten bleiben, aber durch eine standardisierte oder proprietäre Technologie miteinander verbunden sind, die Daten- und Anwendungsportierbarkeit ermöglicht (z.B. Cloud-Zielgruppenverteilung für den Lastenausgleich zwischen Clouds).

[0036] Eine Cloud-Computing-Umgebung ist dienstorientiert mit Fokus auf Statusunabhängigkeit, geringer Kopplung, Modularität und semantischer Interoperabilität. Im Herzen von Cloud-Computing liegt eine Infrastruktur, die ein Netzwerk aus zusammen geschalteten Knoten aufweist.

[0037] Unter Bezugnahme auf **Fig. 1** wird eine schematische Darstellung eines Beispiels für einen Cloud-Computing-Knoten gezeigt. Ein Cloud-Computing-

Knoten **10** ist lediglich ein Beispiel für einen geeigneten Cloud-Computing-Knoten und nicht als eine wie auch immer geartete Beschränkung von Verwendungsumfang oder Funktionalität von Ausführungsformen der hier beschriebenen Erfindung gedacht. Unabhängig davon kann der Cloud-Computing-Knoten **10** mit jeder beliebigen hier dargelegten Funktionalität realisiert sein und/oder diese durchführen.

[0038] In dem Cloud-Computing-Knoten **10** gibt es ein Computersystem/einen Server **12**, das bzw. der mit zahlreichen anderen Universal- oder Spezial-Datenverarbeitungssystemumgebungen oder -konfigurationen betrieben werden kann. Beispiele bekannter Datenverarbeitungssysteme, -umgebungen und/oder -konfigurationen, die für eine Verwendung mit einem Computersystem/Server **12** geeignet sein könnten, sind, ohne jedoch darauf beschränkt zu sein, Personal-Computersysteme, Server-Computersysteme, Thin Clients, Thick Clients, Handheld- oder Laptop-Einheiten, Mehrprozessorsysteme, Systeme auf Grundlage von Mikroprozessoren, Set-Top-Boxen, programmierbare Unterhaltungselektronik, Netzwerk-PCs, Mini-Computersysteme, Mainframe-Computersysteme sowie verteilte Cloud-Computing-Umgebungen, die ein beliebiges der obigen Systeme oder eine beliebige der obigen Einheiten beinhalten, und dergleichen.

[0039] Das Computersystem/der Server **12** lässt sich im allgemeinen Zusammenhang von Anweisungen beschreiben, die durch ein Computersystem ausführbar sind, wie z.B. Programmmoduln, die durch ein Computersystem ausgeführt werden. Allgemein können Programmmoduln Routinen, Programme, Objekte, Komponenten, Logik, Datenstrukturen usw. beinhalten, die bestimmte Aufgaben durchführen oder bestimmte abstrakte Datentypen realisieren. Das Computersystem/der Server **12** kann in verteilten Cloud-Computing-Umgebungen eingesetzt werden, wo Aufgaben von entfernt angeordneten Verarbeitungseinheiten durchgeführt werden, die über ein Datenübertragungsnetzwerk verbunden sind. In einer verteilten Cloud-Computing-Umgebung können sich Programmmoduln sowohl in lokalen als auch in entfernt angeordneten Computersystem-Speichermedien wie beispielsweise Arbeitsspeichereinheiten befinden.

[0040] Fig. 1 zeigt das Computersystem/den Server **12** in dem Cloud-Computing-Knoten **10** als eine Universal-Datenverarbeitungseinheit. Die Komponenten des Computersystems/Servers **12** können eine(n) oder mehrere Prozessoren oder Verarbeitungseinheiten **16**, einen Systemarbeitsspeicher **28** und einen Bus **18** beinhalten, der verschiedene Systemkomponenten wie z.B. den Systemarbeitsspeicher **28** mit dem Prozessor **16** verbindet, ohne jedoch darauf beschränkt zu sein.

[0041] Der Bus **18** steht für mindestens eine von beliebigen mehreren Arten von Busstrukturen, z.B. ein Speicherbus oder ein Arbeitsspeicher-Controller, ein Peripheriebus, ein Accelerated Graphics Port (AGP) und ein Prozessor- oder lokaler Bus, wobei eine beliebige aus einer Vielfalt von Busarchitekturen verwendet werden kann. Beispielhaft und nicht als Beschränkung zu verstehen, beinhalten derartige Architekturen einen ISA-Bus (Industry Standard Architecture), einen MCA-Bus (Micro Channel Architecture), einen EISA-Bus (Enhanced ISA), einen lokalen VESA-Bus (Video Electronics Standards Association) und einen PCI-Bus (Peripheral Component Interconnect).

[0042] Das Computersystem/der Server **12** beinhaltet üblicherweise eine Vielfalt von Medien, die durch ein Computersystem lesbar sind. Derartige Medien können beliebige verfügbare Medien sein, auf die das Computersystem/der Server **12** zugreifen kann, und sowohl flüchtige als auch nicht flüchtige, entfernbare als auch nicht entfernbare Medien beinhalten.

[0043] Der Systemarbeitsspeicher **28** kann ein durch ein Computersystem lesbare Medium in Form eines flüchtigen Arbeitsspeichers wie z.B. eines RAM **30** (Random Access Memory) und/oder eines Caches **32** beinhalten. Das Computersystem/der Server **12** kann des Weiteren andere entfernbare/nicht entfernbare, flüchtige/nicht flüchtige Computersystem-Speichermedien beinhalten. Nur um ein Beispiel zu geben, kann ein Speichersystem **34** zum Lesen von und Schreiben auf ein nicht entfernbare, nicht flüchtiges magnetisches Medium bereitgestellt werden (das nicht gezeigt ist und üblicherweise als eine „Festplatte“ bezeichnet wird). Obwohl hier nicht abgebildet, können ein Magnetplattenlaufwerk zum Lesen von und Schreiben auf eine entfernbare, nicht flüchtige Magnetplatte (z.B. eine „Diskette“) sowie ein optisches Plattenlaufwerk zum Lesen von oder Schreiben auf eine entfernbare, nicht flüchtige optische Platte wie z.B. einen CD-ROM, einen DVD-ROM oder ein anderes optisches Medium bereitgestellt werden. In diesen Fällen kann jedes Laufwerk über eine oder mehrere Datenmedienschnittstellen mit dem Bus **18** verbunden sein. Wie weiter unten ausführlicher dargestellt und beschrieben, kann der Arbeitsspeicher **28** mindestens ein Programmprodukt mit einem Satz von (z.B. mindestens einem) Programmmodulen beinhalten, die konfiguriert sind, um die Funktionen von Ausführungsformen der Erfindung durchzuführen.

[0044] Zum Beispiel und ohne als Beschränkung gedacht zu sein, kann ein Programm/Dienstprogramm **40** mit einem Satz von (mindestens einem) Programmmodulen **42** in dem Arbeitsspeicher **28** gespeichert sein, ebenso wie ein Betriebssystem, ein oder mehrere Anwendungsprogramme, andere Programmmoduln und Programmdaten. Das Betriebssystem, das eine oder die mehreren Anwendungs-

programme, die anderen Programmmodulen und die Programmdaten oder eine Kombination hiervon können jeweils eine Realisierung einer Netzwerkumgebung beinhalten. Die Programmmodulen **42** führen im Allgemeinen die Funktionen und/oder Verfahrensweisen von Ausführungsformen der hier beschriebenen Erfindung aus.

[0045] Das Computersystem/der Server **12** kann zudem mit einer oder mehreren externen Einheiten **14** Daten austauschen, z.B. mit einer Tastatur, einer Zeigeeinheit, einer Anzeige **24** usw.; mit einer oder mehreren Einheiten, die einem Benutzer gestatten, mit dem Computersystem/Server **12** zu interagieren; und/oder mit beliebigen Einheiten (z.B. Netzwerkkarte, Modem usw.), die dem Computersystem/Server **12** ermöglichen, mit einer oder mehreren anderen Datenübertragungseinheiten Daten auszutauschen. Eine derartige Datenübertragung kann über Eingabe-/Ausgabe-Schnittstellen (E/A-Schnittstellen) **22** erfolgen. Des Weiteren kann das Computersystem/der Server **12** über einen Netzwerkadapter **20** mit einem oder mehreren Netzwerken Daten austauschen, z.B. mit einem lokalen Netzwerk (Local Area Network, LAN), einem Weitverkehrsnetzwerk (Wide Area Network, WAN) und/oder einem öffentlichen Netzwerk (z.B. dem Internet). Wie dargestellt, tauscht der Netzwerkadapter **20** über den Bus **18** Daten mit den anderen Komponenten des Computersystems/Servers **12** aus. Dabei sollte klar sein, dass - obwohl sie hier nicht abgebildet sind - auch andere Hardware- und/oder Software-Komponenten in Verbindung mit dem Computersystem/Server **12** verwendet werden könnten. Beispiele hierfür sind, ohne darauf beschränkt zu sein, Mikrocode, Einheitentreiber, redundante Verarbeitungseinheiten, externe Plattenlaufwerksstapel, RAID-Systeme, Bandlaufwerke und Datenarchivierungsspeichersysteme usw.

[0046] Unter Bezugnahme auf **Fig. 2** ist eine veranschaulichende Cloud-Computing-Umgebung **50** dargestellt. Wie gezeigt ist, weist die Cloud-Computing-Umgebung **50** einen oder mehrere Cloud-Computing-Knoten **10** auf, mit denen von Cloud-Nutzern verwendete lokale Datenverarbeitungseinheiten wie der elektronische Assistent (PDA, Personal Digital Assistant) oder das Mobiltelefon **54A**, der Desktop Computer **54B**, der Laptop Computer **54C** und/oder das Automobil-Computer-System **54N** Daten austauschen können. Die Knoten **10** können miteinander Daten austauschen. Sie können physisch oder virtuell in ein oder mehrere Netzwerke wie private, Benutzergemeinschafts-, öffentliche oder hybride Clouds gruppiert werden (nicht gezeigt), wie vorstehend beschrieben wurde, oder in eine Kombination daraus. Dies ermöglicht es der Cloud-Computing-Umgebung **50**, Infrastruktur, Plattformen und/oder Software als Dienst anzubieten, für die ein Cloud-Nutzer keine Ressourcen auf einer lokalen Datenverarbeitungs-

einheit vorhalten muss. Dabei sollte klar sein, dass die in **Fig. 2** gezeigten Arten von Datenverarbeitungseinheiten **54A** bis **N** lediglich veranschaulichend sein sollen und dass die Datenverarbeitungsknoten **10** und die Cloud-Computing-Umgebung **50** über eine beliebige Art Netzwerk und/oder über eine beliebige Art von über ein Netzwerk aufrufbarer Verbindung (z.B. unter Verwendung eines Web-Browsers) mit einer beliebigen Art von computergestützter Einheit Daten austauschen können.

[0047] Unter Bezugnahme auf **Fig. 3** wird ein Satz von funktionsbezogenen Abstraktionsschichten gezeigt, der von der Cloud-Computing-Umgebung **50** (**Fig. 2**) bereitgestellt wird. Dabei sollte von Anfang an klar sein, dass die in **Fig. 3** gezeigten Komponenten, Schichten und Funktionen lediglich zur Veranschaulichung gedacht und Ausführungsformen der Erfindung nicht darauf beschränkt sind. Wie abgebildet ist, werden die folgenden Schichten und entsprechenden Funktionen bereitgestellt.

[0048] Eine Hardware- und Software-Schicht **60** enthält Hardware- und Software-Komponenten. Zu Beispielen für Hardware-Komponenten gehören: Mainframe Computer **61**; auf der RISC-(Reduced Instruction Set Computer) Architektur beruhende Server **62**; Server **63**; Blade-Server **64**; Speichereinheiten **65**; und Netzwerke sowie Netzwerkkomponenten **66**. In einigen Ausführungsformen beinhalten Software-Komponenten eine Netzwerk-Anwendungsserver-Software **67** und eine Datenbank-Software **68**.

[0049] Eine Virtualisierungsschicht **70** stellt eine Abstraktionsschicht bereit, welche die folgenden Beispiele für virtuelle Einheiten zur Verfügung stellen kann: virtuelle Server **71**; virtueller Speicher **72**; virtuelle Netzwerke **73**, die virtuelle private Netzwerke aufweisen; virtuelle Anwendungen und Betriebssysteme **74**; sowie virtuelle Clients **75**.

[0050] In einem Beispiel kann eine Verwaltungsschicht **80** die nachfolgend beschriebenen Funktionen bereitstellen. Eine Ressourcenbereitstellungs-funktion **81** stellt eine dynamische Beschaffung von Datenverarbeitungs- und anderen Ressourcen bereit, mit denen Aufgaben innerhalb der Cloud-Computing-Umgebung durchgeführt werden. Messungs- und Preisermittlungsfunktionen **82** stellen eine Kostenfassung bei der Nutzung von Ressourcen innerhalb der Cloud-Computing-Umgebung sowie eine Fakturierung bzw. Abrechnung für den Verbrauch dieser Ressourcen bereit. In einem Beispiel können diese Ressourcen Anwendungs-Software-Lizenzen aufweisen. Die Sicherheit stellt die Identitätsüberprüfung für Cloud-Nutzer und Aufgaben sowie Schutz für Daten und andere Ressourcen bereit. Ein Benutzerportal **83** stellt Nutzern und Systemadministratoren den Zugang zu der Cloud-Computing-Umgebung bereit. Eine Verwaltung des Dienstumfangs **84** stellt die

Zuordnung und Verwaltung von Cloud-Computing-Ressourcen bereit, so dass die benötigten Dienstziele erreicht werden. Ein Planen und Erfüllen von Vereinbarungen zum Dienstumfang (SLA, Service Level Agreement) **85** stellt die Anordnung vorab und die Beschaffung von Cloud-Computing-Ressourcen, für die eine zukünftige Anforderung vorausgesehen wird, gemäß einem SLA bereit.

[0051] Eine Arbeitslastschicht **90** stellt Beispiele für die Funktionalität bereit, für welche die Cloud-Computing-Umgebung verwendet werden kann. Zu Beispielen für Arbeitslasten und Funktionen, die von dieser Schicht bereitgestellt werden können, gehören: Abbildung und Navigation **91**; Software-Entwicklung und Lebenszyklusverwaltung **92**; Bereitstellung von Ausbildung in virtuellen Klassenzimmern **93**; Datenanalytikverarbeitung **94**; Transaktionsverarbeitung **95**; und kognitive Datei-/Objektverwaltung für verteilte Speicherumgebungen **96**.

[0052] Unter Bezugnahme auf **Fig. 4** ist ein mehrstufiges Speichersystem **400** gemäß einer Ausführungsform gezeigt, das bei manchen Ansätzen für einen öffentlichen, mehrstufigen Objektspeicher stehen kann. Es sei darauf hingewiesen, dass einige der in **Fig. 4** gezeigten Elemente gemäß verschiedenen Ausführungsformen als Hardware und/oder Software realisiert sein können. Das Speichersystem **400** kann einen Speichersystem-Manager **412** für eine Datenübertragung mit einer Mehrzahl von Medien in mindestens einer höheren Speicherstufe **402** und mindestens einer niedrigeren Speicherstufe **406** beinhalten. Die höhere(n) Speicherstufe(n) **402** kann bzw. können vorzugsweise Medien **404** mit wahlfreiem und/oder direktem Zugriff wie z.B. einen nicht flüchtigen Arbeitsspeicher (Non-Volatile Memory, NVM), einen Halbleiterspeicher in Halbleiterlaufwerken (Solid State Drives, SSDs), einen Flash-Speicher, SSD-Arrays, Flash-Speicher-Arrays, Festplatten in Festplattenlaufwerken (Hard Disk Drives, HDDs) usw. und/oder andere hier erwähnte oder nach dem Stand der Technik bekannte Medien beinhalten. Die niedrigere(n) Speicherstufe(n) **406** kann bzw. können vorzugsweise ein oder mehrere weniger leistungsstarke Speichermedien **408** wie z.B. langsamer zugreifende HDDs, Medien mit sequenziellem Zugriff wie Magnetband in Bandlaufwerken und/oder optische Medien usw. und/oder andere hier bekannte oder nach dem Stand der Technik bekannte Medien beinhalten. Eine oder mehrere zusätzliche Speicherstufen **416** können eine beliebige Kombination von Speichermedien beinhalten, wie dies durch einen Konstrukteur des Systems **400** gewünscht wird. Darüber hinaus kann jede der höheren Speicherstufen **402** und/oder der niedrigeren Speicherstufen **406** eine wie auch immer geartete Kombination von Speichereinheiten und/oder Speichermedien beinhalten.

