

Cluster-, Grid- und Cloud-Computing (CGC)

25.1.2011

M.Sc. Christian Baun

Aufgabe 1 (1+1 Punkte)

- a) Erklären Sie, warum Data Grids keine File-Sharing-Anwendungen im klassischen Sinne sind.
- b) Welchen Stellenwert haben Data Grids und Compute Grids? Begründen Sie ihre Antwort.

Aufgabe 2 (4+4 Punkte)

Im LHC Computing Grid fallen jährlich 15 Petabyte Daten an, die gespeichert werden müssen. Wie hoch wäre ein Stapel, wenn zur Speicherung

- a) CDs verwendet würden (Kapazität: 650 Megabyte, Dicke: 1,05 mm)?
- b) Festplatten verwendet werden (Kapazität: 1 Terabyte, Dicke: 1,5 cm)?

Aufgabe 3 (1+2 Punkte)

- a) Was ist die Aufgabe des bekannten Cloud-Diensts `mturk` und welche beiden Rollen gibt es bei `mturk`?
- b) Nennen Sie die vier Arten von Clients beim Client-Server-Modell. Beschreiben Sie die vier Arten. Dabei sollen die Unterschiede klar werden.

Aufgabe 4 (1+1+1+5 Punkte)

- a) Warum führt das Gradientenverfahren bei Optimierungsproblemen nicht immer zum Erfolg?
- b) Wie arbeiten evolutionäre Algorithmen?
- c) Beschreiben Sie die Besonderheiten der funktionalen Programmierung.
- d) Beschreiben Sie den MapReduce-Algorithmus anhand seiner 5 Aktionen und Zwischenergebnisse.

Aufgabe 5 (3 Punkte)

- a) An wie vielen Instanzen kann ein EBS-Volume angehängt sein?
- b) Was ist die ACL und wozu dient sie?
- c) Was ist der ELB und wozu dient er?

Aufgabe 6 (3+1+1 Punkte)

- a) Nennen Sie 6 APIs die die Google App Engine zur Verfügung stellt und beschreiben Sie die Fähigkeiten und Eckdaten dieser APIs.
- b) Beschreiben Sie die zwei Möglichkeiten, einen Drucker mit Google Cloud Print zu verbinden.
- c) Beschreiben Sie wie Cloud Gaming Dienste funktionieren.

Aufgabe 7 (1,5 + 4,5 Punkte)

- a) Nennen Sie die drei unterschiedlichen Arten von P2P-Systemen zum File Sharing, die in der Vorlesung beschrieben wurden.
- b) Beschreiben Sie die drei unterschiedlichen P2P-Systeme anhand jeweils einer konkreten Implementierung. Die Funktion, sowie die Unterschiede, Vor- und Nachteile sollen klar ersichtlich sein.

Aufgabe 8 (4+4 Punkte)

Ein wissenschaftliches Experiment erzeugt 22 Petabyte Daten pro Jahr.

- a) Wie lange dauert die Übertragung in einem Ethernet (LAN) mit 1000 Mbit pro Sekunde?
- b) Wie lange dauert die Übertragung über ADSL mit 15.000 kbit pro Sekunde?

Aufgabe 9 (2+1 Punkte)

- a) Nennen Sie die vier HTTP-Methoden bei REST Web Services, die an die aus dem Datenbanken-Umfeld bekannten CRUD-Aktionen erinnern und beschreiben Sie kurz deren Funktion.
- b) Zusätzlich zu den vier HTTP-Methoden existieren zwei weitere HTTP-Methoden. Nennen Sie diese und beschreiben Sie deren Funktion.

Aufgabe 10 (2+2 Punkte)

Berechnen Sie Stromkosten. Was kostete der jährliche Dauerbetrieb (24/7) bei 0,22 €/kWh für ein Unternehmen XY mit 750 Computerarbeitsplätzen?:

- a) Szenario 1: Fat Clients (PC)
 - Elektrische Anschlussleistung pro Desktopsystem: 280 Watt
 - Elektrische Anschlussleistung pro Bildschirm: 80 Watt
- b) Szenario 2: Thin Clients mit Blade-Servern
 - Elektrische Anschlussleistung pro Thin Client: 30 Watt
 - Elektrische Anschlussleistung pro Bildschirm: 80 Watt
 - Elektrische Anschlussleistung pro Server-Blade: 180 Watt
 - Auf ein Server-Blade passen 30 virtuelle Desktopsysteme

Bedenken Sie, dass es auch Schaltjahre gibt!

