

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

8.7.2011

M.Sc. Christian Baun, Dipl.-Phys. Viktor Mauch

Aufgabe 1 (1+2+1+2+1+2+2+1+2 Punkte)

- In welcher Kategorien von Cloud-Diensten wird menschliche Kreativität zu geringen Kosten oder als Spende Freiwilliger angeboten?
- Warum ist der Begriff Cloud-Betriebssystem in den meisten Fällen irreführend?
- In welcher Kategorie von Cloud-Diensten können die Kunden virtuelle Serverinstanzen betreiben und sogar virtuelle Rechenzentren realisieren?
- Was ist eine PaaS und was kann man damit machen?
- Was brauchen die Kunden, um mit Softwarediensten zu arbeiten?
- Was ist der Hauptunterschied zwischen Public und Private Cloud?
- Nennen Sie je zwei Vor- und Nachteile von Public und Private Clouds
- Was ist eine Hybrid Cloud?
- Warum ist es in manchen Szenarien sinnvoll, Daten in S3 abzulegen, anstatt sie einfach im Speicher einer Instanz oder in einem EBS-Volumen vorzuhalten?

Aufgabe 2 (2+1+2 Punkte)

- Was sagt Amdahls Gesetz aus und wofür wird es verwendet?
- Welche Unwägbarkeiten existieren bei Amdahls Gesetz?
- Was sagt Gustafsons Gesetz aus und was ist der Unterschied zu Amdahls Gesetz?

Aufgabe 3 (6+4 Punkte)

- Nennen Sie die drei Punkte der „Three Point Checklist“ von Ian Foster, die jedes Grid erfüllen muss. Beschreiben Sie auch die Bedeutung jedes der drei Punkte in jeweils 1-2 Sätzen.
- Ordnen Sie die Eigenschaften in der Tabelle jeweils dem Cloud-Computing oder dem Grid-Computing zu. (Es genügt, wenn Sie jeweils „C“ für Cloud Computing und „G“ für Grid Computing eintragen.)

Eigenschaft	Cloud/Grid Computing
Hauptsächlich physische Ressourcen	Grid
Vollautomatisierte Dienste	Cloud
Verteilte, heterogene Ressourcen ohne zentrale Kontrolle	Grid
Basiert auf freier, standardisierter Software und Schnittstellen	Grid
Benutzerfreundliche Bedienung	Cloud
Verbrauchsabhängige Abrechnung	Cloud
Hauptsächlich virtualisierte Ressourcen	Cloud
Finanzierung primär Förderung durch die öffentliche Hand	Grid

Für jede korrekte Antwort gibt es 0.5 Punkte. Für jede falsche Antwort werden 0.5 Punkte abgezogen. Es können maximal 4 Punkte und nicht weniger als 0 Punkte insgesamt erreicht werden.

Aufgabe 4 (2+2+4+2+3 Punkte)

- Was versteht man unter einem Feuerabendcluster?
- Was ist das Ziel beim High Availability Clustering und wie wird dieses Ziel erreicht?
- Es existieren zwei Gruppen von High Availability Clustern, die sich in ihrem Verhalten bei Ausfällen von Knoten unterscheiden.
 - Welche beiden Gruppen sind das?
 - Was sind die Unterschiede zwischen den beiden Gruppen?
- Was versteht man unter Failover und Failback?
- Was versteht man unter einem Split Brain? Wie kann es dazu kommen und was ist das Resultat?

Aufgabe 5 (2+3+3+3 Punkte)

- Was versteht man unter einem Peer-to-Peer-System und was ist die Besonderheit gegenüber anderen verteilten Systemen?
- Welche drei Arten von Peer-to-Peer-Systemen existieren? (*Gesucht sind die Namen!*)
- Worin unterscheiden sich die drei Arten von Peer-to-Peer-Systemen?
- Nennen Sie zu jeder der drei Arten von Peer-to-Peer-Systemen mindestens einen Vor- und einen Nachteil.

Aufgabe 6 (3 Punkte)

Kreuzen Sie bei jeder Aussage in der Tabelle an, ob sie wahr oder falsch ist.