[0053] Der Speichersystem-Manager **412** kann mit den Speichermedien **404, 408** in der bzw. den höheren Speicherstufe(n) **402** und niedrigeren Speicherstufe(n) **406** über ein Netzwerk **410** wie beispielsweise ein Speicherbereichsnetzwerk (Storage Area Network, SAN), wie es z.B. in **Fig. 4** gezeigt wird, oder aber mit einem anderweitigen geeigneten Netzwerktyp Daten austauschen. Der Speichersystem-Manager **412** kann außerdem über eine Host-Schnittstelle **414**, die ein Teil des Speichersystem-Managers **412** sein kann, mit einem oder mehreren (nicht gezeigten) Host-Systemen Daten austauschen. Der Speichersystem-Manager **412** und/oder jedwede andere Komponente des Speichersystems **400** kann in Hardware und/oder Software realisiert sein und einen (nicht gezeigten) Prozessor zum Ausführen von Befehlen eines nach dem Stand der Technik bekannten Typs verwenden, z.B. eine Zentraleinheit (Central Processing Unit, CPU), ein feldprogrammierbares Gatter-Array (Field Programmable Gate Array, FPGA), eine anwendungsspezifische integrierte Schaltung (Application Specific Integrated Circuit, ASIC) usw. Selbstverständlich kann eine beliebige Anordnung eines Speichersystems verwendet werden, wie für den Fachmann beim Lesen der vorliegenden Beschreibung offensichtlich sein dürfte.

[0054] Bei weiteren Ausführungsformen kann das Speichersystem **400** eine beliebige Anzahl von Datenspeicherstufen beinhalten und innerhalb einer jeden Speicherstufe dieselben oder verschiedene Speichermedien beinhalten. Zum Beispiel kann jede Datenspeicherstufe dieselbe Art von Speichermedien wie z.B. HDDs, SSDs, Medien mit sequenziellem Zugriff (Band in Bandlaufwerken, optische Platte in optischen Plattenlaufwerken usw.), Direktzugriffsmedien (CD-ROM, DVD-ROM usw.) oder eine beliebige Kombination von Medienspeichertypen beinhalten. Bei einer solchen Konfiguration kann eine höhere Speicherstufe **402** eine Mehrheit von SSD-Speichermedien (bis hin zu und einschließlich aller SSD-Speichermedien) zum Speichern von Daten in einer leistungsstärkeren Speicherumgebung beinhalten, und die verbleibenden Speicherstufen wie z.B. die niedrigere Speicherstufe **406** und die zusätzlichen Speicherstufen **416** können eine beliebige Kombination von SSDs, HDDs, Bandlaufwerken usw. zum Speichern von Daten in einer weniger leistungsstarken Speicherumgebung beinhalten. Auf diese Weise können Daten, auf die häufiger zugegriffen wird, Daten mit einer höheren Priorität, Daten, auf die schneller zugegriffen werden muss usw. in der höheren Speicherstufe **402** gespeichert werden, während Daten ohne eines dieser Merkmale in den zusätzlichen Speicherstufen **416** wie z.B. der niedrigeren Speicherstufe **406** gespeichert werden können. Selbstverständlich sollte ein Fachmann beim Lesen der vorliegenden Beschreibungen in der Lage sein, gemäß den hier dargelegten Ausführungsformen viele andere Kombinationen von Speichermedientypen für eine

Realisierung in verschiedenen Speicherschemata zu entwickeln.

[0055] Bei einer bestimmten Ausführungsform kann das Speichersystem **400** eine Kombination von SSDs und HDDs beinhalten, wobei die höhere Speicherstufe **402** SSDs (und möglicherweise einen wie auch immer gearteten Pufferspeicher) und die niedrigere Speicherstufe **406** HDDs (und möglicherweise einen wie auch immer gearteten Pufferspeicher) beinhaltet. Gemäß einer weiteren Ausführungsform kann das Speichersystem **400** eine Kombination von SSDs und Magnetband mit Magnetbandlaufwerken beinhalten, wobei die höhere Speicherstufe **402** SSDs (und möglicherweise einen wie auch immer gearteten Pufferspeicher) und die niedrigere Speicherstufe **406** Magnetband (und möglicherweise einen wie auch immer gearteten Pufferspeicher) sowie Magnetbandlaufwerke zum Zugreifen auf Daten auf den Magnetbändern beinhaltet. Bei einer weiteren Ausführungsform kann das Speichersystem **400** eine Kombination von HDDs und Magnetband beinhalten, wobei die höhere Speicherstufe **402** HDDs (und möglicherweise einen wie auch immer gearteten Pufferspeicher) und die niedrigere Speicherstufe **406** Magnetband (und möglicherweise einen wie auch immer gearteten Pufferspeicher) beinhaltet.

[0056] Unter Bezugnahme auf **Fig. 5A** wird ein zentrales, verteiltes und geclustertes Dateisystem **500** (im Folgenden „verteiltes System **500**“) gemäß einer Ausführungsform gezeigt. Das verteilte System **500** kann eine beliebige Anzahl von Dateien und/oder Objekten (im Folgenden „Dateien **504**“) beinhalten, die Informationen und/oder Daten speichern und/oder beinhalten und auf die ein oder mehrere Benutzer **502** des verteilten Systems **500** zugreifen können. Darüber hinaus beinhaltet das verteilte System **500** einen zentralen Cluster **506**, der als ein globales Daten-Repository konfiguriert ist, eine Mehrzahl von lokalen Cache-Clustern **508a**, **508b**, ..., **508n** (im Folgenden „lokale Cluster **508**“, wenn auf sie als Gruppe Bezug genommen wird), die sich in geografischer Hinsicht von dem zentralen Cluster unterscheiden, und ein oder mehrere Netzwerke **510**, welche die verschiedenen lokalen Cluster **508** mit dem zentralen Cluster **506** verbinden. Wie für den Fachmann beim Lesen der vorliegenden Beschreibungen offensichtlich sein dürfte, kann eine beliebige Art von Netzwerk (en) **510** verwendet werden, z.B. das Internet, ein WAN, ein LAN, ein SAN usw., ohne jedoch darauf beschränkt zu sein.

[0057] Zusätzlich beinhalten der zentrale Cluster **506** und jeder der lokalen Cluster **508** eine Hardware-Verarbeitungsschaltung, die konfiguriert ist, um ihr bereitgestellte Programmanweisungen auszuführen. Andere Hardware- und/oder Software-Komponenten, die hier nicht ausdrücklich beschrieben sind, können ebenfalls in dem zentralen Cluster **506** und/oder ei-

nem oder mehreren der lokalen Cluster **508** enthalten sein, wie dem Fachmann bekannt sein dürfte.

[0058] Bei einer Ausführungsform können eine oder mehrere der Hardware-Komponenten innerhalb des zentralen Clusters **506** und/oder einem oder mehreren der lokalen Cluster **508** redundante Komponenten aufweisen, die parallel angeordnet sind, um in Fällen eine redundante Funktionalität durchzuführen, in denen eine primäre Hardware-Komponente ausfällt, nicht mehr mit Strom versorgt wird usw. und nicht verfügbar ist, um die ihr zugewiesene(n) Aufgabe(n) durchzuführen.

[0059] Zusätzlich beinhalten der zentrale Cluster **506** und jeder der lokalen Cluster **508** eine oder mehrere Arten von durch einen Computer lesbaren Speichermedien **512**. In dem zentralen Cluster **506** und den verschiedenen lokalen Clustern **508** kann jede Art von der durch einen Computer lesbaren Speichermedien **512** verwendet werden, z.B., ohne jedoch darauf beschränkt zu sein, Speicher-Einheiten mit einem nicht flüchtigen Arbeitsspeicher (NVM), Direktzugriffsspeicher-Einheiten (Direct Access Storage Devices, DASDs), Arbeitsspeicher-Einheiten usw. Alle beliebigen geeigneten NVM-Einheiten können verwendet werden, z.B. ein Flash-Speicher, ein RAM, ein löscherbarer programmierbarer Nur-Lese-Speicher (Erasable Programmable Read-Only Memory, EPROM), Halbleiterlaufwerkseinheiten (SSDs) usw. Darüber hinaus können beliebige DASDs verwendet werden, z.B. HDDs, Bandmedien zur Verwendung mit einem Bandlaufwerk, optische Laufwerke usw. Zusätzlich kann in den durch einen Computer lesbaren Speichermedien **512** ein Cache oder Puffer vorhanden sein, um vor einer Speicherung in den durch einen Computer lesbaren Speichermedien **512** eine Datenbereitstellung durchzuführen.

[0060] Die in dem zentralen Cluster **506** gespeicherten Dateien **504** und die in den lokalen Clustern **508** gespeicherten Dateien können Informationen von unterschiedlicher Art (Textdaten, Videodaten, Audiodaten, unstrukturierte Daten usw.), Größe, von unterschiedlichem Gehalt oder Inhalt beinhalten, wie dem Fachmann klar sein dürfte. Darüber hinaus können bei manchen Ansätzen Metadaten, die den Dateien **504** zugehörig sind, mindestens einige der Merkmale der verschiedenen Dateien **504** angeben. Allerdings sind die in dem zentralen Cluster **506** gespeicherten Dateien **504** bei herkömmlichen Ansätzen nicht in einer wie auch immer gearteten aussagekräftigen Weise organisiert, die ein effizientes Durchsuchen der Dateien bereitstellen würde.

[0061] Unter Bezugnahme auf **Fig. 5B** wird ein Prozess beschrieben, der ein kognitives Filtern der Dateien **504** in Untercontainer des zentralen Clusters **506** gemäß einer Ausführungsform bereitstellt. Bei verschiedenen Ansätzen verbessert und fördert dieser

Prozess die Fähigkeit, die in dem zentralen Cluster **506** gespeicherten Dateien **504** zu durchsuchen und/oder zu filtern, um Dateien zu ermitteln, die mit größerer Wahrscheinlichkeit relevant für eine Abfrage sind.

[0062] Neben proprietären Algorithmen, die von großen Technologieunternehmen wie Google®, Microsoft®, Baidu®, Tencent® usw. bereitgestellt werden, kann jeder beliebige bekannte Suchalgorithmus verwendet werden, z.B., ohne jedoch darauf beschränkt zu sein, der Rabin-Karp-Zeichenketten-Suchalgorithmus, ein Suchalgorithmus auf Grundlage eines Zustandsautomaten, der Knuth-Morris-Pratt-Algorithmus usw.

[0063] Um die Dateien **504** in dem zentralen Cluster **506** zu organisieren, werden alle in den Dateien **504** gespeicherten unstrukturierten Daten in Text in Dateien auf Textgrundlage oder in mit Text annotierte Dateien konvertiert (im Falle von Bildern und/oder Videodateien können Textannotationen den Originaldateien als Metadaten hinzugefügt werden), die unter Verwendung einer Mehrzahl von herkömmlichen Suchalgorithmen einfach und effizient durchsuchbar sind. Im weiteren Verlauf dieser Beschreibungen werden reine Textdateien und mit Text annotierte Dateien gemischten Inhalts als Dateien auf Textgrundlage bezeichnet. Sobald eine Originaldatei aus den Dateien **504** in eine Datei auf Textgrundlage konvertiert wird, wird eine Verbindung zwischen der ursprünglichen Datei und der erzeugten Datei auf Textgrundlage hergestellt, so dass jede beliebige Suche, welche die Datei auf Textgrundlage zurückgibt, zu der Originaldatei zurückverfolgt werden kann.

[0064] Bei einer Ausführungsform können eine oder mehrere Anwendungsprogrammschnittstellen (Application Program Interfaces, APIs) verwendet werden, um die unstrukturierten Daten in Daten auf Textgrundlage zu konvertieren. Auf jede Art von unstrukturierten Daten kann eine andere Art von API angewendet werden, um die unstrukturierten Daten in Daten auf Textgrundlage zu konvertieren. Bei einer Ausführungsform können IBM® BlueMix® Watson-APIs für die Konvertierung in Daten auf Textgrundlage verwendet werden.

[0065] Gemäß einer Ausführungsform können alle in dem zentralen Cluster **506** gespeicherten Dateien **504** unter Verwendung eines oder mehrerer Suchalgorithmen durchsucht werden, die zum Durchsuchen eines bestimmten Inhalts oder bestimmter unstrukturierter Daten innerhalb der Dateien **504** ausgestaltet sind. In einem Beispiel kann ein Bildsuchalgorithmus konfiguriert sein, um im Besonderen Bilddateien zu durchsuchen und ein bzw. mehrere in der Suche angegebene Bilder zurückzugeben, so dass jede der Dateien, welche die Bilddaten enthalten, unter Verwendung dieses Algorithmus durchsucht werden kann, und diejenigen Dateien, welche das bzw. die

angegebenen Bilder enthalten, durch die Bildsuche zurückgegeben werden. In einem weiteren Beispiel wird auf Dateien, welche die Bilddaten nicht enthalten, dieser Bildsuchalgorithmus nicht angewendet.

[0066] In einem weiteren Beispiel kann ein Audiosuchalgorithmus konfiguriert sein, um im Besonderen Audiodateien zu durchsuchen und in der Suche angegebene Audiodaten zurückzugeben, so dass jede der Dateien, die Audiodaten enthalten, unter Verwendung dieses Algorithmus durchsucht werden kann, und diejenigen Dateien, welche die angegebenen Audiodaten enthalten, durch die Audiosuche zurückgegeben werden. In einem weiteren Beispiel wird auf Dateien, welche die Audiodaten nicht enthalten, dieser Audiosuchalgorithmus nicht angewendet.

[0067] Wie für den Fachmann beim Lesen der vorliegenden Beschreibungen offensichtlich sein dürfte, können in verschiedenen zusätzlichen Ausführungsformen beim Durchsuchen und Organisieren der in dem zentralen Cluster **506** gespeicherten Dateien **504** selbstverständlich viele andere inhaltspezifische Suchalgorithmen und/oder Suchalgorithmen für unstrukturierte Daten verwendet werden, z.B. jene, die für das Durchsuchen von kundenspezifischen und/oder proprietären Datenformen ausgestaltet sind.

[0068] Die Dateien auf Textgrundlage und/oder alle in dem zentralen Cluster **506** gespeicherten Dateien **504** werden analysiert, um eine oder mehrere relevante Kategorien für jede der Dateien auf Textgrundlage und/oder alle in dem zentralen Cluster **506** gespeicherten Dateien zu ermitteln, so dass für die verschiedenen Dateien **504** (und letztlich die zugehörigen Dateien) ein Filtern und Gruppieren durchgeführt werden kann. Die Relevanz der Kategorien wird auf Grundlage von Interessen eines bestimmten Benutzers ausgewählt, z.B. einer Art von Geschäft, an dem der Benutzer beteiligt ist, einem geografischen Standort (z.B. des Benutzers, einer Privatadresse, eines Unternehmens usw.), eines Datums der Abfrage usw.

[0069] In verschiedenen nicht beschränkenden Beispielen können die Dateien auf Textgrundlage und/oder alle in dem zentralen Cluster **506** gespeicherten Dateien **504** gemäß einem den einzelnen Dateien zugehörigen Datum, einem geografischen Standort, der in den einzelnen Dateien erwähnt wird und/oder ihnen zugehörig ist, einem geografischen Standort einer Erzeugung der einzelnen Dateien, einem ähnlichen und/oder gemeinsamen Inhalt der einzelnen Dateien (der auf einem bzw. mehreren Schlüsselwörtern innerhalb der einzelnen Dateien beruhen kann), einer ähnlichen Verwendung der einzelnen Dateien, einer Häufigkeit eines Zugriffs auf die einzelnen Dateien gefiltert und gruppiert werden.