Name:

Vorname:

Matr.Nr.:

Aufgabe 1)

Punkte:

a)

File Sharing Anwendungen basieren in der Regel auf Peer-to-Peer Netzwerken und sind keine Grids. File Sharing Anwendungen sind anwendungsspezifisch. Die Infrastruktur ist nicht flexibel nutzbar. Ein Grid muss nach Foster die folgenden drei Kriterien erfüllen:

Ein Grid...

...koordiniert Ressourcen, die nicht von einer zentralen Kontrollinstanz verwaltet werden...

...und verwendet offene, standardisierte Protokolle und Schnittstellen,...

...um nicht-triviale Dienstgütern bereitzustellen.

b)

Diese Klassifikationen haben in der Realität wenig Relevanz. Grid-Systeme verfolgen in der Realität weniger eingeschränkte, sondern eher universelle Ziele und können selten in derart engen Kategorien klassifiziert werden. Das Ziel von Grids ist in der Regel die gemeinsame Verwendung aller denkbaren IT-Ressourcen. Die Differenzierung in Data Grids, Computational Grids, etc. ist nicht zeitgemäß.

Name:

Vorname:

Matr.Nr.:

Aufgabe 2)

Punkte:

a)

1 Petabyte (PB) = 10^{15} Byte

1.000.000.000.000.000 Byte

Daten im LHC

15.000.000.000.000.000 Byte

Kapazität einer CD

650.000.000 Byte

Anzahl der CDs

$$\frac{15.000.000.000.000.000 \text{ Byte}}{650.000.000 \text{ Byte}} = 23.076.923,08$$

Höhe des CD-Stapel

$$\begin{aligned} 23.076.923,08 * 1,05 \text{ mm} &= 24.230.769,23 \text{ mm} \\ &= 24.230.769,23 \text{ cm} \\ &= 24.230,76923 \text{ m} \\ &= 24,23076923 \text{ km} \end{aligned}$$

b)

1 Petabyte (PB) = 10^{15} Byte

1.000.000.000.000.000 Byte

Daten im LHC

15.000.000.000.000.000 Byte

Kapazität einer 2 TB HDD

1.000.000.000.000 Byte

Anzahl der HDDs

$$\frac{15.000.000.000.000.000 \text{ Byte}}{1.000.000.000.000 \text{ Byte}} = 15.000$$

Höhe des HDD-Stapel

$$\begin{aligned} 15.000 * 1,5 \text{ cm} &= 22.500 \text{ cm} \\ &= 225 \text{ m} \end{aligned}$$

Name:

Vorname:

Matr.Nr.:

Aufgabe 3)

Punkte:

a)

Der Amazon Mechanical Turk ist ein Marktplatz für HuaaS/Crowdsourcing. Es gibt Auftraggeber und Auftragnehmer.

b)

- Bei einer verteilten Anwendung, die auf der Client-Server-Architektur basiert, lassen sich fünf Aufgabenbereiche voneinander trennen:
 - Darstellung der grafischen Benutzeroberfläche
 - Benutzerinterface (GUI)
 - Verarbeitung
 - Datenmanagement
 - Datenspeicherung
- Die Verteilung der einzelnen Aufgaben auf Clients und Server legt fest, um welche Art von Clients es sich handelt
- Entsprechend ihrer Aufgabenbereiche, werden vier Arten von Clients unterschieden:
 - **X-Terminal** oder **Text-Terminal**. Clients sind nur für die Darstellung der (grafischen) Benutzeroberfläche und die Übertragung der Benutzereingaben an den Server zuständig. Bereitstellung der GUI, Verarbeitung, Datenverarbeitung und Datenspeicherung übernimmt der Server
 - **Thin Clients** sind für die Darstellung der grafischen Benutzeroberfläche und Bereitstellung der GUI zuständig
 - **Applet Clients**. Häufig als **Network Computer** bezeichnet. Die Clients übernehmen die Darstellung der grafischen Benutzeroberfläche, Bereitstellung der GUI und einen Teil der Verarbeitung. Die Anwendungen (Applets) verarbeiten die Clients selbst
 - **Fat Clients**. Nur Datenmanagement und Datenhaltung befindet sich auf dem (File- oder Datenbank-)Server

Name:

Vorname:

Matr.Nr.:

Aufgabe 4)

Punkte:

a)

In flachen Plateaus ist der Gradient sehr klein und das Verfahren stagniert nahezu. Ein weiteres Problem ist, dass der Gradient in einem lokalen Optimum hängen bleiben kann und das globale Optimum so nicht findet.

