

Cloud-Computing Seminar - Vergleichende Technologien: Grid-Computing Hochschule Mannheim

Sven Hartlieb

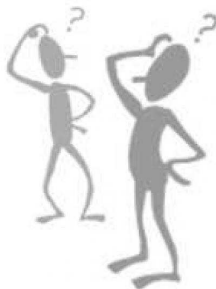
Fakultät für Informatik
Hochschule Mannheim
sven.hartlieb@gmail.com

20.11.2009

Agenda

- 1 Hintergrund
 - Motivation
 - Begriff
 - Geschichte
 - Grid-Projekte
- 2 Grid-Computing
 - Definitionen
 - Abgrenzungen
 - Arten von Grids
 - Merkmale von Grids
- 3 Realisierung von Grid-Systemen
 - Middleware
 - Architektur von Middleware
 - Beispiel: UNICORE

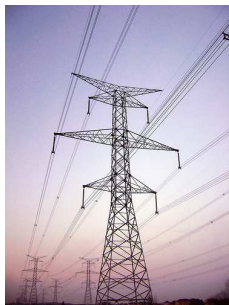
Motivation - Warum Grid-Computing?



Motivation - Warum Grid-Computing?

- Bedarf nach immer größerer Rechenleistung
- Bedarf zum Speichern immer größerer Datenmengen und deren Auswertung
- Bedarf nach standortübergreifenden Projekten mit entsprechender Infrastruktur
- Beispiel: Einer der LHC Detektoren produziert etwa 100 Terrabyte an Daten pro Sekunde
- Suche nach Malaria Wirkstoff mittels des EGEE (Enabling Grids for E-Science) Grids nach zwei Tagen abgeschlossen. Ohne Grid hätte dies ca. 6 Monate gedauert

Der Begriff Grid-Computing



- Englisch: Power Grid = Stromnetz
- Den Benutzer interessiert es nicht, woher der Strom kommt. Er nutzt ihn einfach
- Vision des Grid-Computing: Transparenz bezüglich der Ressourcennutzung

Geschichte

- Mitte der 90er: erste Bestrebungen der Vernetzung verschiedener US-Supercomputer Center. Metacomputing
- 1994: Globus Projekt. Ziel: weltweit vernetzte wissenschaftliche Zusammenarbeit. Erste Version des Globus Toolkits
- Weitere Projekte: FAFNER und I-WAY: legten Technologische Meilensteine. Beeinflussung des Globus Projekts und des seti@home
- 1997: Erstmals Begriff des Grid-Computing
- 1998: wegweisendes Werk: The Grid: Blueprint for a New Computing Infrastructure
- seit 2001: European Data Grid und Nachfolgeprojekt EGEE. Ziel: Schaffung einer nachhaltigen europäischen Grid-Infrastruktur

Das Worldwide LHC Computing Grid

- 2003 ins Leben gerufen
- 170 Organisationen aus 34 Ländern, vorwiegend aus dem universitären Bereich
- Verarbeitet, analysiert und speichert die Daten aus dem LHC
- Daten des LHC werden im CERN gesammelt (Tier 0) und an 11 weitere Rechenzentren verteilt (Tier 1), von hier aus weitere Verteilung an weitere Rechenzentren (Tier 2)

Die D-Grid Initiative

- Gegründet vom Bundesministerium für Bildung und Forschung im Jahr 2004
- Ziel: Förderung von entsprechenden Projekten mit bis zu 100 Millionen Euro
- 2007 wurden etwa 150 Grid Projekte gefördert mit eine Gesamtsumme von ca. 60 Millionen Euro
- 2005 – 2008: Beginn der Entwicklung einer verteilten, integrierten Ressourcenplattform
- 2007-2010: Angebot von IT Services für Wissenschaftler als auch für interessierte aus der Industrie

Definition 1

- Per Definition ist das Grid-Computing eine Technik zur Integration und zur gemeinsamen, institutionsübergreifenden, ortsunabhängigen Nutzung verteilter Ressourcen auf Basis bestehender Kommunikationsinfrastrukturen wie z.B. dem Internet ¹

¹Günther Bengel et al: Masterkurs Parallele und Verteilte Systeme

Definition 2

- A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities ²
- später: Erweiterung der Definition: Zugang zu allen Arten von Ressourcen. Koordinierte Verteilung und Problemlösungsmöglichkeit in dynamischen, multiinstitutionellen virtuellen Organisationen
- Virtuelle Organisation: Organisationen und Gruppen von Menschen, die ein datenintensives Ziel teilen. Zur Erreichung dieser Ziele teilen die Mitglieder ihre Ressourcen in einem Grid

²Ian Foster und Carl Kesselman The Grid: Blueprint for a New Computing Infrastructure

Abgrenzung zu parallelem Rechnen

- Das Hauptaugenmerk beim parallelen Rechnen liegt auf der Aufteilung von Rechenzeit auf mehrere Prozessoren
- Die effiziente Lösung rechenintensiver Aufgaben steht hier also im Vordergrund
- Hierbei gibt es meist eine zentrale Instanz, die die Berechnungen kontrolliert und die Ergebnisse zusammenführt. Dies ist bei Grid-Computing nicht der Fall