[0070] In einem weiteren nicht beschränkenden Beispiel können die Dateien auf Textgrundlage und/oder alle in dem zentralen Cluster **506** gespeicherten Dateien **504** gemäß einem Wert, mehreren Werten und/oder einem Bereich von Werten, die innerhalb der einzelnen Werte gespeichert sind, gefiltert und gruppiert werden, wobei sich die Werte auf einen angegebenen Parameter in der Abfrage beziehen, z.B. ein Datum, eine Währung, eine Uhrzeit, einen virtuellen oder physischen Standort, eine oder mehrere Benutzergruppen, ein Zugriffsrecht oder einen anderweitigen angegebenen Wert, der für den Erzeuger der Abfrage interessant ist. Der angegebene interessante Wert kann alles sein, was in Zusammenhang mit einem Geschäft, einer Ausbildung, einem Vorhaben und/oder Interesse von Bedeutung ist, für das bzw. die in dem verteilten Speichersystem Daten gespeichert sind. In mehreren Beispielen kann ein angegebener interessanter Wert bei einer Anwendung in der Medizinbranche eine oder mehrere Patientenkategorisierungen (z.B. Alter, Geschlecht, ethnische Zugehörigkeit, bestehende Vorerkrankungen usw.), eine oder mehrere Krankheitskategorisierungen (z.B. Krebs, Asthma, Strep-tokokken-Infektion, Arthritis usw.), eine oder mehrere Untersuchungskategorisierungen (z.B. Röntgenuntersuchung, Genuntersuchung, körperliche Untersuchung usw.) usw. sein; bei einer Anwendung in der Finanzbranche kann ein angegebener interessanter Wert ein oder mehrere Ticker-Symbole (z.B. MSFT, INTL, T usw.), eine oder mehrere Branchenkategorisierungen (z.B. Technologie, Software, Pharmazeutik, Produktion usw.), eine oder mehrere Bewertungsmaßzahlen (z.B. Marktkapitalisierung von über 1 Milliarde, Small-, Mid-, Large-Cap usw.) usw. sein; bei einer Anwendung in der Pharmabranche kann ein angegebener interessanter Wert eine oder mehrere Kategorisierungen auf Arzneimittelgrundlage (z.B. Statin, Koffein, Benzodiazepin, Fentanyl, Paracetamol, Morphin, Opiat, Oxycodon usw.), ein oder mehrere Wissenschaftler, die einer Behandlung zugehörig sind, eine oder mehrere Behandlungsanwendungen (Schmerzmittel, abschwellendes Medikament, Lokalanästhetikum usw.), eine oder mehrere Dosierung usw. sein.

[0071] Wie einem Fachmann beim Lesen der vorliegenden Beschreibungen klar sein dürfte, können bei verschiedenen Ausführungsformen ein oder mehrere spezifische interessante Werte auf Grundlage einer bestimmten Anwendung und der spezifischen Abfragen, die den bzw. die spezifischen interessanten Werte verwenden, kundenspezifisch definiert sein.

[0072] In einem weiteren nicht beschränkenden Beispiel können die Dateien auf Textgrundlage und/oder alle in dem zentralen Cluster **506** gespeicherten Dateien **504** gemäß einer Popularität der in den einzelnen Dateien enthaltenen Informationen, bezogen auf den zentralen Cluster **506** und darauf, wie häufig ei-

ne bestimmte Datei in einen der lokalen Cluster **508** dupliziert wird, gefiltert und gruppiert werden.

[0073] Gemäß einer weiteren Ausführungsform können innerhalb einer Gruppe Untergruppen erzeugt werden, und innerhalb einer bestimmten Gruppe können mehrere zusätzliche Ebenen von Untergruppen in einer Baumstruktur vorhanden sein. Auf diese Weise können Dateien, die in die bestimmte Gruppe eingesortiert werden, des Weiteren in Untergruppen eingesortiert werden, die noch feiner differenziert sind als die bestimmte Gruppe. Beim Gruppieren nach Standort können die obersten Gruppen mit dem Land gefüllt sein, gefolgt von Untergruppen für Bundesstaaten oder Provinzen, danach von Untergruppen für Städte, gefolgt von Untergruppen für Stadtteile innerhalb von Städten usw.

[0074] Bei einer Ausführungsform kann diese Filter- und Gruppierungsoperation als ein fortlaufender Hintergrundprozess ausgeführt werden, der in dem zentralen Cluster **506** ausgeführt wird, so dass beim Hinzufügen von neuen Dateien zu dem zentralen Cluster **506** diese (bei Bedarf) konvertiert und effizient und mit wenig Auswirkung auf die übrige Funktionalität des verteilten Systems **500** kategorisiert werden können. Bei einer weiteren Ausführungsform kann die Filter- und Gruppierungsoperation in regelmäßigen Abständen oder als Reaktion auf das Stattfinden eines auslösenden Ereignisses ausgeführt werden. Jedes geeignete auslösende Ereignis kann verwendet werden, z.B. ein Hinzufügen von einer oder mehreren neuen Dateien zu dem zentralen Cluster **506**, ein Modifizieren von einer oder mehreren Dateien in dem zentralen Cluster **506**, ein Hinzufügen und/oder ein Modifizieren einer vorgegebenen Schwellenanzahl von Dateien zu dem bzw. in dem zentralen Cluster **506**, eine ausdrückliche Anforderung von einem Administrator usw.

[0075] Bei einer Ausführungsform können die Dateien auf Textgrundlage und/oder alle in dem zentralen Cluster **506** gespeicherten Dateien **504** gemäß einem oder mehreren Schlüsselwörtern gefiltert und gruppiert werden. Das eine bzw. die mehreren Schlüsselwörter können auf Grundlage einer Eingabe von einer Anzahl von Benutzern, die mit den verschiedenen lokalen Clustern **508** interagieren, automatisch erzeugt werden und können für Begriffe stehen, die über eine vorbestimmte Zeitspanne hinweg am häufigsten abgefragt werden.

[0076] Wenn eine Abfrage in einem lokalen Cluster ausgeführt wird, wird die Abfrage oder eine wie auch immer geartete Angabe zu der Abfrage üblicherweise an den zentralen Cluster **506** gesendet, um zu ermitteln, welche der Dateien **504** der Abfrage entsprechen. Bei der Angabe zu der Abfrage kann es sich um die Abfrage selbst, um einen Satz von Abfragen, die zuvor empfangen wurden und/oder deren Emp-

fang in der Zukunft erwartet wird, und/oder um Interessen handeln, die durch einen oder mehrere Benutzer ausgedrückt werden, welche die Grundlage für eine künftige Abfrage bilden können und die dazu verwendet werden können, Daten für eine solche künftige Abfrage vorabzurufen. Dabei kann jede beliebige Art von Abfrage verwendet werden, z.B. eine Suchabfrage, eine Analyseabfrage, die etwas über die zugrunde liegenden Daten ermittelt, die durch die Analyseabfrage zurückgegeben werden (z.B. einen oder mehrere Aspekte, ein oder mehrere Merkmale, eine oder mehrere Ähnlichkeiten und/oder ein oder mehrere Unterschiede zwischen Daten in der Abfrage). Herkömmlicherweise werden alle Dateien **504** in den lokalen Cluster dupliziert, der die Abfrage anfordert. Bei einer weiteren herkömmlichen Ausführungsform kann die Abfrage in dem zentralen Cluster **506** ausgeführt werden, wodurch wertvolle Ressourcen des zentralen Clusters **506** aufgebraucht werden, anstatt die Arbeitslast an einen der lokalen Cluster **508** auszulagern.

[0077] Wie in Fig. 5C gezeigt, können als Reaktion auf Gruppen **516**, die in dem zentralen Cluster **506** vorab vorhanden sind und die kognitiv erzeugt werden, um gängigen Abfragen zu entsprechen, für eine bestimmte Abfrage **514** relevante Dateien schnell ermittelt und in einen lokalen Cluster **518**, von dem die Abfrage **514** stammt, dupliziert werden, anstatt alle Dateien **504** in den lokalen Cluster **518** zu duplizieren, von dem die Abfrage **514** stammt. Dies spart beträchtliche Ressourcen gegenüber herkömmlichen Prozessen, da es nicht notwendig ist, Ergebnisse unter Verwendung von Ressourcen (z.B. Arbeitsspeicherplatz, Prozessorkapazität usw.) des zentralen Clusters **506** zu filtern bzw. alle Dateien zu filtern, nachdem sie in den lokalen Cluster **518** dupliziert wurden, um zu ermitteln, welche der Dateien **504** für die Abfrage relevant sind. Dies erfolgt teilweise, weil Ressourcen des lokalen Clusters **518** nur zum Filtern eines Teilsatzes der Dateien **504** genutzt werden, z.B. denjenigen, die sich innerhalb einer oder mehrerer bestimmter Gruppen **520** in dem zentralen Cluster **506** befinden, die sich auf die Abfrage beziehen, wie durch den zentralen Cluster **506** beim Empfangen der Abfrage **514** ermittelt wurde.

[0078] Nachdem für mindestens einige der Dateien **504** des zentralen Clusters **504** ein Filtern und Gruppieren durchgeführt wird, lernen die verschiedenen Filter, die zum Gruppieren der Dateien **504** verwendet werden, auf Grundlage der Nützlichkeit von vorhandenen Gruppen und der Nutzungsmuster der lokalen Cluster **508** für Gruppen in dem zentralen Cluster **506** mit der Zeit und passen sich an. Dieses Lernen passt die Filter an, so dass sie Gruppen bereitstellen können, die Filter beinhalten, welche relevanter für die empfangenen Abfragen sind, so dass als Reaktion auf den künftigen Empfang von Abfragen weniger Ressourcen des zentralen Clusters **506** ver-

braucht werden. Auch wenn selbstverständlich nicht vollständig vorhergesagt werden kann, welche Abfragen künftig empfangen werden, sind intelligente, lernende Filter in der Lage, Gruppen bereitzustellen, die dafür verwendet werden können, mehr als 90 % aller empfangenen Abfragen zu beantworten, wobei auf die verbleibenden Abfragen mit Ergebnissen geantwortet wird, nachdem die Dateien **504** in dem zentralen Cluster **506** gefiltert wurden. Das Lernen und die Anpassung der Filter kann dazu führen, dass die verschiedenen Gruppen **516** als Reaktion auf die Änderungen an den Dateien **504** abgewandelt und/oder modifiziert werden, z.B. durch ein Hinzufügen von einer oder mehreren neuen Gruppen, ein Modifizieren der Dateien, die sich innerhalb einer oder mehrerer Gruppen befinden, ein Entfernen von einer oder mehreren vorhandenen Gruppen usw.

[0079] Bei einer weiteren Ausführungsform können die einzelnen Dateien, die in einer einzigen Gruppe zusammengefasst sind (z.B. der Gruppe **520**), als Reaktion auf die Änderungen im Laufe der Zeit abgewandelt und/oder modifiziert werden, z.B. durch ein Hinzufügen von einer oder mehreren neuen Dateien zu einer Gruppe, ein Entfernen von einer oder mehreren vorhandenen Dateien aus einer Gruppe, ein Modifizieren der Gruppen, zu denen eine bestimmte Datei gehört, usw.

[0080] Etwaige relevante Änderungen können in der Gruppierung von einzelnen Dateien und in den Gruppen **516** selbst in dem zentralen Cluster **506** berücksichtigt werden, z.B. Änderungen an Rohdaten, die den gefilterten Ergebnissen zugrunde liegen (so können Änderungen an Rohdaten z.B. dazu führen, dass eine Datei nicht mehr relevant für eine Gruppe ist und/oder dass die Datei für eine Gruppe relevant wird, zu der sie noch nicht hinzugefügt wurde), eine Maßzahl auf Leistungsgrundlage, die einen Gruppierungserfolg misst (z.B. wie erfolgreich das Gruppieren beim Einsparen von Ressourcen für Abfragen ist, die von den lokalen Clustern **508** abgeschickt wurden), eine Popularität von Dateien innerhalb einer bestimmten Gruppe (z.B. wie häufig eine bestimmte Gruppe in einen lokalen Cluster dupliziert wird, gegenüber einem normalisierten Durchschnittswert aller Gruppen), Taxonomieänderungen usw.

[0081] Bei einer weiteren Ausführungsform können Richtlinien für die Cache-Leerung verwendet werden, um zu ermitteln, welche Gruppen **516** in einem beliebigen bestimmten lokalen Cluster (oder bei einer weiteren Ausführungsform in allen lokalen Clustern **508**) verbleiben sollen und welche Gruppen gelöscht (oder überschreibbar gemacht) werden sollen, um Platz für Informationen freizugeben, auf die häufiger zugegriffen wird oder die häufiger angefordert werden. Nachdem eine Gruppe (wie z.B. die Gruppe **520**) als Reaktion auf eine Ausführung einer Abfrage in dem lokalen Cluster in einen lokalen Cluster (wie z.B. den

lokalen Cluster **518**) dupliziert wurde, kann die Gruppe zum Beispiel für eine vorbestimmte Zeitspanne, wie sie durch eine Richtlinie für die Cache-Leerung vorgegeben wird, z.B. 1 Woche, 1 Tag, 12 Stunden, 3 Stunden usw., innerhalb des lokalen Clusters beibehalten werden. Als Reaktion darauf, dass die Gruppe für eine weitere Abfrage verwendet wird (ob alle Dateien in der Gruppe oder eine Untergruppierung hiervon), kann die vorbestimmte Zeitspanne für diese bestimmte Gruppe ausgesetzt werden, so dass sie - im Gegensatz zu einer Gruppe, die nur für die erste Abfrage verwendet wird - für eine erweiterte Zeitspanne (Zeit nach der ersten Abfrage bis zur zweiten Abfrage + vorbestimmte Zeitspanne) in dem lokalen Cluster verbleibt.

[0082] Die Zeitspanne für ein Beibehalten einer Gruppe (wie z.B. der Gruppe **520**) in dem lokalen Cluster (wie z.B. dem lokalen Cluster **518**), wie sie durch eine oder mehrere Richtlinien für die Cache-Leerung vorgegeben wird, kann bei einer Ausführungsform so festgelegt werden, dass sie für einen einzigen lokalen Cluster gilt, bei einer weiteren Ausführungsform kann sie so festgelegt werden, dass sie für einen Teilsatz von lokalen Clustern gilt und bei einer weiteren Ausführungsform so, dass sie global für alle lokalen Cluster **508** gilt. Ein Administrator kann die Richtlinie für die Cache-Leerung nach Belieben festlegen.

[0083] Auf diese Weise kann eine Gruppe (wie z.B. die Gruppe **520**) einmal oder mehrmals wiederverwendet werden, nachdem eine anfängliche Abfrage ausgeführt wird, solange sie in dem lokalen Cluster (wie z.B. dem lokalen Cluster **518**) verbleibt, so dass die zugrunde liegenden Dateien der Gruppe nicht jedes Mal, wenn eine Abfrage einen Zugriff auf die Dateien der Gruppe vorsieht, erneut in die lokalen Cluster dupliziert werden müssen.

[0084] In einem in **Fig. 6** gezeigten Beispiel für ein verteiltes System **600**, das medizinische Daten **602** bereitstellt, soll eine Abfrage **606** der in dem verteilten System **600** gespeicherten Daten auf medizinische Daten ausgeführt werden, die „Röntgenbilder“ betreffen. Darüber hinaus soll der zentrale Cluster **604** die darin gespeicherten medizinischen Daten **602** bereits nach medizinischem Typ gruppiert haben, z.B. „Röntgenbilder“ **608**, „PET-Bilder“ **610**, „CT-Bilder“ **612**, „Ultraschallbilder“ **614** usw. Als Reaktion auf die Abfrage von „Röntgenbilder“ werden nur die Daten, die sich innerhalb der Gruppe „Röntgenbilder“ **608** befinden, an den lokalen Cluster **616** gesendet, von dem die Abfrage ausgegangen ist. Danach werden anstatt der Ressourcen des zentralen Clusters **604** die Ressourcen des lokalen Clusters **616** verwendet, um die Analyseabfrage der Daten auszuführen, die sich innerhalb der Gruppe „Röntgenbilder“ **608** befinden. Im Anschluss an die ursprüngliche Abfrage kann eine Richtlinie für die Cache-Leerung darüber hinaus ei-

nen Zeitpunkt bestimmen, zu dem die Daten, die sich innerhalb der Gruppe „Röntgenbilder“ **608** befinden, aus dem lokalen Cluster **616** gelöscht werden, so dass die Daten bei zusätzlichen Abfragen in dem lokalen Cluster **616** wiederverwendet werden können.

[0085] Unter Bezugnahme auf **Fig. 7** wird ein Verfahren **700** gemäß einer Ausführungsform gezeigt. Das Verfahren **700** kann gemäß der vorliegenden Erfindung in jeder der Umgebungen durchgeführt werden, die u.a. in den **Fig. 1** bis **Fig. 6** in verschiedenen Ausführungsformen dargestellt werden. Selbstverständlich können in dem Verfahren **700** mehr oder weniger Operationen als in **Fig. 7** konkret beschrieben enthalten sein, wie dem Fachmann beim Lesen der vorliegenden Beschreibungen klar sein dürfte.

[0086] Jeder der Schritte des Verfahrens **700** kann durch jede geeignete Komponente der Betriebsumgebung durchgeführt werden. Zum Beispiel kann bei verschiedenen Ausführungsformen das Verfahren **700** teilweise oder vollständig durch einen Mikroprozessor, einen Server, einen Cluster von Datenverarbeitungseinheiten (z.B. einen zentralen Cluster), eine Verarbeitungsschaltung mit einem oder mehreren darin befindlichen Prozessoren oder eine anderweitige Einheit durchgeführt werden, die einen oder mehrere Prozessoren aufweist. Die Verarbeitungsschaltung, z.B. Prozessor(en), Chip(s) und/oder Modul(e), die in Hardware und/oder in Software realisiert werden und vorzugsweise mindestens eine Hardware-Komponente enthalten, kann in jeder Einheit genutzt werden, um einen oder mehrere Schritte des Verfahrens **700** durchzuführen. Veranschaulichende Prozessoren beinhalten, ohne jedoch darauf beschränkt zu sein, eine MPU, eine CPU, eine ASIC, ein FPGA usw., Kombinationen hiervon oder jede andere geeignete Datenverarbeitungseinheit nach dem Stand der Technik.