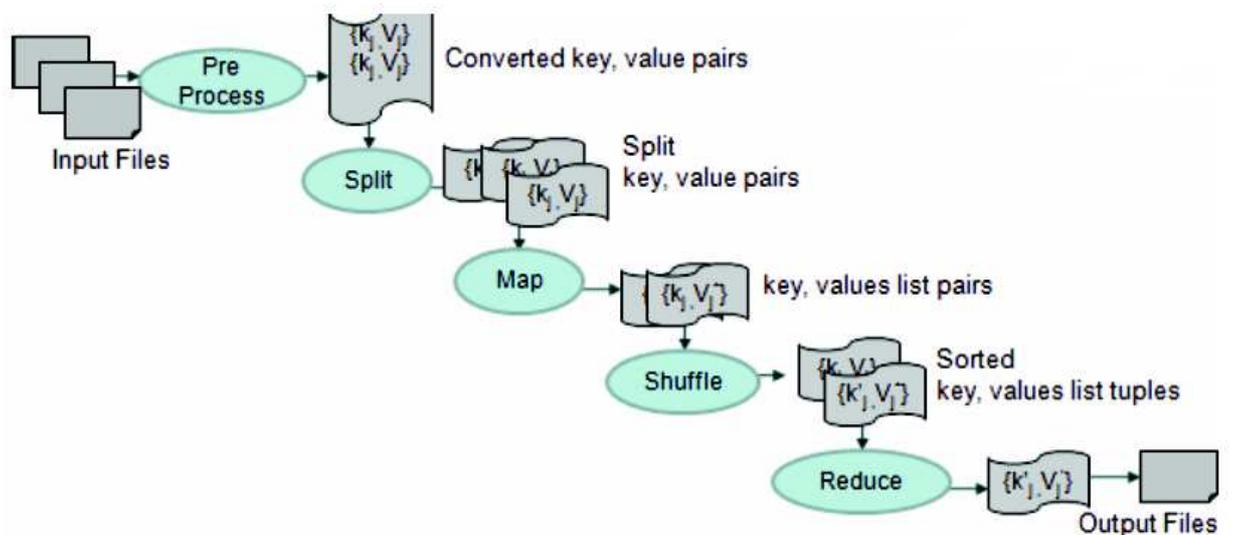
b)

Ein Evolutionärer Algorithmus hat die biologische Evolution als Vorbild. Individuen werden durch ihre Eigenschaften beschrieben und müssen sich bzgl. der Selektionsbedingungen als möglichst geeignet behaupten. Dementsprechend dürfen sie ihre Eigenschaften vererben oder nicht vererben. Durch mehrere Durchläufe von Mutation, Rekombination und Selektion entwickelt sich so die Bevölkerung immer näher an das gesuchte Optimum.

c)

Datenstrukturen werden nicht verändert. Stattdessen werden neue Datenstrukturen erzeugt. Die Reihenfolge von Operationen spielt keine Rolle für das Endergebnis. Dadurch gute Eignung für parallele Ausführung.

d)



Name:

Vorname:

Matr.Nr.:

Aufgabe 5)

Punkte:

a)

Ein EBS-Volume kann immer an genau eine Instanz angehängt werden. Dafür müssen sich Volume und Instanz zu innerhalb der gleichen Verfügbarkeitszone befinden.

b)

Für jedes Objekt (Key) und jeden Bucket können die Zugriffsrechte mit Hilfe der Access Control List (ACL) festgelegt werden.

c)

Der Dienst Elastic Load Balancing (ELB) von Amazon ist ein Teil von EC2. Innerhalb jeder Verfügbarkeitszone können Benutzer elastische Lastverteiler erzeugen. Jedem Lastverteiler weist der Benutzer einen Pool an Instanzen zu. Ein elastischer Lastverteiler verteilt automatisch die eingehenden Anfragen über die EC2-Instanzen seines Pools.

