



Grid, Cloud und Peer to Peer

Hochleistungs-Ein-/Ausgabe

Michael Kuhn

Wissenschaftliches Rechnen
Fachbereich Informatik
Universität Hamburg

2016-06-17

1 Grid, Cloud und Peer to Peer

- Orientierung
- Einleitung
- Grid
- Cloud
- Peer to Peer
- Zusammenfassung

2 Quellen

Motivation

- Hohe Leistungsfähigkeit immer häufiger nötig
 - Komplexe Simulationen etc.
- Nicht immer ist ein Supercomputer vor Ort verfügbar
 - Oder Computer vor Ort alleine nicht leistungsfähig genug
- Verschiedene Konzepte zur Nutzung fremder Ressourcen
 - Konkret Grid, Cloud und Peer to Peer

Peer to Peer

- Koordination Gleichgestellter (Peers)
 - Im Gegensatz zu Client-Server-Systemen
- Viele Einsatzgebiete
 - Streaming, Bitcoin, ...
 - Hauptsächlich Bereitstellung von Daten
- Sehr weite Verbreitung
 - $\approx 10\%$ des weltweiten Datentransfers

Probleme...

- Neben Durchsatz auch Kapazität ein Problem
 - Daten sind möglicherweise sehr groß
- Originaldaten befinden sich auf Rechner A
 - Rechner A hat großes Speichersystem
- Daten werden auf Rechner B transferiert
 - Temporäre Speicherung belegt Platz
 - Ausreichend Platz für Ergebnisdaten notwendig

Typen

- Rechen-Grids (Computing)
 - Rechenkapazität aus der Steckdose
- Daten-Grids (Data)
 - Datenbestände werden verknüpft
 - Zugang häufig über Grid-Portal
- Ressourcen-Grids (Resource)
- Dienstleistungs-Grids (Service)
- Wissens-Grids (Knowledge)

GridFTP

- Benötigt ein Zertifikat der virtuellen Institution
 - Teilweise recht aufwendig
 - Vorlegen des Personalausweises etc.
- Danach Erzeugung eines temporären Proxys
 - Aufruf: `grid-proxy-init`
 - Status anzeigen mit `grid-proxy-info`
- Danach Datentransfer möglich
 - Aufruf: `globus-url-copy source destination`

GridFTP...

- Setzt üblicherweise auf normalen Dateisystemen auf
 - Erweiterung von FTP
 - Unterstützung von Grid-Authentifizierung
- Kompatibilität zwischen verschiedenen Institutionen
 - Immer noch institutionsspezifische Pfade etc.
 - Jobscripte müssen unter Umständen angepasst werden
 - Im Hochleistungsrechnen üblicherweise mehrere Konfigurationen für verschiedene Hochleistungsrechner

MPICH-G2

- MPICH mit Grid-Unterstützung
 - Veraltet, basiert auf MPI 1.1
- Grid-Techniken für bessere Integration
 - Starten von Prozessen auf entfernten Systemen
 - Staging von Programmen und Daten
 - Sicherheit
- Automatische Auswahl der Kommunikationsmethode
 - Hochleistungsvernetzung innerhalb des Clusters
 - IP zwischen Clustern



Überblick

- Ähnliches Konzept wie Grid
 - Berechnung und Daten „in der Wolke“
- Keine genaue Kenntnis über Ressourcen notwendig
 - Automatische Ausführung auf verfügbaren Ressourcen
- Im Gegensatz zu Grid zentralisierter Ansatz
 - Anbieter kontrolliert Ressourcen
- Populär im kommerziellen Sektor
 - Amazon, Google, Microsoft, Backblaze, ...

Everything as a Service [2]

- Infrastructure (IaaS)
 - Zugriff auf (virtualisierte) Hardware
 - Eigenes Betriebssystem etc.
- Platform (PaaS)
 - Durch Cloud-Anbieter definierte Plattform
 - Erlaubt Anwendungen darauf zu entwickeln
- Software (SaaS)
 - Zugriff auf Software
 - Auch Software on Demand



Everything as a Service...