[0087] Wie in **Fig. 7** gezeigt, kann das Verfahren **700** mit einer Operation **702** beginnen, in der eine Mehrzahl von Dateien gefiltert werden, die in einem zentralen Cluster eines verteilten Dateisystems gespeichert sind, um vor einem Empfangen einer Abfrage zu der Mehrzahl von Dateien unabhängige Teile der Mehrzahl von Dateien unter Verwendung von Filtern, die für ein solches Filtern ausgestaltet sind, in einer Mehrzahl von Gruppen zu platzieren. Bei diesem Filterprozess werden Dateien innerhalb einer jeden aus der Mehrzahl von Gruppen gruppiert, da sie ein gemeinsames, suchbares Merkmal teilen, z.B. ein Datum, einen oder mehrere Benutzer, einen geografischen Standort, einen Inhalt (z.B. eine Art von Information, die in der Datei enthalten ist, wie beispielsweise eine Person, ein Ort, ein Konzept, eine Idee, ein Objekt usw., auf die/den/das sich die Datei bezieht, zu der/dem sie Informationen enthält, die/das/den sie beschreibt usw.), einen Typ (Text, Audio, Video usw.), eine Größe, einen Speicherort usw.

[0088] In einer Operation **704** wird eine Angabe zu der Abfrage in dem zentralen Cluster empfangen. Diese Angabe kann die Abfrage selbst, eine Datei oder ein Verweis sein, die bzw. der die Abfrage und/oder den beabsichtigten Zweck der Abfrage beschreibt (z.B. eine Art der Abfrage, Dateien, auf die abgezielt wird, und/oder Dateitypen für die Abfrage usw.).

[0089] In einer Operation **706** wird auf die Abfrage geantwortet, indem Dateien aus einer oder mehreren der Mehrzahl von Gruppen, die der Abfrage entsprechen, in einen lokalen Cluster des verteilten Dateisystems dupliziert werden, der die Angabe zu der Abfrage bereitgestellt hat. Der lokale Cluster unterscheidet sich in geografischer Hinsicht von dem zentralen Cluster, z.B. befinden sie sich an verschiedenen geografischen Standorten.

[0090] Welche Dateien sich auf die Abfrage beziehen, kann auf Grundlage einer Schlüsselwortgruppierung der Dateien in dem zentralen Cluster ermittelt werden. Wenn zum Beispiel bei einer Abfrage nach allen Grundbucheinträgen eines bestimmten Postleitzahlbereichs gesucht wird, kann der zentrale Cluster die Dateien aus einer Gruppe zurückgeben, die alle Dateien mit Grundbucheinträgen in diesem Bereich enthält, die durch eine Untergruppe eingegrenzt wurden, welche nur die Grundbucheinträge innerhalb des angeforderten Postleitzahlbereichs enthält.

[0091] Bei einer weiteren Ausführungsform kann das Verfahren **700** ein Empfangen der Mehrzahl von Dateien aus einer oder mehreren Quellen in dem zentralen Cluster beinhalten. Die Quellen können jede geeignete Quelle von Daten und Dateien beinhalten, z.B. lokale Cluster des verteilten Dateisystems, Hosts, Server, das Internet, ein SAN, das dem verteilten Dateisystem Daten bereitstellt, usw. Bei manchen Ansätzen beinhaltet die Mehrzahl von Dateien Text- und unstrukturierte Daten. Die unstrukturierten Daten können Bilddaten, Audiodaten, Datendateien in einem proprietären Format usw. beinhalten. Darüber hinaus kann das Verfahren **700** ein Speichern der Mehrzahl von Dateien in dem zentralen Cluster und ein Konvertieren der unstrukturierten Daten in Text in dem zentralen Cluster beinhalten. Auf diese Weise können die Filter auf den Text der Dateien angewendet werden, nachdem dieser aus den unstrukturierten Daten konvertiert wurde, so dass selbst Bilder, Videos, Audiodateien usw. in logische Untercontainer in dem zentralen Cluster gruppiert werden können, um einem lokalen Cluster als Reaktion auf eine Abfrage hiervon bereitgestellt zu werden.

[0092] Gemäß einer weiteren Ausführungsform kann das Verfahren **700** ein Erzeugen der Filter auf Grundlage von einem oder mehreren Faktoren beinhalten, um die Mehrzahl von Gruppen zu erzeugen, die Dateien beinhalten, welche in Abfragen, die in

dem zentralen Cluster empfangen werden, häufig angefordert werden. Die Faktoren können beinhalten, ohne jedoch darauf beschränkt zu sein: ein Datum, das einer oder mehreren Dateien zugehörig ist, einen oder mehrere Benutzer, die einer oder mehreren Dateien zugehörig sind, ausgewählten Inhalt und/oder ein oder mehrere Schlüsselwörter einer oder mehrerer Dateien, einen geografischen Standort, der einer oder mehreren Dateien zugehörig ist, usw. Andere Faktoren können ein oder mehrere Schlüsselwörter beinhalten, auf die die Dateien durchsucht werden können, um ähnlichen und/oder gemeinsamen Inhalt innerhalb von Dateien zu ermitteln, die in einer gemeinsamen Gruppe zusammengefasst sind. Des Weiteren können die Filter im Laufe der Zeit angepasst werden, um Dateien zu gruppieren, die in Abfragen, welche in dem zentralen Cluster empfangen werden, häufig angefordert werden. Mit steigender Zahl der empfangenen Abfragen können die Filter somit die optimale Gruppierung für die in dem zentralen Cluster gespeicherten Dateien lernen, um effizient auf die Abfragen zu reagieren. Wenn zum Beispiel eine Abfrage eines bestimmten Teilsatzes von Dateien mehrmals empfangen wird, kann eine Gruppe erzeugt werden, die den Teilsatz von Dateien enthält, so dass sie bei jeder darauffolgenden Abfrage des selben Teilsatzes von Dateien zurückgegeben werden. Diese Art des Lernens und Anpassens ist von Nutzen, um den Ressourcenverbrauch in dem zentralen Cluster zu verringern, der erforderlich wäre, um diese Ergebnisse bei jeder Anforderung der Abfrage zu filtern, anstatt lediglich einmal zum Erzeugen der Gruppe.

[0093] Die Filter können manuell erzeugt werden, z.B. durch einen Administrator des zentralen Clusters oder ein wie auch immer geartetes anderes System, das Zugriff auf die Dateien hat, um eine Wahrscheinlichkeit zu verbessern, dass Dateien innerhalb einer bestimmten Gruppe für eine oder mehrere darauffolgende Abfragen relevant sind. Bei einer alternativen Ausführungsform können die Filter kognitiv und intelligent durch den zentralen Cluster oder ein wie auch immer geartetes anderes System, das Zugriff auf die Filter hat, automatisch erzeugt werden, wodurch eine Eingabe von Benutzern des bzw. der Systeme ganz oder größtenteils vermieden wird, die darauf abzielt, die Filter zu erzeugen und die Wahrscheinlichkeit zu maximieren, dass Dateien innerhalb bestimmter Gruppen für darauffolgend empfangene Abfragen in dem zentralen Cluster relevant sind.

[0094] Bei einer weiteren Ausführungsform kann das Verfahren **700** ein Aktualisieren der Mehrzahl von Gruppen beinhalten, um Änderungen der Mehrzahl von Dateien zu berücksichtigen, die in dem globalen Cluster gespeichert sind. Wenn sich die Dateien im Laufe der Zeit ändern, ändern sich auch die Gruppen und die Zugehörigkeit von Dateien zu den Gruppen. Das Aktualisieren der Mehrzahl von Gruppen kann

somit in regelmäßigen Abständen auf Grundlage eines vorgegebenen Zeitplans, fortlaufend als eine Hintergrundoperation oder als Reaktion auf ein auslösendes Ereignis stattfinden. Eine Reihe von unterschiedlichen Aktionen können durchgeführt werden, um die Mehrzahl von Gruppen zu aktualisieren, z.B., ohne jedoch darauf beschränkt zu sein: ein Entfernen von einer oder mehreren Gruppen, ein Hinzufügen von einer oder mehreren Gruppen, ein Hinzufügen von einer oder mehreren Dateien zu einer bestimmten Gruppe und ein Entfernen von einer oder mehreren Dateien aus der bestimmten Gruppe. Jedes geeignete auslösende Ereignis kann verwendet werden, um die Gruppen und die Zugehörigkeit hierzu zu aktualisieren. Bei einer weiteren Ausführungsform kann das auslösende Ereignis beinhalten, ohne jedoch darauf beschränkt zu sein: eine Löschung einer vorhandenen Datei in dem zentralen Cluster, eine Hinzufügung einer neuen Datei zu dem zentralen Cluster, eine Hinzufügung eines neuen Dateityps zu dem zentralen Cluster und eine Aktualisierung eines Textkonvertierungsprozesses des zentralen Clusters.

[0095] Das Verfahren **700** kann in einem System, einem Verfahren und/oder einem Computerprogrammprodukt realisiert sein. Zum Beispiel kann ein System eine Verarbeitungsschaltung und eine mit der Verarbeitungsschaltung integrierte Logik beinhalten, die durch die Verarbeitungsschaltung ausführbar ist oder die mit der Verarbeitungsschaltung integriert und durch diese ausführbar ist. „Integriert mit“ bedeutet, dass die Verarbeitungseinheit ein Hardware-Prozessor ist, der fest codierte Logik beinhaltet, z.B. eine ASIC, ein FPGA usw. „Ausführbar durch“ bedeutet, dass der Prozessor zum Ausführen von Software-Logik konfiguriert ist, um eine durch die Software-Logik vorgegebene Funktionalität zu realisieren, wobei der Prozessor eine MPU, eine CPU, ein Mikroprozessor usw. sein kann. Die Logik ist konfiguriert, um die Verarbeitungsschaltung zum Durchführen des Verfahrens **700** zu veranlassen.

[0096] Bei einem weiteren Beispiel kann ein Computerprogrammprodukt ein durch einen Computer lesbare Speichermedium mit darin enthaltenen Programmanweisungen beinhalten. Das durch einen Computer lesbare Speichermedium kann jedes geeignete Speichermedium sein, dass nach dem Stand der Technik bekannt und konfiguriert ist, um Daten zu speichern und einem Computer einen Zugriff auf darin gespeicherte Daten zu ermöglichen. Die enthaltenen Programmanweisungen sind durch eine Verarbeitungsschaltung ausführbar, um die Verarbeitungsschaltung zum Durchführen des Verfahrens **700** zu veranlassen.

[0097] Unter Bezugnahme auf **Fig. 8** wird ein Verfahren **800** gemäß einer Ausführungsform gezeigt. Das Verfahren **800** kann gemäß der vorliegenden Erfin-

dung in jeder der Umgebungen durchgeführt werden, die u.a. in den **Fig. 1** bis **Fig. 6** in verschiedenen Ausführungsformen dargestellt werden. Selbstverständlich können in dem Verfahren **800** mehr oder weniger Operationen als in **Fig. 8** konkret beschrieben enthalten sein, wie dem Fachmann beim Lesen der vorliegenden Beschreibungen klar sein dürfte.

[0098] Jeder der Schritte des Verfahrens **800** kann durch jede geeignete Komponente der Betriebsumgebung durchgeführt werden. Zum Beispiel kann bei verschiedenen Ausführungsformen das Verfahren **800** teilweise oder vollständig durch einen Mikroprozessor, einen Server, einen Cluster von Datenverarbeitungseinheiten (z.B. einen lokalen Cluster), eine Verarbeitungsschaltung mit einem oder mehreren darin befindlichen Prozessoren oder eine anderweitige Einheit durchgeführt werden, die einen oder mehrere Prozessoren aufweist. Die Verarbeitungsschaltung, z.B. Prozessor(en), Chip(s) und/oder Modul(e), die in Hardware und/oder in Software realisiert werden und vorzugsweise mindestens eine Hardware-Komponente enthalten, kann in jeder Einheit genutzt werden, um einen oder mehrere Schritte des Verfahrens **800** durchzuführen. Veranschaulichende Prozessoren beinhalten, ohne jedoch darauf beschränkt zu sein, eine MPU, eine CPU, eine ASIC, ein FPGA usw., Kombinationen hiervon oder jede andere geeignete Datenverarbeitungseinheit nach dem Stand der Technik.

[0099] Wie in **Fig. 8** gezeigt, kann das Verfahren **800** mit einer Operation **802** beginnen, in der ein lokaler Cluster eines verteilten Dateisystems durchsucht oder anderweitig auf Dateien untersucht wird, die für eine Abfrage relevant sind. Diese Operation findet statt, bevor eine Angabe zu der Abfrage an einen zentralen Cluster des verteilten Dateisystems gesendet wird, um sicherzustellen, dass die Interaktion mit dem zentralen Cluster minimiert wird, falls die Dateien bereits in dem lokalen Cluster vorhanden sein sollten.

[0100] Nach einem Ermitteln, dass für die Abfrage relevante Dateien bereits in dem lokalen Cluster vorhanden sind, kann bei einer weiteren Ausführungsform darüber hinaus eine separate Abfrage an den zentralen Cluster gesendet werden, um zu ermitteln, ob eine beliebige der vorhandenen Dateien in dem zentralen Cluster aktualisiert wurde, seit sie in den lokalen Cluster dupliziert wurde, und wenn dies der Fall sein sollte, werden solche aktualisierten Dateien in den lokalen Cluster dupliziert, um veraltete Dateien zu ersetzen, die in dem lokalen Cluster gespeichert sind.

[0101] In einer Operation **804** wird die Angabe zu der Abfrage als Reaktion auf eine Ermittlung an den zentralen Cluster gesendet, dass die relevanten Dateien nicht in dem lokalen Cluster gespeichert sind. Bei einer Ausführungsform kann die Abfrage selbst als die

Angabe gesendet werden. Bei anderen Ausführungsformen beinhaltet die Angabe sachdienliche Informationen, die der zentrale Cluster benötigt, um Dateien zu ermitteln, die für die Abfrage relevant wären, z.B. einen oder mehrere Parameterwerte, eine oder mehrere Datumsangaben, eine oder mehrere Uhrzeiten, einen oder mehrere Standorte, einen oder mehrere Benutzer usw.

[0102] In einer Operation **806** wird eine Gruppe von Dateien, die für die Abfrage relevant sind, in dem lokalen Cluster empfangen. Bei einer Ausführungsform wird diese Gruppe von Dateien als Reaktion darauf durch den zentralen Cluster gesendet, dass die Angabe zu der Abfrage an den zentralen Cluster gesendet und von diesem empfangen wird. Bei einer Ausführungsform kann diese Gruppe von Dateien in dem zentralen Cluster bereits vorhanden sein, bevor die Angabe zu der Abfrage an den zentralen Cluster gesendet wird, wodurch eine schnelle Antwort auf die Abfrage bereitgestellt wird und der lokale Cluster in die Lage versetzt wird, die Abfrage zu der empfangenen Gruppe von Dateien durchzuführen, anstatt eine Durchführung einer solcher Aktion durch den zentralen Cluster zu erfordern.

[0103] In einer Operation **808** wird die Abfrage zu der Gruppe von Dateien unter der Bedingung ausgeführt, dass die Gruppe von Dateien nicht alle Dateien enthält, die in dem zentralen Cluster gespeichert sind. Jede geeignete Abfrage kann ausgeführt werden, z.B. eine Analyseabfrage, die eine komplexe Funktionalität und/oder mehrere Analyseprozesse innerhalb einer einzigen Abfrage enthalten kann, eine Suchabfrage, eine Anfügeabfrage, eine Löschabfrage, eine Tabellenerstellungsabfrage usw.

[0104] In einer Operation **810** wird die Gruppe von Dateien für eine vorbestimmte Zeitspanne in dem lokalen Cluster gespeichert, die im Einklang mit einer Richtlinie für die Cache-Leerung mit einem letzten Zugriff auf die Gruppe von Dateien beginnt. Sollte durch den lokalen Cluster eine darauffolgende Abfrage ausgegeben werden, die denselben oder einen überlappenden Teil der Gruppe von Dateien verwendet, muss auf diese Weise die Gruppe von Dateien nicht aus dem zentralen Cluster dupliziert werden, wodurch Ressourcen des zentralen Clusters und des Netzwerks im Allgemeinen geschont werden.