Name:

Vorname:

Matr.Nr.:

Aufgabe 6)

Punkte:

a)

- Authentifizierung
 - Authentifizierung/Authorisierung ist über einen Google-Account möglich
 - Keine Implementierung einer eigenen Authentifizierungslösung notwendig
- Datastore
 - Persistenter Speicher, realisiert als Key/Value-Datenbank
 - Transaktionen sind atomar
 - Definition, Abfrage und Manipulation von Daten erfolgt über eine eigene Sprache, die GQL (Google Query Language)
- Memcache
 - Hochperformanter temporärer Datenspeicher aus Hauptspeicher
 - Es wird eine Verfallszeit in Sekunden angegeben, wann der Eintrag aus dem Memcache entfernt werden soll
 - Daten werden je nach Auslastung des Memcache früher wieder verdrängt
- URL Fetch
 - Zugriff auf Inhalte im Internet
 - Kommunikation über RESTful Web Services möglich
- Mail
 - Emails versenden und empfangen über das GMail Gateway
 - Als Absender ist nur die E-Mail Adresse des eingeloggtten Benutzers oder des Administrators der Anwendung erlaubt
- Bildmanipulation
 - Funktionen um Bilder zu drehen, spiegeln, beschneiden und zum Verändern der Bildgröße
- Nachrichtenübermittlung via XMPP
 - Nachrichten können an jedes XMPP-kompatibles Nachrichtensystem (z.B. Google Talk) gesendet und von diesem empfangen werden

b)

- **Google Cloud Print kompatible Netzwerkdrucker.** Der Drucker ist beim Dienst angemeldet. Druckjobs werden an einen Dienst gesendet. Der Dienst bereitet die Druckaufträge auf und leitet diese an den Drucker weiter.
- **Legacy Drucker** die nicht zu Google Cloud Print kompatibel sind können über einen Proxy angesprochen werden. Der Proxy registriert den Drucker und sendet Druckaufträge an den Dienst. Aufbereitete Druckjobs werden über den Proxy an den Drucker gesendet.

c)

Cloud Gaming bezeichnet Dienste, um High-End Videospiele auf Low-End Geräten (ältere PCs, Fernseher, mobile internetfähige Endgeräte) verfügbar zu machen. Videospiele werden in den Serverfarmen des Anbieters ausgeführt. Ausschließlich die Darstellung erfolgt auf den lokalen Zielgeräten. Die Darstellung wird als komprimierter Videostrom übertragen. Benutzereingaben werden zum Anbieter gesendet und dort ausgewertet.

Name:

Vorname:

Matr.Nr.:

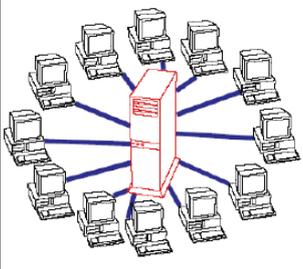
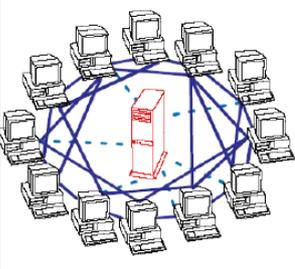
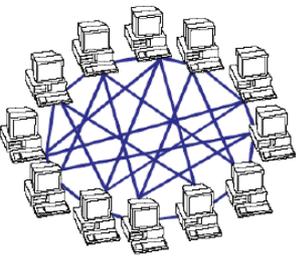
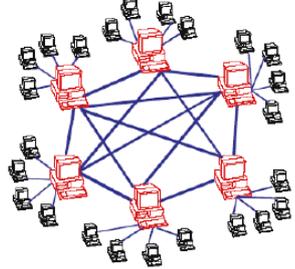
Aufgabe 7)

Punkte:

a)

- Zentralisiertes P2P
- Reines bzw. Pures P2P
- Hybrides P2P

b)

<i>Client-Server</i>	<i>Peer-to-Peer</i>		
	1. Resources are shared between the peers 2. Resources can be accessed directly from other peers 3. Peer is provider and requestor (Servent concept)		
	<i>1st Generation</i>		<i>2nd Generation</i>
1. Server is the central entity and only provider of service and content. → Network managed by the Server 2. Server as the higher performance system. 3. Clients as the lower performance system Example: WWW	<i>Centralized P2P</i>	<i>Pure P2P</i>	<i>Hybrid P2P</i>
	1. All features of Peer-to-Peer included 2. Central entity is necessary to provide the service 3. Central entity is some kind of index/group database Example: Napster	1. All features of Peer-to-Peer included 2. Any terminal entity can be removed without loss of functionality 3. → No central entities Examples: Gnutella 0.4, Freenet	1. All features of Peer-to-Peer included 2. Any terminal entity can be removed without loss of functionality 3. → dynamic central entities Example: Gnutella 0.6, JXTA
			

Quelle: Jörg Eberspächer und Rüdiger Schollmeier. *First and Second Generation of Peer-to-Peer Systems* (2005). LNCS 3485

Name:

Vorname:

Matr.Nr.:

Aufgabe 8)

Punkte:

a)