- `/dev/null` 😊
 - 25 GB pro Monat kostenlos
 - “We support BigData!”
 - “Run huge Map-Reduce jobs on the data you won’t see anymore!”
 - LD_PRELOAD-Bibliothek für transparente Nutzung
 - Webseite: <https://devnull-as-a-service.com/>

Liefermodelle [2]

- Public Cloud
 - Öffentlich zugänglich
 - Üblicherweise verbrauchsabhängige Bezahlung
- Private Cloud
 - Infrastruktur innerhalb der eigenen Organisation
- Hybrid Cloud
 - Eine Kombination aus Public und Private Cloud
- Community Cloud
 - Wie bei Public Cloud, allerdings kleinerer Nutzerkreis
 - Beispiel: Sciebo (Campuscloud)

Charakteristika [2]

- 1 *“On-demand self-service”*
 - Benutzer können automatisiert Ressourcen anfordern
 - Keine menschliche Interaktion notwendig
- 2 *“Broad network access”*
 - Verfügbarkeit über das Netzwerk
 - Zugang über Standardmechanismen und unterschiedliche Plattformen
- 3 *“Resource pooling”*
 - Ressourcen befinden sich in einem Pool und können von mehreren Benutzern in Anspruch genommen werden
 - Dynamische Zuteilung nach Bedarf

Charakteristika... [2]

- 4 “*Rapid elasticity*”
 - Ressourcen können nach Bedarf dynamisch skaliert werden
 - Verfügbare Ressourcen erscheinen unlimitiert
- 5 “*Measured service*”
 - Ressourcen werden automatisiert kontrolliert und optimiert
 - Benutzung kann überwacht und gemeldet werden

Daten

- Daten sind kein so großes Problem wie bei Grid
 - Datentransfer über große Entfernungen problematisch
- Berechnung und Daten oft beim selben Anbieter
 - Keine Migration notwendig
 - Üblicherweise gute Anbindung
 - Teilweise mit garantiertem Durchsatz
- Häufig kein normales Dateisystem
 - Stattdessen Objektspeicher
 - Zugriff oft über HTTP

Daten...

- Amazon Simple Storage Service (S3) sehr beliebt
 - Teil der Amazon Web Services (AWS)
 - reddit, Dropbox, Minecraft, Tumblr, ...
- S3-Schnittstelle ist ein häufig verwendeter Standard
 - Google Cloud Storage
 - OpenStack Swift
 - Ceph mit RADOS-Gateway

HPC

- Inzwischen auch Cloud-HPC
 - Früher Fokus auf Komfort
- Amazon Elastic Compute Cloud (EC2)
 - C4-Instanzen für das Hochleistungsrechnen
 - Intel Xeon (Haswell) mit Zugriff auf Intel AVX/AVX2, Intel Turbo Boost und Enhanced Networking
 - Optimierte Anbindung an Elastic Block Storage (EBS)
 - Unterstützung für das Erstellen von Clustern

HPC...

- Überlegung: Was kostet ein Supercomputer in der Cloud?
 - Beispiel: DKRZ, Mistral
 - Ca. 3.000 Knoten mit jeweils 24 Kernen
- Zwischen c4.4xlarge und c4.8xlarge
 - $\approx 0,75$ \$/h bei Laufzeit von 3 Jahren und Vorauszahlung
 - Entspricht 2.250 \$/h, 54.000 \$/d und 19.710.000 \$/a
 - 98.550.000 \$ bei einer Laufzeit von 5 Jahren ($\approx 88.000.000$ €)
 - 197.100.000 \$ ($\approx 176.000.000$ €) bei On-Demand-Instanzen
- Vergleich: Kosten für Mistral
 - 40.000.000 € Anschaffung
 - 2.000.000 €/a Betrieb
 - 50.000.000 € bei einer Laufzeit von 5 Jahren

HPC...