[0105] Die Richtlinie für die Cache-Leerung kann durch einen Administrator erzeugt und/oder angepasst werden, um die in dem lokalen Cluster gespeicherten Dateien effizient zu verwalten, um den begrenzten Speicherplatz des lokalen Clusters besser zu nutzen und dessen Ressourcenauslastung zu maximieren sowie die Ressourcenauslastung des zentralen Clusters zu minimieren.

[0106] Das Verfahren **800** kann in einem System, einem Verfahren und/oder einem Computerprogrammprodukt realisiert sein. Zum Beispiel kann ein System eine Verarbeitungsschaltung und eine mit der Verarbeitungsschaltung integrierte Logik beinhalten, die durch die Verarbeitungsschaltung ausführbar ist oder die mit der Verarbeitungsschaltung integriert und durch diese ausführbar ist. „Integriert mit“ bedeutet, dass die Verarbeitungseinheit ein Hardware-Prozessor ist, der fest codierte Logik beinhaltet, z.B. eine ASIC, ein FPGA usw. „Ausführbar durch“ bedeutet, dass der Prozessor zum Ausführen von Software-Logik konfiguriert ist, um eine durch die Software-Logik vorgegebene Funktionalität zu realisieren, wobei der Prozessor eine MPU, eine CPU, ein Mikroprozessor usw. sein kann. Die Logik ist konfiguriert, um die Verarbeitungsschaltung zum Durchführen des Verfahrens **800** zu veranlassen.

[0107] Bei einem weiteren Beispiel kann ein Computerprogrammprodukt ein durch einen Computer lesbare Speichermedium mit darin enthaltenen Programmanweisungen beinhalten. Das durch einen Computer lesbare Speichermedium kann jedes geeignete Speichermedium sein, dass nach dem Stand der Technik bekannt und konfiguriert ist, um Daten zu speichern und einem Computer einen Zugriff auf darin gespeicherte Daten zu ermöglichen. Die enthaltenen Programmanweisungen sind durch eine Verarbeitungsschaltung ausführbar, um die Verarbeitungsschaltung zum Durchführen des Verfahrens **800** zu veranlassen.

[0108] Bei der vorliegenden Erfindung kann es sich um ein System, ein Verfahren und/oder ein Computerprogrammprodukt handeln. Das Computerprogrammprodukt kann (ein) durch einen Computer lesbare(s) Speichermedium (oder -medien) beinhalten, auf dem/denen durch einen Computer lesbare Programmanweisungen gespeichert sind, um einen Prozessor dazu zu veranlassen, Aspekte der vorliegenden Erfindung auszuführen.

[0109] Bei dem durch einen Computer lesbaren Speichermedium kann es sich um eine physische Einheit handeln, die Anweisungen zur Verwendung durch eine Anweisungsausführungseinheit behalten und speichern kann. Bei dem durch einen Computer lesbaren Speichermedium kann es sich zum Beispiel um eine elektronische Speichereinheit, eine magnetische Speichereinheit, eine optische Speichereinheit, eine elektromagnetische Speichereinheit, eine Halbleiterspeichereinheit oder jede geeignete Kombination daraus handeln, ohne auf diese beschränkt zu sein. Zu einer nicht erschöpfenden Liste spezifischer Beispiele des durch einen Computer lesbaren Speichermediums gehören die Folgenden: eine tragbare Computerdiskette, eine Festplatte, ein Direktzugriffsspeicher (RAM), ein Nur-Lese-Speicher (ROM), ein löscherbarer programmierbarer Nur-Lese-

Speicher (EPROM bzw. Flash-Speicher), ein statischer Direktzugriffsspeicher (SRAM), ein tragbarer Compact-Disc-Nur-Lese-Speicher (CD-ROM), eine DVD (Digital Versatile Disc), ein Speicher-Stick, eine Diskette, eine mechanisch kodierte Einheit wie zum Beispiel Lochkarten oder erhabene Strukturen in einer Rille, auf denen Anweisungen gespeichert sind, und jede geeignete Kombination daraus. Ein durch einen Computer lesbare Speichermedium soll in der Verwendung hierin nicht als flüchtige Signale an sich aufgefasst werden, wie zum Beispiel Funkwellen oder andere sich frei ausbreitende elektromagnetische Wellen, elektromagnetische Wellen, die sich durch einen Wellenleiter oder ein anderes Übertragungsmedium ausbreiten (z.B. durch ein Lichtwellenleiterkabel geleitete Lichtimpulse) oder durch einen Draht übertragene elektrische Signale.

[0110] Hierin beschriebene, durch einen Computer lesbare Programmanweisungen können von einem durch einen Computer lesbaren Speichermedium auf jeweilige Datenverarbeitungs-/Verarbeitungseinheiten oder über ein Netzwerk wie zum Beispiel das Internet, ein lokales Netzwerk, ein Weitverkehrsnetzwerk und/oder ein drahtloses Netzwerk auf einen externen Computer oder eine externe Speichereinheit heruntergeladen werden. Das Netzwerk kann Kupferübertragungskabel, Lichtwellenübertragungsleiter, drahtlose Übertragung, Leitwegrechner, Firewalls, Vermittlungseinheiten, Gateway-Computer und/oder Edge-Server aufweisen. Eine Netzwerkadapterkarte oder Netzwerkschnittstelle in jeder Datenverarbeitungs-/Verarbeitungseinheit empfängt durch einen Computer lesbare Programmanweisungen aus dem Netzwerk und leitet die durch einen Computer lesbaren Programmanweisungen zur Speicherung in einem durch einen Computer lesbaren Speichermedium innerhalb der entsprechenden Datenverarbeitungs-/Verarbeitungseinheit weiter.

[0111] Bei computerlesbaren Programmbefehlen zum Durchführen von Operationen der vorliegenden Erfindung kann es sich um Assembler-Befehle, ISA-Befehle (Instruction-Set-Architecture), Maschinenbefehle, maschinenabhängige Befehle, Mikrocode, Firmware-Befehle, einen Zustand festlegende Daten oder aber entweder um Quellcode oder um Objektcode handeln, der in einer beliebigen Kombination von einer oder mehreren Programmiersprachen wie z.B. einer objektorientierten Programmiersprache wie Smalltalk, C++ oder dergleichen, sowie in herkömmlichen prozeduralen Programmiersprachen wie z.B. der Programmiersprache „C“ oder ähnlichen Programmiersprachen geschrieben ist. Die durch einen Computer lesbaren Programmanweisungen können vollständig auf dem Computer des Benutzers, teilweise auf dem Computer des Benutzers, als eigenständiges Software-Paket, teilweise auf dem Computer des Benutzers und teilweise auf einem entfernt angeordneten Computer oder voll-

ständig auf dem entfernt angeordneten Computer oder Server ausgeführt werden. In letzterem Fall kann der entfernt angeordnete Computer mit dem Computer des Benutzers durch eine beliebige Art Netzwerk verbunden sein, darunter ein lokales Netzwerk (LAN) oder ein Weitverkehrsnetzwerk (WAN), oder die Verbindung kann mit einem externen Computer hergestellt werden (zum Beispiel über das Internet unter Verwendung eines Internet-Dienstanbieters). In einigen Ausführungsformen können elektronische Schaltungen, darunter zum Beispiel programmierbare Logikschaltungen, im Feld programmierbare Gatter-Anordnungen (FPGA, Field Programmable Gate Arrays) oder programmierbare Logikandornungen (PLA, Programmable Logic Arrays) die durch einen Computer lesbaren Programmanweisungen ausführen, indem sie Zustandsinformationen der durch einen Computer lesbaren Programmanweisungen nutzen, um die elektronischen Schaltungen zu personalisieren, um Aspekte der vorliegenden Erfindung durchzuführen.

[0112] Aspekte der vorliegenden Erfindung sind hierin unter Bezugnahme auf Ablaufpläne und/oder Blockschaltbilder bzw. Schaubilder von Verfahren, Vorrichtungen (Systemen) und Computerprogrammprodukten gemäß Ausführungsformen der Erfindung beschrieben. Es wird darauf hingewiesen, dass jeder Block der Ablaufpläne und/oder der Blockschaltbilder bzw. Schaubilder sowie Kombinationen von Blöcken in den Ablaufplänen und/oder den Blockschaltbildern bzw. Schaubildern mittels durch einen Computer lesbare Programmanweisungen ausgeführt werden können.

[0113] Diese durch einen Computer lesbaren Programmanweisungen können einem Prozessor eines Universalcomputers, eines Spezialcomputers oder einer anderen programmierbaren Datenverarbeitungsvorrichtung bereitgestellt werden, um eine Maschine zu erzeugen, so dass die über den Prozessor des Computers bzw. der anderen programmierbaren Datenverarbeitungsvorrichtung ausgeführten Anweisungen ein Mittel zur Umsetzung der in dem Block bzw. den Blöcken der Ablaufpläne und/oder der Blockschaltbilder bzw. Schaubilder festgelegten Funktionen/Schritte erzeugen. Diese durch einen Computer lesbaren Programmanweisungen können auch auf einem durch einen Computer lesbaren Speichermedium gespeichert sein, das einen Computer, eine programmierbare Datenverarbeitungsvorrichtung und/oder andere Einheiten so steuern kann, dass sie auf eine bestimmte Art funktionieren, so dass das durch einen Computer lesbare Speichermedium, auf dem Anweisungen gespeichert sind, ein Herstellungsprodukt aufweist, darunter Anweisungen, welche Aspekte der/des in dem Block bzw. den Blöcken des Ablaufplans und/oder der Blockschaltbilder bzw. Schaubilder angegebenen Funktion/Schritts umsetzen.

[0114] Die durch einen Computer lesbaren Programmieranweisungen können auch auf einen Computer, eine andere programmierbare Datenverarbeitungsvorrichtung oder eine andere Einheit geladen werden, um das Ausführen einer Reihe von Prozessschritten auf dem Computer bzw. der anderen programmierbaren Vorrichtung oder anderen Einheit zu verursachen, um einen durch einen Computer umgesetzten Prozess zu erzeugen, so dass die auf dem Computer, einer anderen programmierbaren Vorrichtung oder einer anderen Einheit ausgeführten Anweisungen die in dem Block bzw. den Blöcken der Ablaufpläne und/oder der Blockschaltbilder bzw. Schaubilder festgelegten Funktionen/Schritte umsetzen.

[0115] Die Ablaufpläne und die Blockschaltbilder bzw. Schaubilder in den Figuren veranschaulichen die Architektur, die Funktionalität und den Betrieb möglicher Ausführungen von Systemen, Verfahren und Computerprogrammprodukten gemäß verschiedenen Ausführungsformen der vorliegenden Erfindung. In diesem Zusammenhang kann jeder Block in den Ablaufplänen oder Blockschaltbildern bzw. Schaubildern ein Modul, ein Segment oder einen Teil von Anweisungen darstellen, die eine oder mehrere ausführbare Anweisungen zur Ausführung der bestimmten logischen Funktion(en) aufweisen. In einigen alternativen Ausführungen können die in dem Block angegebenen Funktionen in einer anderen Reihenfolge als in den Figuren gezeigt stattfinden. Zwei nacheinander gezeigte Blöcke können zum Beispiel in Wirklichkeit im Wesentlichen gleichzeitig ausgeführt werden, oder die Blöcke können manchmal je nach entsprechender Funktionalität in umgekehrter Reihenfolge ausgeführt werden. Es ist ferner anzumerken, dass jeder Block der Blockschaltbilder bzw. Schaubilder und/oder der Ablaufpläne sowie Kombinationen aus Blöcken in den Blockschaltbildern bzw. Schaubildern und/oder den Ablaufplänen durch spezielle auf Hardware beruhende Systeme umgesetzt werden können, welche die festgelegten Funktionen oder Schritte durchführen, oder Kombinationen aus Spezial-Hardware und Computeranweisungen ausführen.

[0116] Darüber hinaus kann ein System gemäß verschiedenen Ausführungsformen einen Prozessor und Logik enthalten, die mit dem Prozessor integriert und/oder durch diesen ausführbar ist, wobei die Logik so konfiguriert werden kann, dass sie einen oder mehrere der hier genannten Prozessschritte ausführt. Dabei bedeutet „integriert mit“, dass der Prozessor über Logik verfügt, die in Form von Hardware-Logik in ihn eingebettet ist, wie z.B. eine ASIC, ein FPGA usw. „Durch den Prozessor ausführbar“ bedeutet, dass es sich bei der Logik um Hardware-Logik; Software-Logik wie z.B. Firmware, ein Teil eines Betriebssystems, ein Teil eines Anwendungsprogramms; usw. oder um eine Kombination von Hardware- und Software-Logik handelt, auf die der Prozessor zugreifen kann und die

so konfiguriert wird, dass sie den Prozessor veranlasst, bei Ausführung durch den Prozessor eine Funktionalität durchzuführen. Software-Logik kann in einem lokalen und/oder entfernt angeordneten Arbeitsspeicher eines beliebigen Arbeitsspeichertyps nach dem Stand der Technik gespeichert werden. Jeder Prozessor nach dem Stand der Technik kann verwendet werden, z.B. ein Software-Prozessormodul und/oder ein Hardware-Prozessormodul wie beispielsweise eine ASIC, ein FPGA, eine CPU, eine integrierte Schaltung (Integrated Circuit, IC), ein Grafikprozessor (Graphics Processing Unit, GPU) usw.

[0117] Es dürfte offensichtlich sein, dass die verschiedenen Merkmale der vorgenannten Systeme und/oder Methodiken auf eine beliebige Weise kombiniert werden können, so dass sich aus den oben dargelegten Beschreibungen eine Vielfalt von Kombinationen ergibt.

[0118] Obwohl oben verschiedene Ausführungsformen beschrieben wurden, sollte deutlich sein, dass diese lediglich als Beispiel und nicht als Einschränkung dargelegt wurden. Umfang und inhaltliche Reichweite einer bevorzugten Ausführungsform werden daher durch keine der oben beschriebenen beispielhaften Ausführungsformen beschränkt, sondern lediglich durch die folgenden Ansprüche und deren Entsprechungen definiert.

Patentansprüche

1. Verfahren, aufweisend:
 ein Filtern einer Mehrzahl von Dateien, die in einem zentralen Cluster eines verteilten Dateisystems gespeichert sind, um vor einem Empfangen einer Abfrage zu der Mehrzahl von Dateien unabhängige Teile der Mehrzahl von Dateien unter Verwendung von Filtern in einer Mehrzahl von Gruppen zu platzieren, wobei Dateien innerhalb einer jeden aus der Mehrzahl von Gruppen ein gemeinsames, suchbares Merkmal teilen;
 ein Empfangen einer Angabe zu der Abfrage in dem zentralen Cluster; und
 ein Antworten auf die Abfrage durch ein Duplizieren von Dateien aus einer oder mehreren der Mehrzahl von Gruppen, die der Abfrage entsprechen, in einen lokalen Cluster des verteilten Dateisystems, der die Angabe zu der Abfrage bereitgestellt hat und sich in geografischer Hinsicht von dem zentralen Cluster unterscheidet.

2. Verfahren nach Anspruch 1, aufweisend:
 ein Empfangen der Mehrzahl von Dateien aus einer oder mehreren Quellen in einem zentralen Cluster, wobei die Mehrzahl von Dateien Text- und unstrukturierte Daten beinhaltet;
 ein Speichern der Mehrzahl von Dateien in dem zentralen Cluster; und

ein Konvertieren der unstrukturierten Daten in Text in dem zentralen Cluster, wobei die Filter auf den Text der Dateien angewendet werden, nachdem dieser aus den unstrukturierten Daten konvertiert wurde.

3. Verfahren nach Anspruch 1, aufweisend: ein Erzeugen der Filter auf Grundlage von einem oder mehreren Faktoren, um die Mehrzahl von Gruppen zu erzeugen, die Dateien enthalten, welche in Abfragen, die in dem zentralen Cluster empfangen werden, häufig angefordert werden, wobei die Faktoren aus einem Satz von Faktoren ausgewählt werden, aufweisend: ein Datum, das einer oder mehreren Dateien zugehörig ist, einen oder mehrere Benutzer, die einer oder mehreren Dateien zugehörig sind, ausgewählten Inhalt und/oder ein oder mehrere Schlüsselwörter einer oder mehrerer Dateien sowie einen geografischen Standort, der einer oder mehreren Dateien zugehörig ist.

4. Verfahren nach Anspruch 3, aufweisend: ein Anpassen der Filter im Laufe der Zeit, um Dateien zu gruppieren, die in Abfragen, welche in dem zentralen Cluster empfangen werden, häufig angefordert werden.