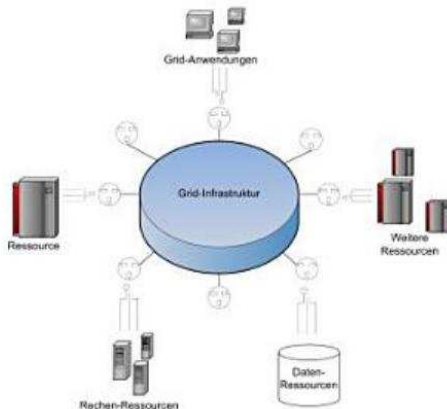
Abgrenzung zu Cluster-Computing

- Die Ressourcen in einem Grid gehören in der Regel mehreren, unabhängigen Organisationen an
- In einem Cluster gehören die Ressourcen für gewöhnlich nur einer organisatorischen Einheit an
- Weitere entscheidende Merkmale, die bei Cluster Computing fehlen, sind die verteilte Kommunikation sowie die offenen Schnittstellen

Abgrenzung zu Cloud-Computing

- Beim Grid-Computing werden weniger, aber aufwändige Anfragen an die Ressourcen stellt
- Beim Cloud-Computing werden dagegen häufige, aber dafür weniger aufwändige Ressourcenanfragen
- Cloud-Computing unterliegt der zentralen Kontrolle des Betreibers, bei Grids liegt keine zentrale Kontrolle vor
- Grid-Middleware muss lokal installiert werden um Ressourcen zu nutzen. In einer Cloud werden alle Ressourcen durch Netzwerkdienste zur Verfügung gestellt

Arten von Grids

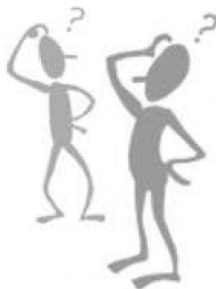


Quelle: Jürgen Dunkel et al.: Systemarchitekturen für verteilte Anwendungen

Arten von Grids

- Information Grid
 - stellt lediglich Daten bereit
 - damit kann das Internet als allseits bekanntes Grid bezeichnet werden, dass durch seine Dienste wie z.B. world wide web Informationen liefert
- Ressource Grid
 - liefert Ressourcen, die von anderen genutzt werden können
 - zum Beispiel Rechenleistung und Speicherkapazität
- Computational Grid
 - Ressourcen mehrerer Rechner zusammengefasst, um die Gesamtrechenleistung zu erhöhen
- Data Grid
 - gemeinsame Nutzung und Verarbeitung großer Datenmengen im Vordergrund
- Service Grid
 - nicht mehr direkten Zugriff auf Informationen und Ressourcen
 - diese werden durch bereitgestellte Dienste (Services) zur Verfügung gestellt, z.B. mittels OGSA oder SOAP

Arten von Grids



Arten von Grids

- ..viele Grid Arten..scheinbar! Denn von Anfang an hatten alle Grids immer mehrere Aufgaben
- ..ja was ist denn nun ein Grid?

Merkmale von Grids

- Ein Grid koordiniert dezentrale Ressourcen
 - Das heisst, dass die Ressourcen von mehreren Benutzern in Anspruch genommen werden können
 - Zudem existiert keine zentrale Einheit zur Koordination, sondern die Abstimmung untereinander wird durch die Prozesse untereinander geregelt
- Grids verwenden offene, standardisierte Protokolle und Schnittstellen
 - Dieser Grund ist deswegen unverzichtbar, da die Koordination von getrennt verwalteten Rechnern nur dann geschehen kann, wenn die hierzu nötigen Schnittstellen bekannt und hinreichend verbreitet sind
- Grids bieten nicht-triviale Dienste an
 - Durch das Angebot mehrerer Dienste bietet das Grid Möglichkeiten, die mehrere einzelne Rechner nicht bieten könnten

Middleware

- Vermittlungs- und Verwaltungssoftware für Grid-Systeme
- Standards für die Kommunikation benötigt, z.B. Protokolle
- Open Grid Forum: Gremium, das Standards und Spezifikationen hierfür liefert
- Standardisierte Architektur: Open Grid Services Architecture
- Bekannteste Middleware Systeme: Globus Toolkit, gLite, Unicore

Architektur von Middleware



Quelle: Jürgen Dunkel et al.: Systemarchitekturen für verteilte Anwendungen

Architektur von Middleware

- Die OGSA spezifiziert hierbei 4 aufeinander aufbauende Schichten:
- Ressourcenschicht
 - Ressourcen, die nach außen hin angeboten werden
- Open Grid Services Infrastructure
 - stellt die Verbindung zwischen den Ressourcen der untersten Schicht her und den den zentralen benötigten Grid Diensten der Schicht darüber
- Die OGSA Platform Services
 - stellen einerseits den Einstiegspunkt für Anwendungen dar, und bieten andererseits zentrale, für das Grid benötigte Dienste an. Hierzu gehören zum Beispiel Optimierung der Ressourcennutzung oder Verwaltung der Grid Ressourcen
- Oberste Schicht
 - besteht dann aus den Anwendungen selbst, die das Grid benutzen möchten

Beispiel: UNICORE

- Seit 1997 entwickelt. Entwickelt am Forschungszentrum Jülich
- Vorteile: leicht benutzbar
- Ist Open Source
- Orientiert sich an OGSA
- Hat grafische Oberfläche
- betriebssystemunabhängig, damit Java realisiert.
- Vorführung: 'Hello world' und Grafikerstellung

Fragen

- Danke für Eure Aufmerksamkeit!
- ..noch Fragen?