1 Gigabit (Gbit) = 10^9 Bit 1.000.000.000 Bit

Daten im Experiment 22.000.000.000.000.000 Byte

Bandbreite des Ethernet (1 Gbit/s) 1.000.000.000 Bit/s

Bandbreite des Ethernet in Byte/s 125.000.000 Byte/s

Dauer der Datenübertragung [s] = 176.000.000 : 60

Dauer der Datenübertragung [m] $\approx 2.933.333,33 : 60$

Dauer der Datenübertragung [h] $\approx 48.888,89 : 24$

Dauer der Datenübertragung [d] $\approx 2.037,037 : 365,25$

Dauer der Datenübertragung [y] $\approx 5,577$

b)

1 Megabit (Mbit) = 10^6 Bit 1.000.000 Bit

Daten im Experiment 22.000.000.000.000.000 Byte

Bandbreite des ADSL (16 Mbit/s) 15.000.000 Bit/s

Bandbreite des Ethernet in Byte/s 1.875.000 Byte/s

Dauer der Datenübertragung [s] $\approx 1,1733 * 10^{10} : 60$

Dauer der Datenübertragung [m] $\approx 195.555.555,5 : 60$

Dauer der Datenübertragung [h] $\approx 3.259.259,258 : 24$

Dauer der Datenübertragung [d] $\approx 135.802,4691 : 365,25$

Dauer der Datenübertragung [y] $\approx 371,8068969$

Name:

Vorname:

Matr.Nr.:

Aufgabe 9)

Punkte:

a)

HTTP	CRUD	SQL	Beschreibung
PUT	Create	INSERT	Neue Ressource erzeugen oder deren Inhalt ersetzen
GET	Read/Retrieve	SELECT	Ressource bzw. deren Repräsentation anfordern
POST	Update	UPDATE	Ressource etwas hinzufügen
DELETE	Delete/Destroy	DELETE	Eine Ressource löschen

b)

- **HEAD** fordert vom Server nur den Header einer Ressource (Datei) an
 - So kann sich der Benutzer des Web-Service über die Metadaten informieren, ohne die eigentlichen Ressource zu übertragen
 - Es wird der gleiche Header zurückgeliefert wie bei **GET**
- **OPTIONS** prüft welche Methoden auf einer Ressource verfügbar sind

Name:

Vorname:

Matr.Nr.:

Aufgabe 10)

Punkte:

a)

- Stromkosten (mit Schaltjahr) pro Computerarbeitsplatz pro Jahr

$$0,36 \text{ kW} * 24 \frac{\text{h}}{\text{Tag}} * 365,25 \frac{\text{Tag}}{\text{Jahr}} * 0,22 \frac{\text{€}}{\text{kWh}} = 694,2672 \frac{\text{€}}{\text{Jahr}}$$

- Stromkosten für 750 Computerarbeitsplätze pro Jahr

$$750 * 694,2672 \frac{\text{€}}{\text{Jahr}} = 520.700,4 \frac{\text{€}}{\text{Jahr}}$$

b)

- Stromkosten (mit Schaltjahr) pro Computerarbeitsplatz (ohne Server) pro Jahr

$$0,11 \text{ kW} * 24 \frac{\text{h}}{\text{Tag}} * 365,25 \frac{\text{Tag}}{\text{Jahr}} * 0,22 \frac{\text{€}}{\text{kWh}} = 212,1372 \frac{\text{€}}{\text{Jahr}}$$

- Stromkosten (mit Schaltjahr) pro Server-Blade pro Jahr

$$0,18 \text{ kW} * 24 \frac{\text{h}}{\text{Tag}} * 365,25 \frac{\text{Tag}}{\text{Jahr}} * 0,22 \frac{\text{€}}{\text{kWh}} = 347,1336 \frac{\text{€}}{\text{Jahr}}$$

- Stromkosten für 750 Computerarbeitsplätze (ohne Server) pro Jahr

$$750 * 212,1372 \frac{\text{€}}{\text{Jahr}} = 159.102,9 \frac{\text{€}}{\text{Jahr}}$$

- Stromkosten für 25 Server-Blades pro Jahr

$$25 * 347,1336 \frac{\text{€}}{\text{Jahr}} = 8.678,34 \frac{\text{€}}{\text{Jahr}}$$

- Stromkosten für Computerarbeitsplätze und Server-Blades pro Jahr

$$159.102,9 \frac{\text{€}}{\text{Jahr}} + 8.678,34 \frac{\text{€}}{\text{Jahr}} = 167.781,24 \frac{\text{€}}{\text{Jahr}}$$