5. Verfahren nach Anspruch 3, wobei die Filter erzeugt werden, um eine Wahrscheinlichkeit zu maximieren, dass Dateien innerhalb einer bestimmten Gruppe für eine oder mehrere empfangene Abfragen relevant sind.

6. Verfahren nach Anspruch 1, aufweisend: ein Aktualisieren der Mehrzahl von Gruppen, um Änderungen der Mehrzahl von Daten zu berücksichtigen, die in dem zentralen Cluster gespeichert sind, wobei das Aktualisieren der Mehrzahl von Gruppen in regelmäßigen Abständen auf Grundlage eines vorgegebenen Zeitplans, fortlaufend als eine Hintergrundoperation oder als Reaktion auf ein auslösendes Ereignis stattfindet und wobei das Aktualisieren der Mehrzahl von Gruppen eine Aktion beinhaltet, die aus einem Satz von Aktionen ausgewählt wird, aufweisend: ein Entfernen von einer oder mehreren Gruppen, ein Hinzufügen von einer oder mehreren Dateien zu einer bestimmten Gruppe und ein Entfernen von einer oder mehreren Dateien aus der bestimmten Gruppe.

7. Verfahren nach Anspruch 6, wobei das auslösende Ereignis aus einem Satz ausgewählt wird, aufweisend: eine Löschung einer vorhandenen Datei in dem zentralen Cluster, eine Hinzufügung einer neuen Datei zu dem zentralen Cluster, eine Hinzufügung eines neuen Dateityps zu dem zentralen Cluster und eine Aktualisierung eines Textkonvertierungsprozesses des zentralen Clusters.

8. Computerprogrammprodukt, wobei das Computerprogrammprodukt ein durch einen Computer lesbares Speichermedium mit darauf enthaltenen Programmanweisungen aufweist, wobei das durch einen Computer lesbare Speichermedium kein flüchtiges Signal an sich ist, wobei die Programmanweisungen durch eine Verarbeitungsschaltung ausführbar sind, um die Verarbeitungsschaltung zu veranlassen: durch die Verarbeitungsschaltung eine Mehrzahl von Dateien zu filtern, die in einem zentralen Cluster eines verteilten Dateisystems gespeichert sind, um vor einem Empfangen einer Abfrage zu der Mehrzahl von Dateien unabhängige Teile der Mehrzahl von Dateien unter Verwendung von Filtern in einer Mehrzahl von Gruppen zu platzieren, wobei Dateien innerhalb einer jeden aus der Mehrzahl von Gruppen ein gemeinsames, suchbares Merkmal teilen; durch die Verarbeitungsschaltung eine Angabe zu der Abfrage in dem zentralen Cluster zu empfangen; und durch die Verarbeitungsschaltung auf die Abfrage zu antworten, indem Dateien aus einer oder mehreren der Mehrzahl von Gruppen, die der Abfrage entsprechen, in einen lokalen Cluster des verteilten Dateisystems dupliziert werden, der die Angabe zu der Abfrage bereitgestellt hat.

9. Computerprogrammprodukt nach Anspruch 8, wobei die enthaltenen Programmanweisungen des Weiteren durch die Verarbeitungsschaltung ausführbar sind, um die Verarbeitungsschaltung zu veranlassen:

die Mehrzahl von Dateien aus einer oder mehreren Quellen durch die Verarbeitungsschaltung in einem zentralen Cluster eines verteilten Dateisystems zu empfangen, wobei die Mehrzahl von Dateien Text- und unstrukturierte Daten beinhaltet; die Mehrzahl von Dateien durch die Verarbeitungsschaltung in dem zentralen Cluster zu speichern; und die unstrukturierten Daten durch die Verarbeitungsschaltung in Daten auf Textgrundlage in dem zentralen Cluster zu konvertieren, wobei die Filter auf den Text der Dateien angewendet werden, nachdem dieser aus den unstrukturierten Daten konvertiert wurde.

10. Computerprogrammprodukt nach Anspruch 8, wobei die enthaltenen Programmanweisungen des Weiteren durch die Verarbeitungsschaltung ausführbar sind, um die Verarbeitungsschaltung zu veranlassen:

die Filter auf Grundlage von einem oder mehreren Faktoren durch die Verarbeitungsschaltung zu erzeugen, um die Mehrzahl von Gruppen zu erzeugen, die Dateien enthalten, welche in Abfragen, die in dem zentralen Cluster empfangen werden, häufig angefordert werden, wobei die Faktoren aus einem Satz von Faktoren ausgewählt werden, aufweisend: ein Datum, das einer oder mehreren Dateien zugehörig ist, einen oder mehrere Benutzer, die einer oder

mehreren Dateien zugehörig sind, ausgewählten Inhalt und/oder ein oder mehrere Schlüsselwörter einer oder mehrerer Dateien sowie einen geografischen Standort, der einer oder mehreren Dateien zugehörig ist.

11. Computerprogrammprodukt nach Anspruch 10, wobei die Filter erzeugt werden, um eine Wahrscheinlichkeit zu maximieren, dass Dateien innerhalb einer bestimmten Gruppe für eine oder mehrere empfangene Abfragen relevant sind.

12. Computerprogrammprodukt nach Anspruch 10, wobei die enthaltenen Programmanweisungen des Weiteren durch die Verarbeitungsschaltung ausführbar sind, um die Verarbeitungsschaltung zu veranlassen:

die Filter durch die Verarbeitungsschaltung im Laufe der Zeit anzupassen, um Dateien zu gruppieren, die in Abfragen, welche in dem zentralen Cluster empfangen werden, häufig angefordert werden.

13. Computerprogrammprodukt nach Anspruch 8, wobei die enthaltenen Programmanweisungen des Weiteren durch die Verarbeitungsschaltung ausführbar sind, um die Verarbeitungsschaltung zu veranlassen:

die Mehrzahl von Gruppen durch die Verarbeitungsschaltung zu aktualisieren, um Änderungen der Mehrzahl von Dateien zu berücksichtigen, die in dem zentralen Cluster gespeichert sind, wobei die enthaltenen Programmanweisungen, welche die Verarbeitungsschaltung veranlassen, die Mehrzahl von Gruppen zu aktualisieren, die Verarbeitungsschaltung des Weiteren veranlassen, eine Aktion aus einem Satz von Aktionen auszuwählen, aufweisend: ein Entfernen von einer oder mehreren Gruppen, ein Hinzufügen von einer oder mehreren Dateien zu einer bestimmten Gruppe und ein Entfernen von einer oder mehreren Dateien aus der bestimmten Gruppe, und wobei die enthaltenen Programmanweisungen, welche die Verarbeitungsschaltung veranlassen, die Mehrzahl von Gruppen zu aktualisieren, in regelmäßigen Abständen auf Grundlage eines vorgegebenen Zeitplans, fortlaufend als eine Hintergrundoperation oder als Reaktion auf ein auslösendes Ereignis ausgeführt werden.

14. Computerprogrammprodukt nach Anspruch 13, wobei das auslösende Ereignis aus einem Satz ausgewählt wird, aufweisend: eine Löschung einer vorhandenen Datei in dem zentralen Cluster, eine Hinzufügung einer neuen Datei zu dem zentralen Cluster, eine Hinzufügung eines neuen Dateityps zu dem zentralen Cluster und eine Aktualisierung eines Textkonvertierungsprozesses des zentralen Clusters.

15. System, aufweisend:
eine Verarbeitungsschaltung;

einen Arbeitsspeicher; und
in dem Arbeitsspeicher gespeicherte Logik, die bei Ausführung durch die Verarbeitungsschaltung die Verarbeitungsschaltung veranlasst:

eine Mehrzahl von Dateien zu filtern, die in einem zentralen Cluster eines verteilten Dateisystems gespeichert sind, um vor einem Empfangen einer Abfrage zu der Mehrzahl von Dateien unabhängige Teile der Mehrzahl von Dateien unter Verwendung von Filtern in einer Mehrzahl von Gruppen zu platzieren, wobei Dateien innerhalb einer jeden aus der Mehrzahl von Gruppen ein gemeinsames, suchbares Merkmal teilen;

eine Angabe zu der Abfrage in dem zentralen Cluster zu empfangen; und

auf die Abfrage zu antworten, indem Dateien aus einer oder mehreren der Mehrzahl von Gruppen, die der Abfrage entsprechen, in einen lokalen Cluster des verteilten Dateisystems dupliziert werden, der die Angabe zu der Abfrage bereitgestellt hat.

16. System nach Anspruch 15, wobei die Logik die Verarbeitungsschaltung des Weiteren veranlasst: die Mehrzahl von Dateien aus einer oder mehreren Quellen in dem zentralen Cluster zu empfangen, wobei die Mehrzahl von Dateien Text- und unstrukturierte Daten beinhaltet;

die Mehrzahl von Dateien in dem zentralen Cluster zu speichern; und

die unstrukturierten Daten in Text in dem zentralen Cluster zu konvertieren,

wobei die Filter auf den Text der Dateien angewendet werden, nachdem dieser aus den unstrukturierten Daten konvertiert wurde.

17. System nach Anspruch 15, wobei die Logik die Verarbeitungsschaltung des Weiteren veranlasst: die Filter auf Grundlage von einem oder mehreren Faktoren zu erzeugen, um die Mehrzahl von Gruppen zu erzeugen, die Dateien enthalten, welche in Abfragen, die in dem zentralen Cluster empfangen werden, häufig angefordert werden, wobei die Faktoren aus einem Satz von Faktoren ausgewählt werden, aufweisend: ein Datum, das einer oder mehreren Dateien zugehörig ist, einen oder mehrere Benutzer, die einer oder mehreren Dateien zugehörig sind, ausgewählten Inhalt und/oder ein oder mehrere Schlüsselwörter einer oder mehrerer Dateien sowie einen geografischen Standort, der einer oder mehreren Dateien zugehörig ist.

18. System nach Anspruch 17, wobei die Filter erzeugt werden, um eine Wahrscheinlichkeit zu maximieren, dass Dateien innerhalb einer bestimmten Gruppe für eine oder mehrere empfangene Abfragen relevant sind.

19. System nach Anspruch 17, wobei die Logik die Verarbeitungsschaltung des Weiteren veranlasst:

die Filter im Laufe der Zeit anzupassen, um Dateien zu gruppieren, die in Abfragen, welche in dem zentralen Cluster empfangen werden, häufig angefordert werden.

20. System nach Anspruch 15, wobei die Logik die Verarbeitungsschaltung des Weiteren veranlasst: die Mehrzahl von Gruppen zu aktualisieren, um Änderungen der Mehrzahl von Daten zu berücksichtigen, die in regelmäßigen Abständen auf Grundlage eines vorgegebenen Zeitplans, fortlaufend als eine Hintergrundoperation oder als Reaktion auf ein auslösendes Ereignis in dem zentralen Cluster gespeichert werden, und wobei die Logik, welche die Verarbeitungsschaltung veranlasst, die Mehrzahl von Gruppen zu aktualisieren, eine Aktion durchführt, die aus einem Satz von Aktionen ausgewählt wird, aufweisend: ein Entfernen von einer oder mehreren Gruppen, ein Hinzufügen von einer oder mehreren Dateien zu einer bestimmten Gruppe und ein Entfernen von einer oder mehreren Dateien aus der bestimmten Gruppe.

21. System nach Anspruch 20, wobei das auslösende Ereignis aus einem Satz ausgewählt wird, aufweisend: eine Löschung einer vorhandenen Datei in dem zentralen Cluster, eine Hinzufügung einer neuen Datei zu dem zentralen Cluster, eine Hinzufügung eines neuen Dateityps zu dem zentralen Cluster und eine Aktualisierung eines Textkonvertierungsprozesses des zentralen Clusters.

22. Verfahren, aufweisend:
 ein Empfangen einer Mehrzahl von Dateien aus einer oder mehreren Quellen in einem zentralen Cluster eines verteilten Dateisystems, wobei die Mehrzahl von Dateien Text- und unstrukturierte Daten beinhaltet;
 ein Speichern der Mehrzahl von Dateien in dem zentralen Cluster;
 ein Konvertieren der unstrukturierten Daten in Text in dem zentralen Cluster; und
 ein Filtern der Mehrzahl von Dateien, um unter Verwendung von Filtern vor einem Empfangen einer Abfrage zu der Mehrzahl von Dateien unabhängige Teile der Mehrzahl von Dateien in einer Mehrzahl von Gruppen zu platzieren, wobei Dateien innerhalb einer jeden aus der Mehrzahl von Gruppen ein gemeinsames, suchbares Merkmal teilen und wobei die Filter auf den Text der Dateien angewendet werden, nachdem dieser aus den unstrukturierten Daten konvertiert wurde.

23. Verfahren nach Anspruch 22, aufweisend:
 ein Empfangen einer Angabe zu der Abfrage in dem zentralen Cluster;
 ein Antworten auf die Abfrage, indem Dateien aus einer oder mehreren der Mehrzahl von Gruppen, die der Abfrage entsprechen, in einen lokalen Cluster des verteilten Dateisystems dupliziert werden, der die Angabe zu der Abfrage bereitgestellt hat; und

ein Anpassen der Filter im Laufe der Zeit, um Dateien zu gruppieren, die in Abfragen, welche in dem zentralen Cluster empfangen werden, häufig angefordert werden.

24. Verfahren nach Anspruch 22, aufweisend:
 ein Erzeugen der Filter auf Grundlage von einem oder mehreren Faktoren, um die Mehrzahl von Gruppen zu erzeugen, die Dateien enthalten, welche in Abfragen, die in dem zentralen Cluster empfangen werden, häufig angefordert werden, wobei die Faktoren aus einem Satz von Faktoren ausgewählt werden, aufweisend: ein Datum, das einer oder mehreren Dateien zugehörig ist, einen oder mehrere Benutzer, die einer oder mehreren Dateien zugehörig sind, ausgewählten Inhalt und/oder ein oder mehrere Schlüsselwörter einer oder mehrerer Dateien sowie einen geografischen Standort, der einer oder mehreren Dateien zugehörig ist; und
 ein Anpassen der Filter im Laufe der Zeit, um Dateien zu gruppieren, die in Abfragen, welche in dem zentralen Cluster empfangen werden, häufig angefordert werden.

25. Verfahren, aufweisend:
 ein Durchsuchen eines lokalen Clusters eines verteilten Dateisystems nach Dateien, die für eine Abfrage relevant sind, bevor eine Angabe zu der Abfrage an einen zentralen Cluster des verteilten Dateisystems gesendet wird;
 ein Senden der Angabe zu der Abfrage an den zentralen Cluster als Reaktion auf eine Ermittlung, dass die relevanten Dateien nicht in dem lokalen Cluster gespeichert sind;
 ein Empfangen einer Gruppe von Dateien, die für die Abfrage relevant sind, in dem lokalen Cluster;
 ein Ausführen der Abfrage zu der Gruppe von Dateien unter der Bedingung, dass die Gruppe von Dateien nicht alle Dateien enthält, die in dem zentralen Cluster gespeichert sind; und
 ein Speichern der Gruppe von Dateien für eine vorbestimmte Zeitspanne in dem lokalen Cluster, die im Einklang mit einer Richtlinie für die Cache-Leerung mit einem letzten Zugriff auf die Gruppe von Dateien beginnt.