Aussage	wahr	falsch
Die Google App Engine ist eine „Software as a Service“		X
Buckets in S3 haben einen hierarchischen Namensraum		X
Man kann die Firewall-Einstellungen von EC2-Instanzen mit Hilfe von Sicherheitsgruppen konfigurieren	X	
MapReduce ist eine freie Implementierung des Google File Systems (GFS)		X
Im GFS verwaltet der GFS-Master die Metadaten (z.B. Anzahl und Lokation der Chunks)	X	
EBS-Volumen können zu jedem Zeitpunkt nur an eine Instanz angehängt sein	X	

Für jede korrekte Antwort gibt es 0.5 Punkte. Für jede falsche Antwort werden 0.5 Punkte abgezogen. Es können maximal 3 Punkte und nicht weniger als 0 Punkte insgesamt erreicht werden.

Aufgabe 7 (4 Punkte)

Bewerten Sie die Eigenschaften in der Tabelle dahingehend, ob sie eher zu asymmetrischen Kryptosystemen oder symmetrischen Kryptosystemen gehören.

(Es genügt, wenn Sie jeweils „A“ für asymmetrische und „S“ für symmetrische Kryptosysteme eintragen.)

Eigenschaft	A / S
Beide Kommunikationsteilnehmer nutzen zur Ver- und Entschlüsselung den selben Schlüssel	S
Nachrichtenkanal findet nur in eine Richtung statt, viele-zu-eins	A
Wird auch für das Signieren von Nachrichten verwendet	A
Bekanntes Verfahren: RSA	A
Ermöglicht Kanal in beide Richtungen zweier Kommunikationsteilnehmer	S
Es wird ein Schlüsselpaar (Public und Private Key) verwendet	A
Ver- und Entschlüsselung sind auf Geschwindigkeit optimiert	S
Bekanntes Verfahren: AES	S

Für jede korrekte Antwort gibt es 0.5 Punkte. Für jede falsche Antwort werden 0.5 Punkte abgezogen. Es können maximal 4 Punkte und nicht weniger als 0 Punkte insgesamt erreicht werden.

Name:

Vorname:

Matr.Nr.:

Aufgabe 1)

Punkte:

- a) Humans as a Service (HuaaS)
- b) Auch für die Nutzung eines Cloud-Betriebssystems ist ein Rechner mit Browser und daher ein zugrunde liegendes Betriebssystem nötig. Das native Betriebssystem wird nicht ersetzt. Es werden nur die Benutzeranwendungen und Benutzerdaten ausgelagert.
- c) Infrastructure as a Service (IaaS)
- d) Platform as a Server. Ein Anbieter betreibt skalierbare Laufzeitumgebungen für eine oder mehr Programmiersprachen. Die Kunden können Ihre Webanwendungen auf den Servern des Anbieters betreiben.
- e) Die Kunden benötigen nur einen Browser.
- f) Bei einer Public Cloud gehören Anbieter und Kunden unterschiedlichen Organisationen an. Bei einer Private Cloud gehören Anbieter und Benutzer der gleichen Organisation an.
- g)
 - Public Cloud
 - Keine Kosten für Anschaffung, Betrieb und Wartung eigener Hardware.
 - Ressourcen sind sofort einsatzbereit und unbegrenzt verfügbar.
 - Angst vor mangelnder Datensicherheit und Lock-in.
 - Auslagerung von Ressourcen ist nicht immer möglich (Datenschutz).
 - Private Cloud
 - Keine Probleme mit Lock-in und Datenschutz.
 - Kosten ähnlich einer nicht-Cloud-basierten Architektur.
 - Softwarequalität der freien Projekte teilweise verbesserungswürdig.
- h) Public und Private Cloud-Dienste werden gemeinsam verwendet.
- i) Daten in S3 sind immer über das Internet erreichbar und unabhängig von Serverinstanzen. Beim Instanzspeicher sind die Daten weg, wenn die Instanz weg ist. Bei EBS braucht man immer eine laufende Instanz mit Serverdiensten, um die Daten über das Internet zu erreichen.