- Dafür nur knapp die Hälfte an Arbeitsspeicher
 - 30 GiB bzw. 60 GiB pro Instanz
 - Mistral insgesamt 240 TB
- Außerdem noch keine Speicherkosten enthalten
 - Mistral: Lustre-Dateisystem mit 50 PB
- Speicher über Elastic Block Storage
 - 0,054 \$/GB pro Monat
 - 2.700.000 \$ pro Monat, 32.400.000 \$/a
 - 162.000.000 \$ für 5 Jahre (\approx 145.000.000 €)

Ceph

- Ceph ist eine Speicherplattform
 - Bietet Datei-, Objekt- und Blockspeicher
 - Kein Single Point of Failure
 - Skalierbar bis in den Exabyte-Bereich
 - Fehlertoleranz durch Replikation
- Kein Cloud-System, wird aber häufig als Basis verwendet
 - S3-kompatible Schnittstelle

Ceph...

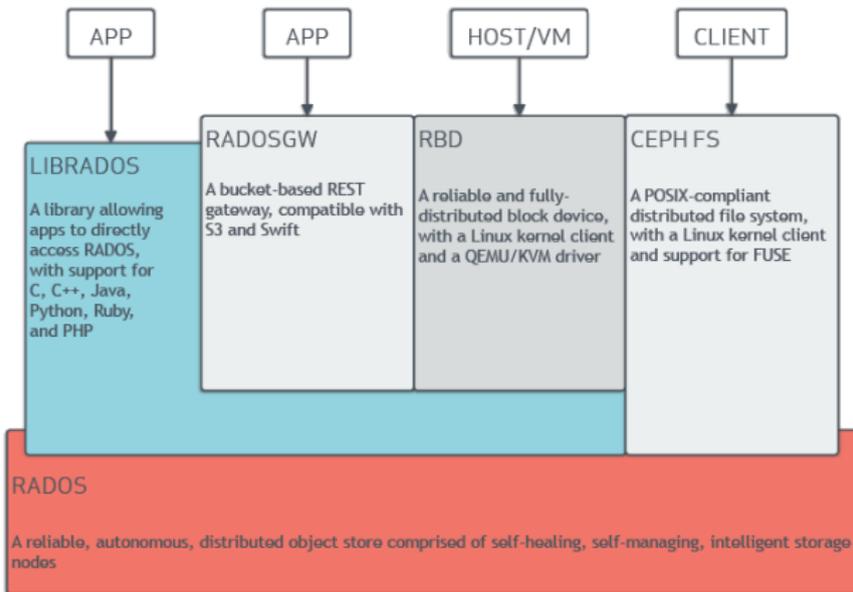


Abbildung: Ceph-Stack [6]

Ceph...

- Basis ist ein verteilter Object Store
 - Komplette Verwaltung wird von RADOS übernommen
 - Darauf aufbauend zusätzliche Funktionalitäten
 - Oder direkter Zugriff auf den Object Store
- Verteilte Blockgeräte
 - Kann für lokale Dateisysteme genutzt werden
- POSIX-Dateisystem
 - CephFS stellt Dateisystemfunktionalität bereit



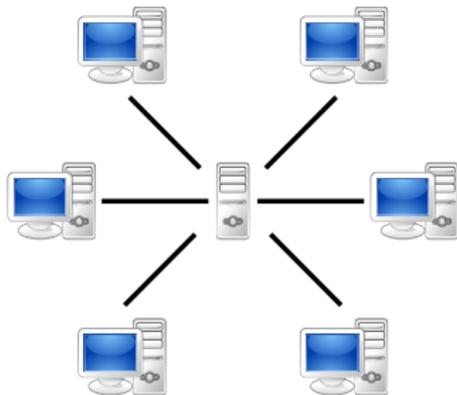
Ceph...