Es folgen 10 Seiten Zeichnungen

Anhängende Zeichnungen

10 ↘

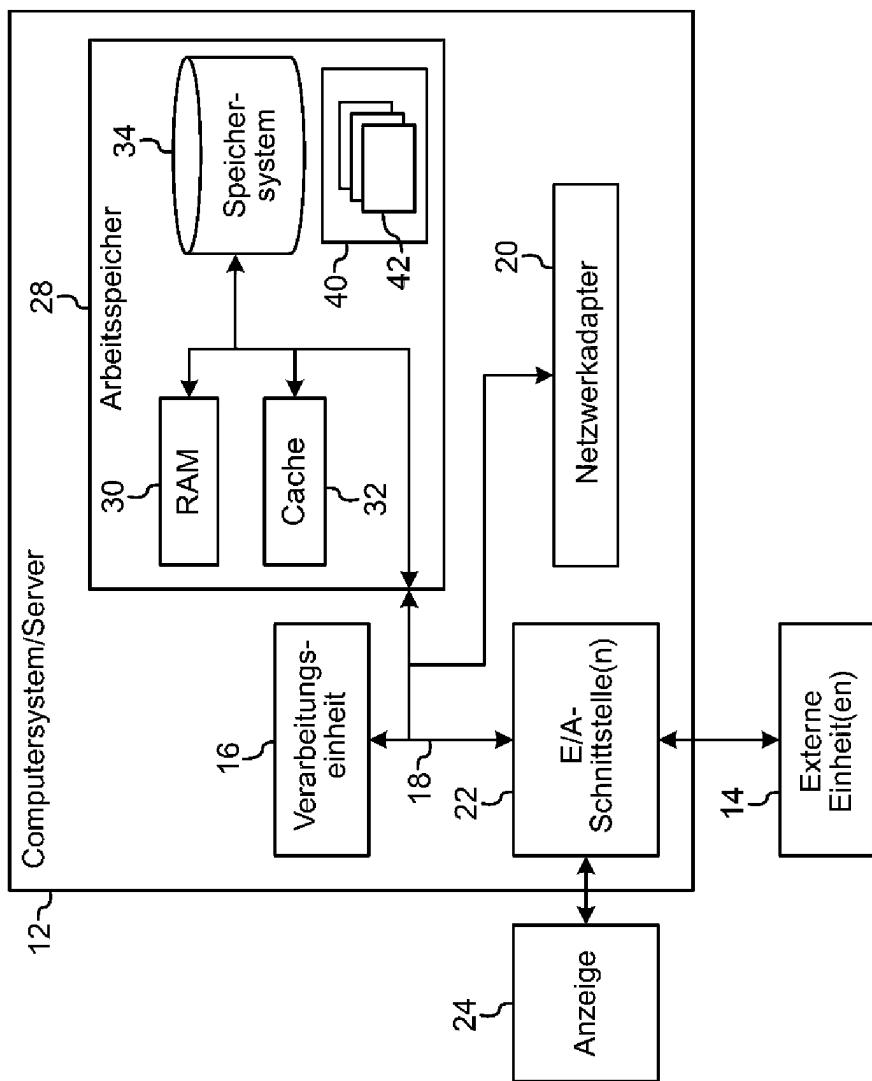


FIG. 1

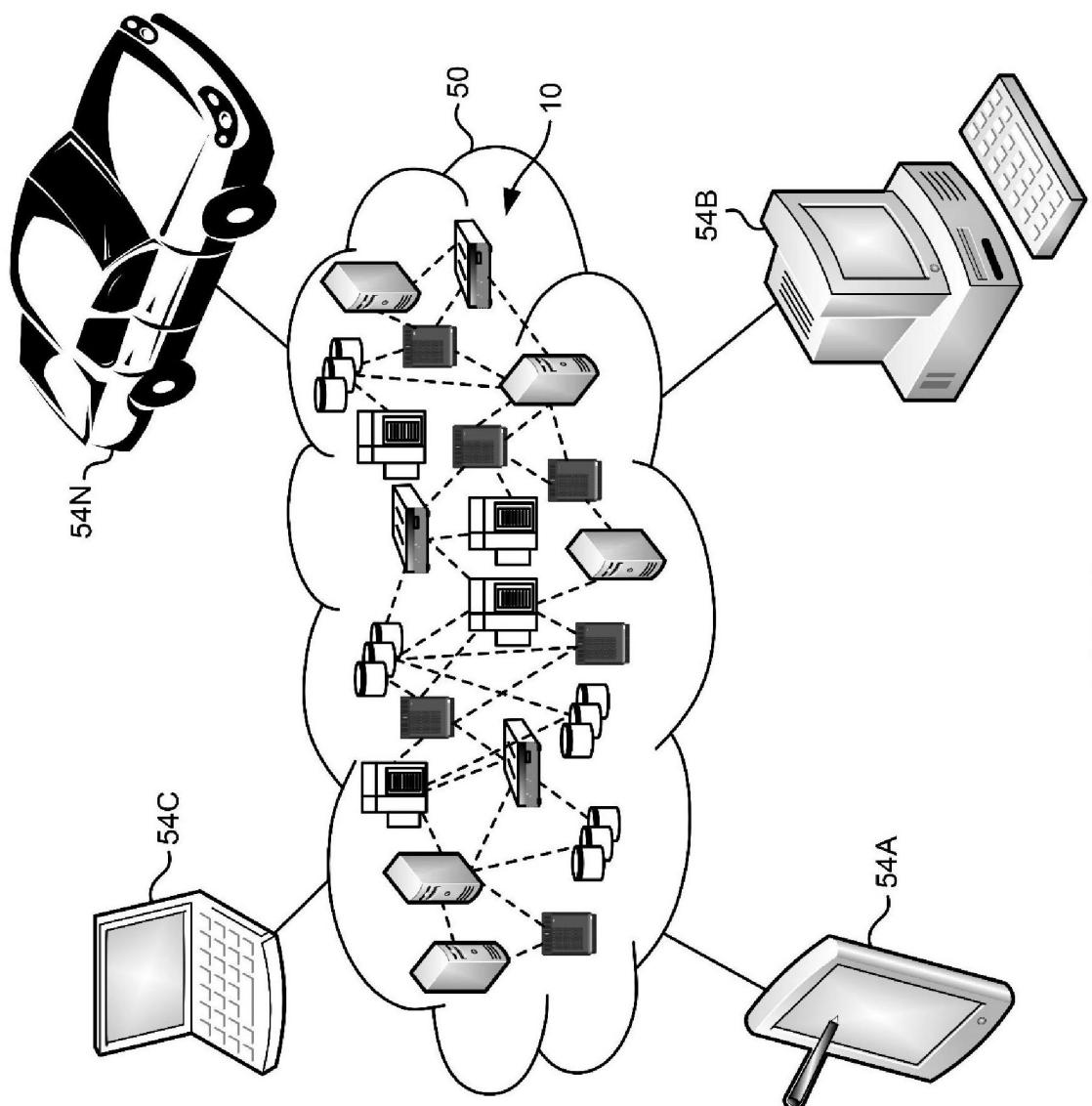


FIG. 2

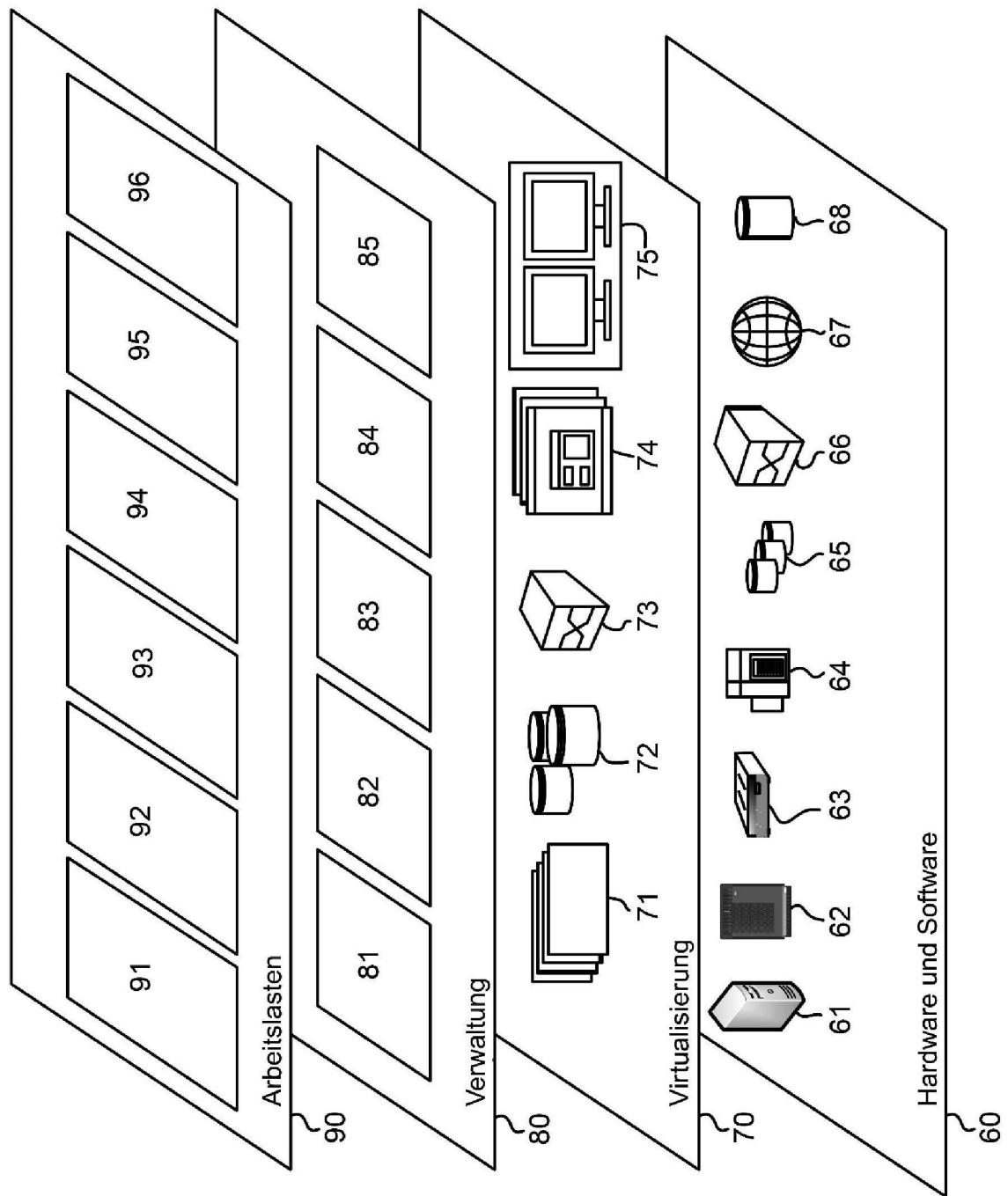


FIG. 3

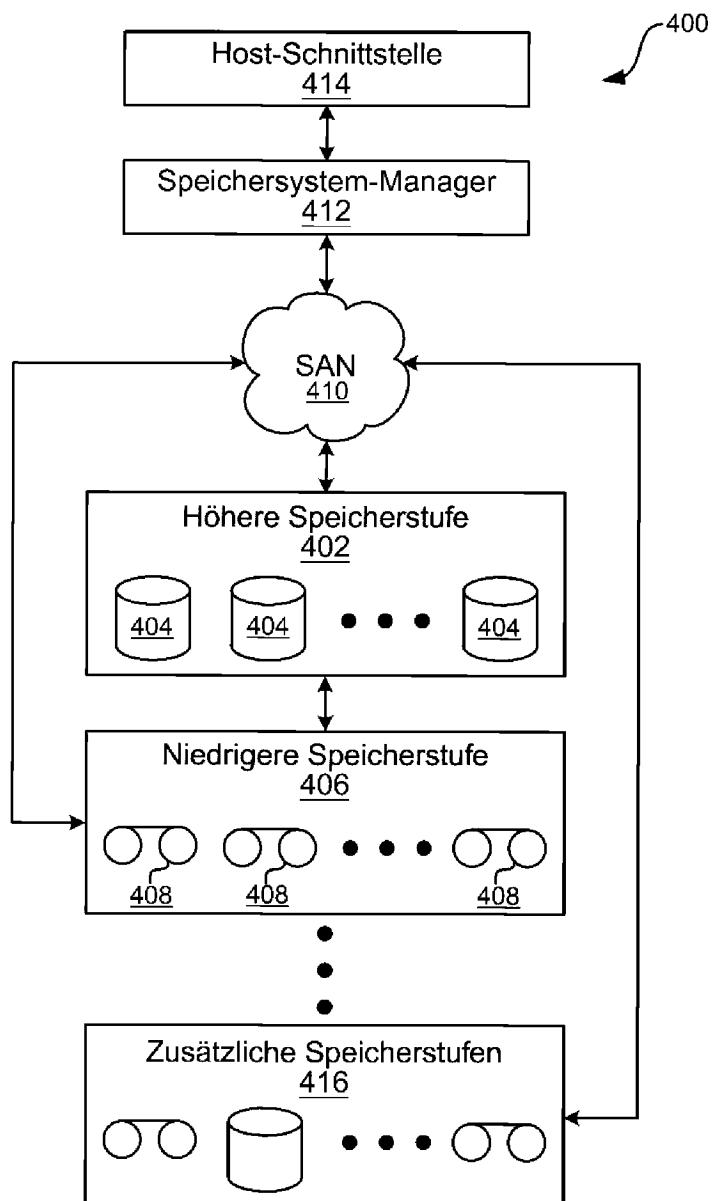


FIG. 4

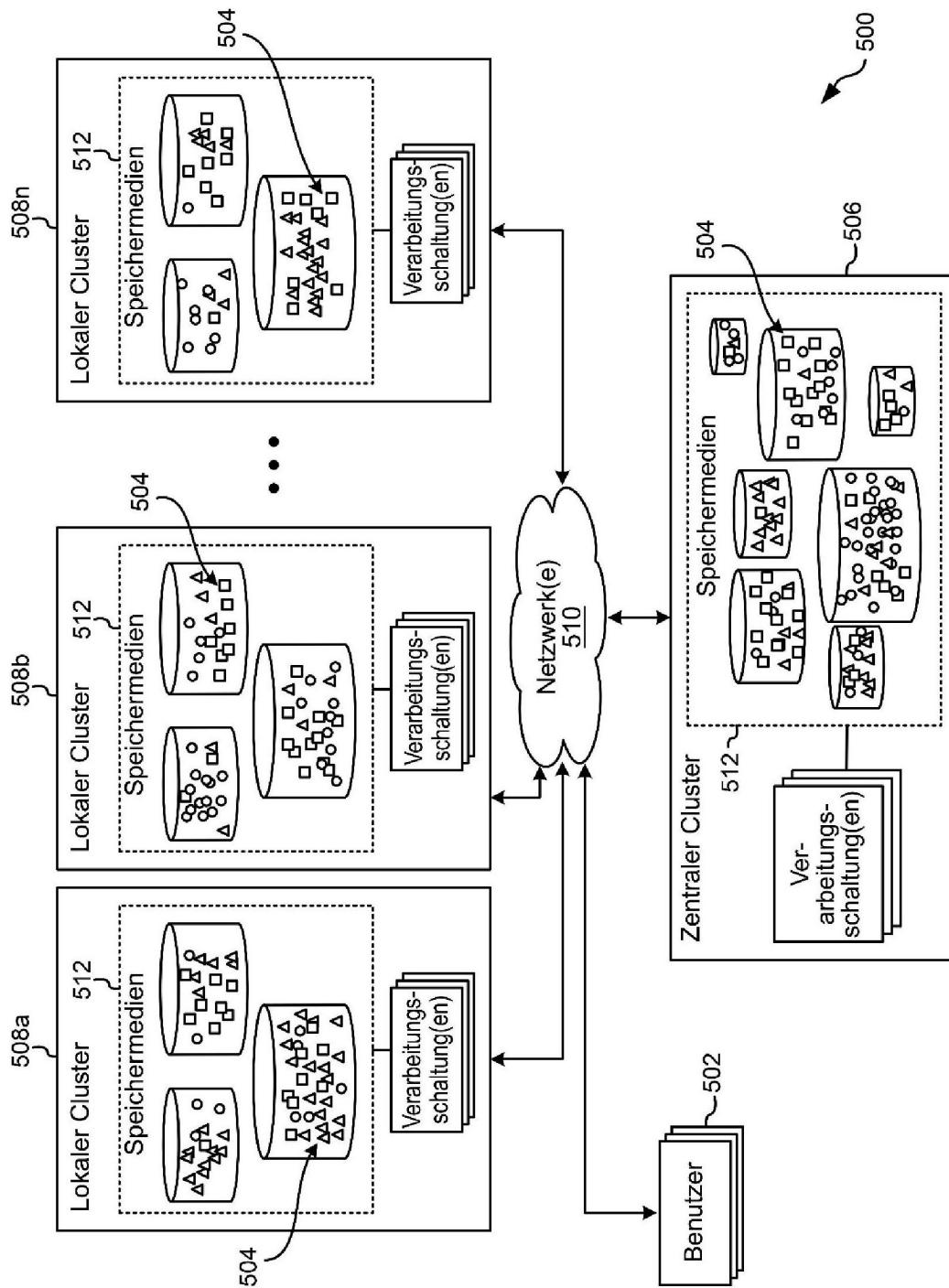