Name:

Vorname:

Matr.Nr.:

Aufgabe 2)

Punkte:

- a) Mit dem Amdahlschen Gesetz wird die maximale zu erwartende Beschleunigung von Programmen durch parallele Ausführung auf mehreren Prozessoren berechnet. Nach Amdahl wird der Geschwindigkeitszuwachs vor allem durch den sequentiellen Anteil des Problems beschränkt. Ein Programm kann nie vollständig parallel ausgeführt werden.
- b) Amdahls Gesetz berücksichtigt nicht den Cache und die damit verbundenen Auswirkungen innerhalb der Praxis. Durch eine steigende Anzahl an CPUs vergrößert sich auch die verfügbare Menge des schnellen Speichers. Im günstigsten Fall ist es möglich, die gesamte Problemgröße im Cache anstatt im langsamen Hauptspeicher zu halten. In einem solchen Fall kann es (sehr selten!) zum super-linearen SpeedUp kommen, also einer Beschleunigung, die über die zusätzliche, reine Rechenleistung hinausgeht.
- c) Gustafsons Gesetz besagt, dass ein genügend großes Problem effizient parallelisiert werden kann. Der Unterschied zu Amdahl ist, dass der parallele Anteil mit der Anzahl der Prozessoren wächst. Der sequentielle Teil wirkt nicht beschränkend, da er mit zunehmender Anzahl an Prozessoren unbedeutender wird. Geht die Anzahl der Prozessoren gegen unendlich, wächst der SpeedUp linear mit der Anzahl der Prozessoren.

Name:

Vorname:

Matr.Nr.:

Aufgabe 3)

Punkte:

a) Ein Grid...

- **...koordiniert Ressourcen, die nicht von einer zentralen Kontrollinstanz verwaltet werden...**
 - Ressourcen können sein: Cluster, Speicher, Daten(-banken), Anwendungen (Lizenzen), Datenquellen (wissenschaftliche Messgeräte),...
- **...und verwendet offene, standardisierte Protokolle und Schnittstellen,...**
 - Protokolle und Schnittstellen müssen offen sein und Funktionen für Authentifizierung, Autorisierung, Ressourcen-Ermittlung und -Zugriff bereitstellen
- **...um nicht-triviale Dienstgütern bereitzustellen.**
 - Dienstgütern können sich u.a. auf Datendurchsatz, Antwortzeit, Erreichbarkeit der Dienste und den Grad der Datensicherheit beziehen

Name:

Vorname:

Matr.Nr.:

Aufgabe 4)

Punkte:

- a) Die Knoten eines Feierabendclusters stehen nur zu bestimmten Zeiten zur Verfügung. Die Knoten werden während der üblichen Arbeitszeiten als normale Arbeitsrechner eingesetzt.
- b) Beim High Availability Clustering sind Systeme mit einer hohen Ausfallsicherheit, also Verfügbarkeit, gefordert. Erreicht wird das durch die Redundanz der Knoten und ihrer Komponenten sowie das Ausschließen eines Single Point of Failure
- c) Active/Passive-Cluster und Active/Active-Cluster
 - **Active/Passive-Cluster**
 - Im Normalbetrieb ist mindestens ein Knoten im Zustand Passiv
 - Knoten im Zustand Passiv übernehmen im Normalbetrieb keine Dienste
 - Fällt ein Knoten aus, übernimmt ein Passiv-Knoten dessen Dienste
 - Vorteil: Die Dienste müssen nicht für Cluster-Betrieb ausgelegt sein
 - Nachteil: Viel potentielle Leistung bleibt im Normalbetrieb ungenutzt
 - **Active/Active-Cluster**
 - Auf allen Knoten laufen die gleichen Dienste
 - Alle Knoten sind im Zustand Aktiv
 - Fallen Knoten aus, müssen die noch aktiven Knoten deren Aufgaben mit übernehmen
 - Vorteil: Bessere Lastverteilung zwischen den Knoten
 - Nachteil: Dienste müssen für den Betrieb im Cluster ausgelegt sein, da zeitgleich alle Knoten auf die gleichen Ressourcen (Daten!) zugreifen
- d) Als Failover bezeichnet man die Fähigkeit, beim Ausfall eines Knotens alle Aufgaben automatisch einem anderen Knoten zu übergeben und so die Ausfallzeit zu minimieren. Sind ausgefallene Knoten wieder einsatzbereit, melden sich diese am Lastverteiler zurück und erhalten wieder Jobs. Dieses Verhalten wird als Failback bezeichnet.
- e) Split Brain bezeichnet die Trennung der Kommunikation zwischen Knoten, obwohl die Rechner selbst einwandfrei funktionieren. Kontrollprogramme wie Heartbeat gehen davon aus, dass es zum Ausfall von Knoten kam. Jeder Node erklärt sich zum Primary Node. Bei Active/Passive-Clustern führt das zum Ausfall des Clusters bzw. der angebotenen Dienste. Bei gemeinsamem Speicher versucht jeder Knoten auf den Speicher zu schreiben. Bei verteiltem Speicher verteilen sich die Schreibzugriffe und es entstehen unterschiedliche Datenbestände. Die Konsistenz der Daten ist nicht mehr gewährleistet. Eine Reperatur ist schwierig (zeitaufwendig) bis unmöglich.