- Typische Komponenten
 - Object Storage Daemon für Daten
 - Metadata Daemon für Metadaten
- Daten werden in lokalem Dateisystem gespeichert
 - Früher EBOFS, wird nicht mehr unterstützt
 - Aktuell btrfs, dadurch POSIX auf zwei Schichten

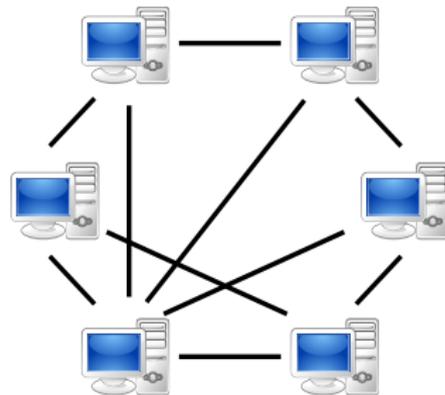
Einleitung

- Peer to Peer (P2P) bekannt aus Tauschbörsen
 - Peers sind Gleichgestellte im Netzwerk
 - Im Gegensatz zu Client-Server-Systemen
- Teilnehmer können Dienste anbieten und in Anspruch nehmen
 - Üblicherweise aber Zuweisung bestimmter Dienste
 - Häufig beschränkt auf Datenaustausch
- Teilnehmer kommunizieren direkt untereinander
 - Keine zentrale Instanz, die Flaschenhals sein könnte

Einleitung...



(a) Client-Server-System [7]



(b) Peer-to-Peer-System [7]

Probleme

- Teilnehmer sind sehr heterogen
 - Unterschiedliche Rechenleistung, Durchsatz, Latenz etc.
 - Teilnehmer können dem Netzwerk beliebig beitreten
- Verfügbarkeit der Teilnehmer ist nicht garantiert
 - Redundanz zwingend notwendig
 - Beitreten und Verlassen wird als *Churn* bezeichnet

Dezentralisierung

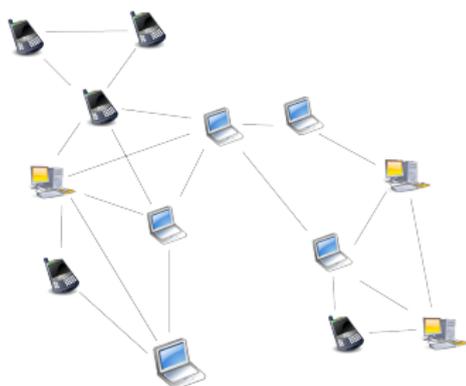
- Keine zentrale Datenbasis
 - Jeder Teilnehmer stellt Daten und Speicherplatz bereit
 - Teilnehmer kennen nicht gesamten Datenbestand
- Keine zentrale Kontrollinstanz
 - Manchmal aber Vermittler für bessere Leistung
 - Beispiel: BitTorrent-Tracker
- Unterschiedliche Grade der Dezentralisierung

Dezentralisierung...

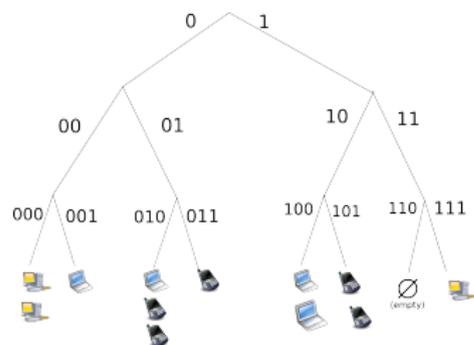
- Vollständig zentralisiert
 - Benötigt zentrale Instanzen um zu funktionieren
 - Üblicherweise Verwaltung der Peers und Daten
- Hybride Zentralisierung
 - Einige Peers nehmen Sonderaufgaben wahr
 - Sogenannte Supernodes
- Vollständig dezentralisiert
 - Alle Peers nehmen dieselben Aufgaben wahr
 - Theoretisch beliebig skalierbar
 - Hohe Fehlertoleranz

Strukturierung

- Peers bilden ein sogenanntes **Overlay-Netzwerk**
 - Beschreibt die Verbindungen zwischen den Teilnehmern
 - Üblicherweise unabhängig vom physikalischen Netzwerk
- Unterschiedliche Grade der Strukturierung



(c) Unstrukturiertes
Overlay-Netzwerk [7]



(d) Strukturiertes
Overlay-Netzwerk [7]



Strukturierung...