FIG. 5A

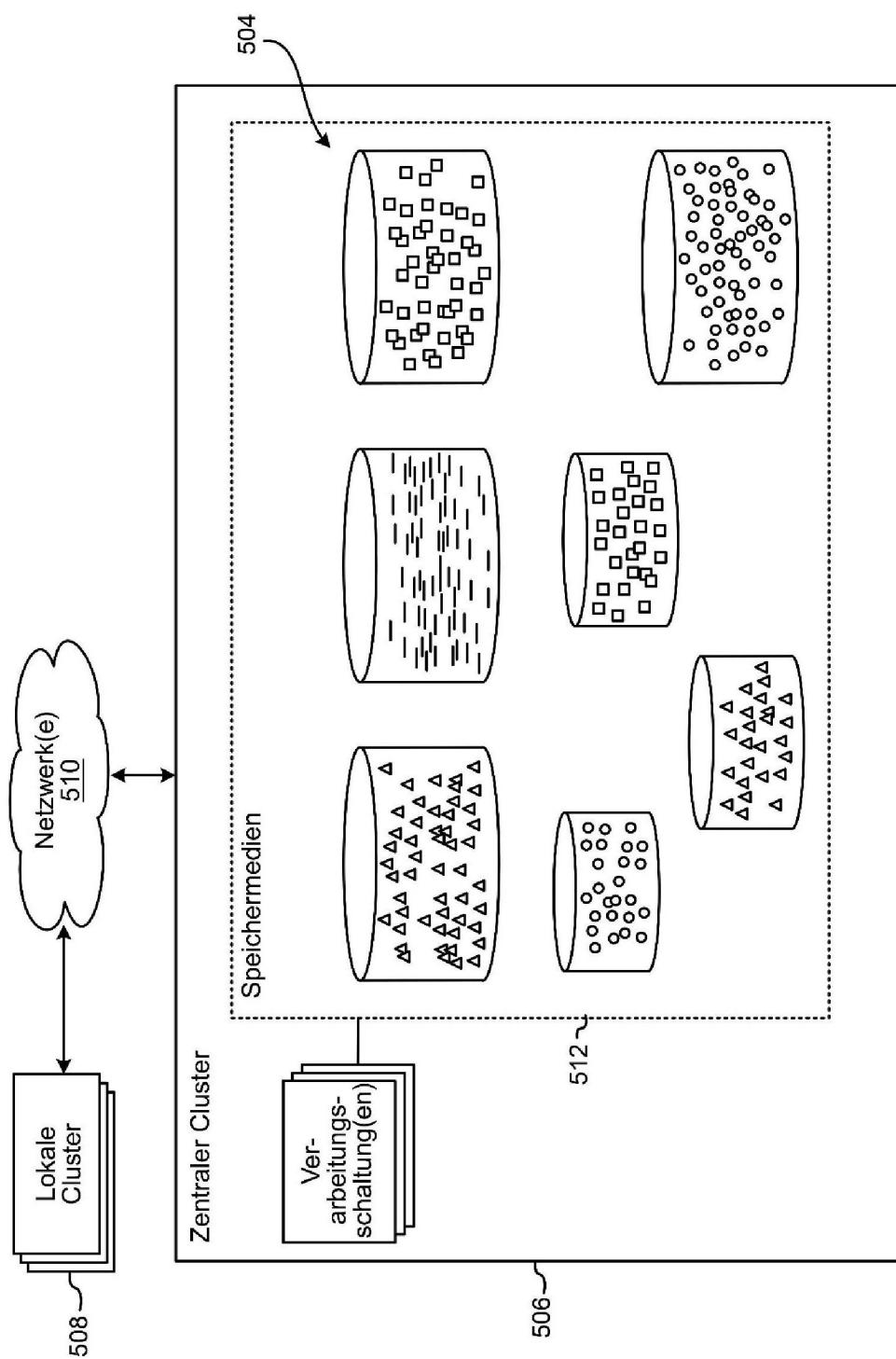


FIG. 5B

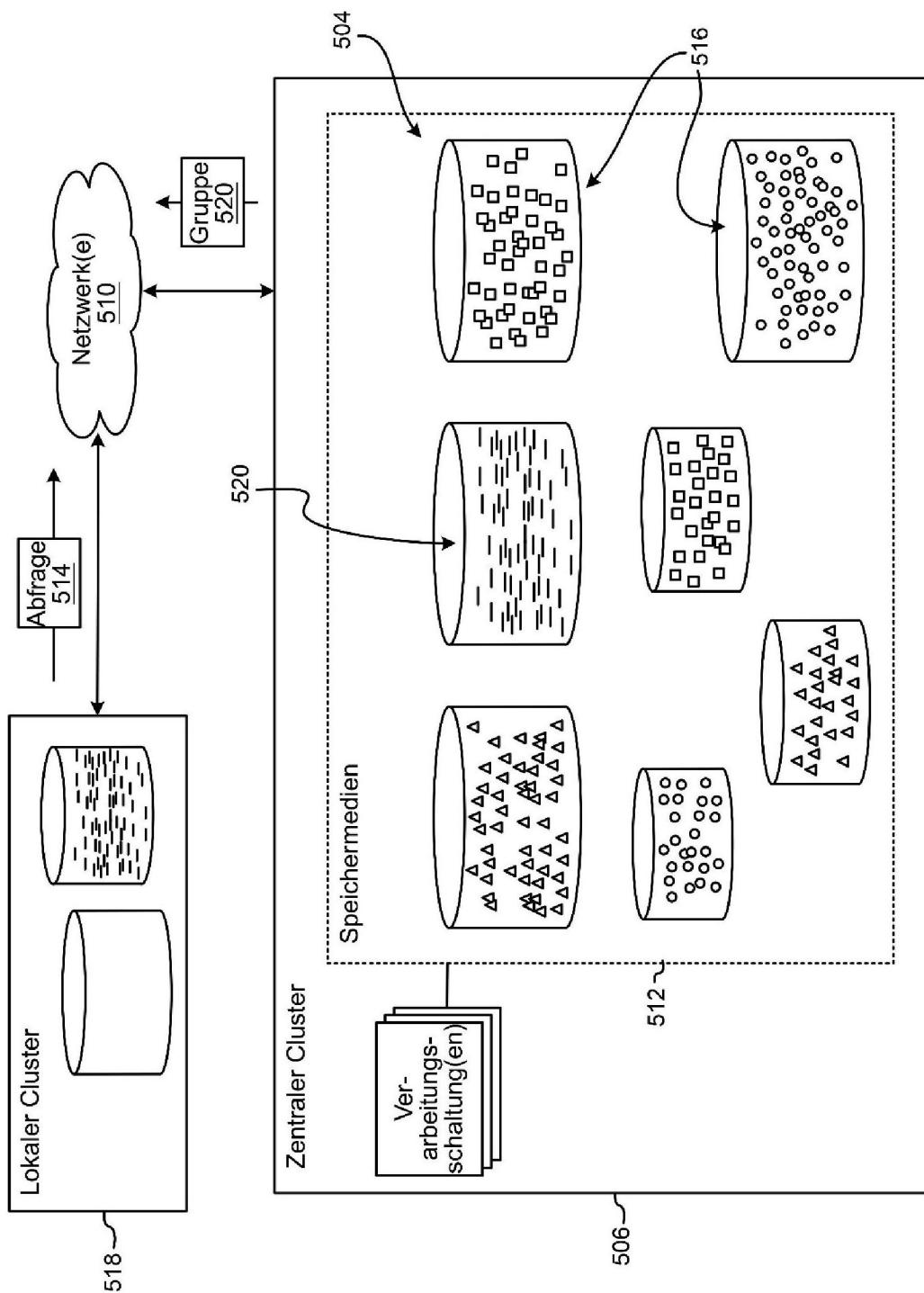


FIG. 5C

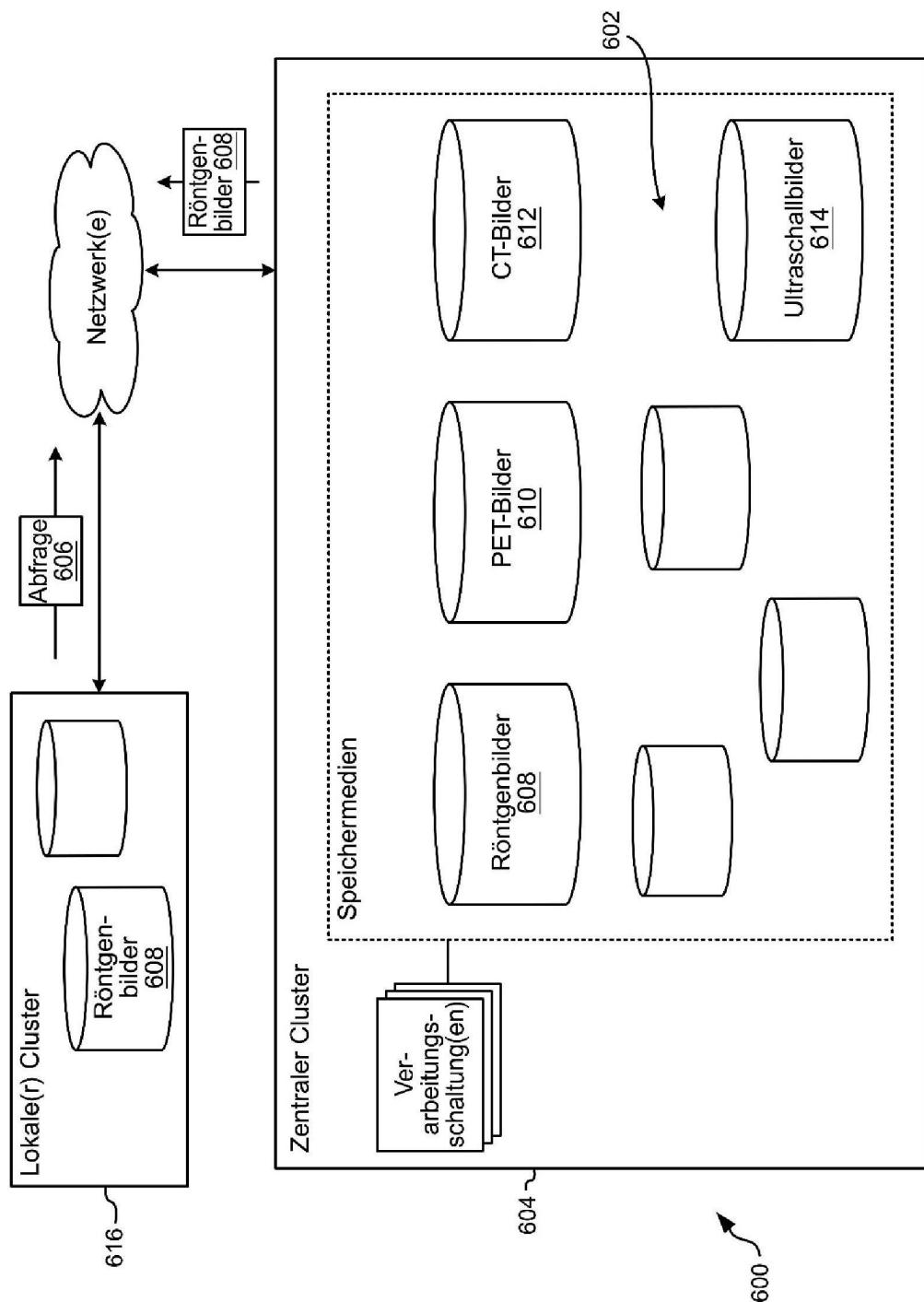
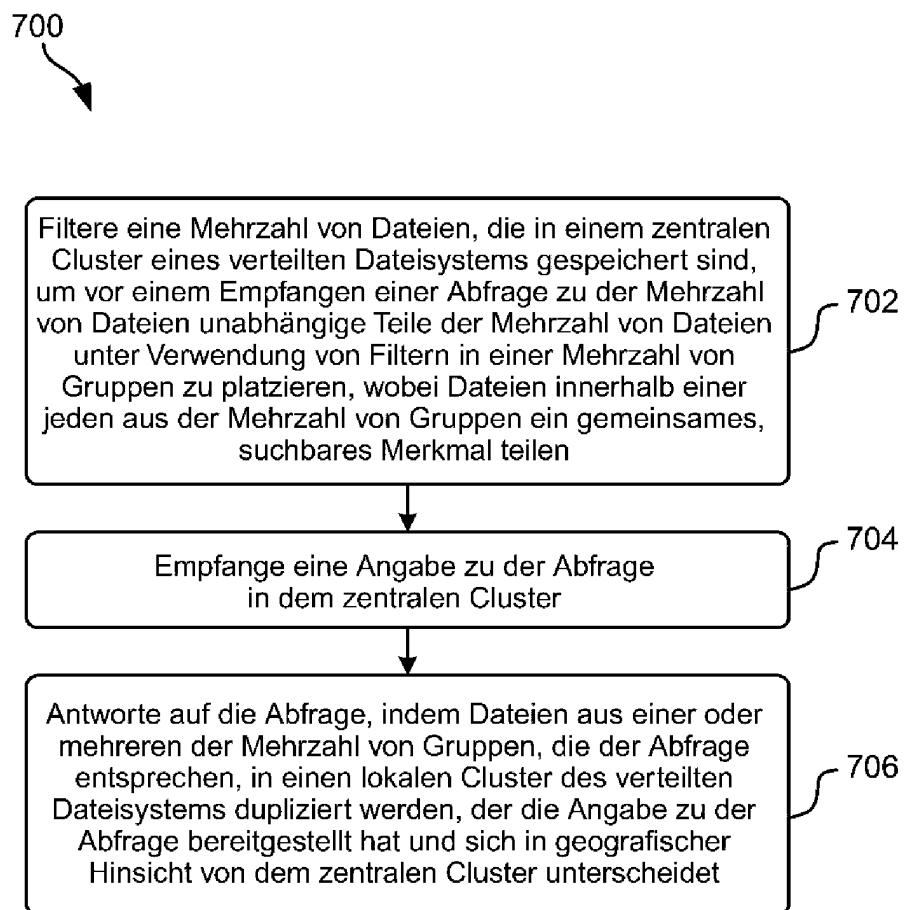
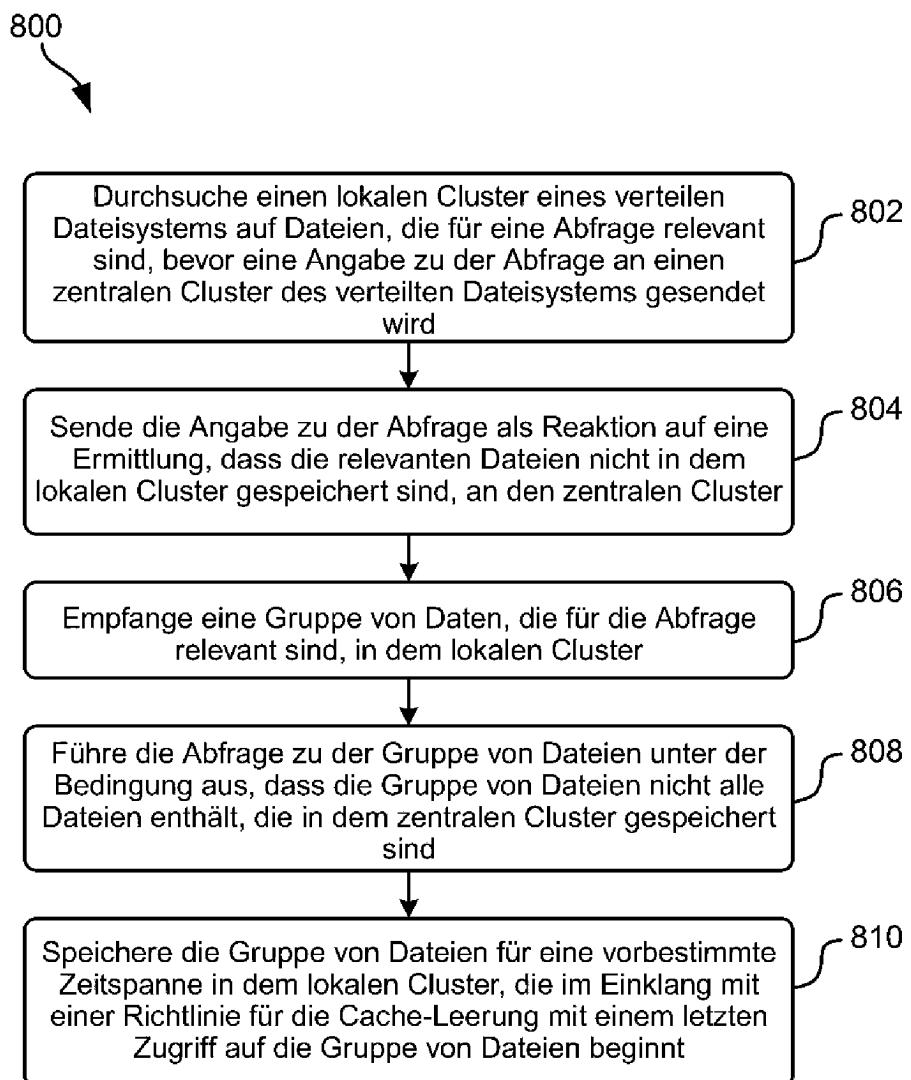


FIG. 6

**FIG. 7**

**FIG. 8**