Name:

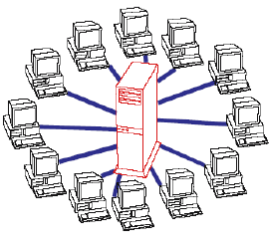
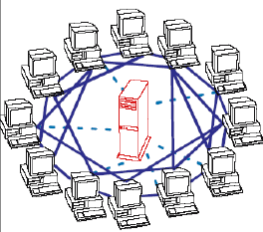
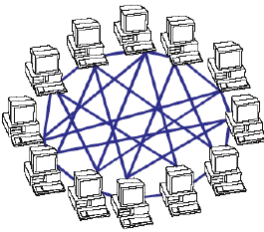
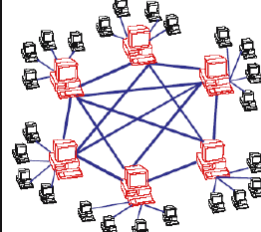
Vorname:

Matr.Nr.:

Aufgabe 5)

Punkte:

- a) Ein Peer-to-Peer-System ist ein Verbund gleichberechtigter Knoten. Die Knoten werden als Peers bezeichnet und machen sich gegenseitig Ressourcen zugänglich. Jeder Knoten ist gleichzeitig Client und Server.
- b) Zentralisiertes P2P, pures/reines P2P und hybrides P2P
- c)

Client-Server	Peer-to-Peer		
	1. Resources are shared between the peers 2. Resources can be accessed directly from other peers 3. Peer is provider and requestor (Serving concept)		
	1st Generation		2nd Generation
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
			

d)

- Vor- und Nachteile von zentralisiertem P2P
 - Vorteile:
 - * Peers kommunizieren direkt miteinander (Server hat nur wenige Aufgaben).
 - * Gute Erweiterbarkeit. Jeder weitere Peer verursacht nur wenig Last auf dem Server.
 - * Ausfall einzelner oder mehrerer Peers kann dem Netzwerk und der Verfügbarkeit der Dienste nichts anhaben.
 - * Zentralisierte Dienste bieten eine hohe Leistung.
 - * Zentralisierter Dienst ist ein bekannter Einstiegspunkt. Neue Peers können leicht Teil des Systems werden.
 - Nachteile:
 - * Server verursachen Kosten (Strom, Stellfläche, Administration,...)
 - * Ohne die zentralisierten Dienste funktioniert das System nicht.
 - * Zentralisierte Dienste sind immer ein Angriffspunkt \implies Juristische Angriffsfläche.
- Vor- und Nachteile von reinem/purem P2P
 - Vorteile:
 - * Keine Abhängigkeit von Servern. Peers kommunizieren direkt miteinander.
 - * Keine zentralen Angriffspunkte.
 - * System ist kaum zu zerstören. Ausfall einzelner Peers kann dem System und der Verfügbarkeit der Dienste nichts anhaben.
 - * Höhere Selbstbestimmung der Benutzer. Bei Client-Server muss bei steigender Anzahl von Clients auch die Serverleistung steigen.
 - * Kein Administrationsaufwand, da keine Server.
 - Nachteile:
 - * Schwieriger Einstieg. Es muss mindestens ein Peer bekannt sein.
 - * Suchanfragen sind eine starke Belastung für das Netzwerk.
 - * Listen von Daten und Peers müssen verteilt werden (hoher Overhead).
- Vor- und Nachteile von hybridem P2P
 - Vorteile:
 - * Die Vorteile von reinem P2P bleiben erhalten.
 - Nachteile:
 - * Es gibt (dynamische) Server. Das System wird dadurch komplexer.