- Unstrukturiert
 - Einfach zu realisieren
 - Peers verbinden sich zufällig miteinander
 - Alle Peers sind gleich, kein Problem bei hohem Churn
 - Schwierig bezüglich Suche
- Strukturiert
 - Bestimmte Topologie vorgegeben
 - Meistens über eine verteilte Hashtabelle (DHT) realisiert
 - Hoher Churn problematischer durch ständige Neuorganisation der Struktur

Vergleich

- Ähnliches Modell wie bei Grid und Cloud denkbar
 - Daten sind von überall zugreifbar
 - Daten sind einigermaßen sicher
 - Keine eigene Hardware notwendig
- Bezahle Anbieter für Datenhaltung
 - Früher: Wuala
 - Daten werden verschlüsselt, aufgeteilt und an Peers verteilt
 - Konnte sich nie wirklich durchsetzen

Dateisysteme

- Übliche Anwendung ist File Sharing
 - Daten werden einmal eingestellt und dann geteilt
 - Aufteilung in Blöcke für parallelen Transfer
 - Keine nachträglichen Modifikationen möglich
- Dateisysteme sind aber auch möglich
 - Teilnehmer können Daten ändern
 - Nur Forschungsprototypen verfügbar

Dateisysteme...

- Ivy
 - Unterstützt mehrere Benutzer
 - Unterstützung für Lesen und Schreiben
 - Auf Basis von Logs und DHT
- Shark
 - Daten stammen von zentralem Server
 - Kooperatives Caching verteilt Last
- OceanStore
 - Starke Konsistenz durch Commit-Protokoll
 - Konsistenzanforderungen können gelockert werden

Zusammenfassung

- Grid
 - Bereitstellung von Ressourcen zur entfernten Nutzung
 - Komplexe Benutzung (Zertifikate, virtuelle Organisationen)
- Cloud
 - Bereitstellung von Ressourcen zur entfernten Nutzung
 - Deutlich einfachere Handhabung (webbasiert)
- Peer to Peer
 - Bereitstellung von Informationen (meistens Dateien)
 - Üblicherweise alles öffentlich

Zusammenfassung...

- Grids, Clouds und Peer to Peer haben ähnliche Konzepte
- Berechnung ist einigermaßen einfach zu realisieren
 - Clouds sind teurer als ein eigener Hochleistungsrechner
- Daten sind problematisch
 - Daten zu bewegen ist teurer als sie zu berechnen
 - Müssen zum/vom Ort der Berechnung transportiert werden
 - Einschränkungen bezüglich Kapazität und Durchsatz
- Bei Peer-to-Peer-Systemen unterschiedliche Grade der Dezentralisierung und Strukturierung
 - Meistens mit verteilten Hashtabellen

1 Grid, Cloud und Peer to Peer

- Orientierung
- Einleitung
- Grid
- Cloud
- Peer to Peer
- Zusammenfassung

2 Quellen



Quellen I

- [1] Ian Foster. What is the Grid? A Three Point Checklist. <http://www.mcs.anl.gov/~itf/Articles/WhatIsTheGrid.pdf>, 07 2002.
- [2] Peter M. Mell and Timothy Grance. SP 800-145. The NIST Definition of Cloud Computing. Technical report, Gaithersburg, MD, United States, 2011.
- [3] Athicha Muthitachoen, Robert Morris, Thomer M. Gil, and Benjie Chen. Ivy: A Read/Write Peer-to-peer File System. *SIGOPS Oper. Syst. Rev.*, 36(SI):31–44, December 2002.
- [4] University of Chicago. Globus Toolkit. <http://toolkit.globus.org/toolkit/>.



Quellen II

- [5] University of Chicago. MPICH-G2. http://toolkit.globus.org/grid_software/computation/mpich-g2.php.
- [6] Wikipedia. Ceph (software). [https://en.wikipedia.org/wiki/Ceph_\(software\)](https://en.wikipedia.org/wiki/Ceph_(software)).
- [7] Wikipedia. Peer-to-peer. <https://en.wikipedia.org/wiki/Peer-to-peer>.