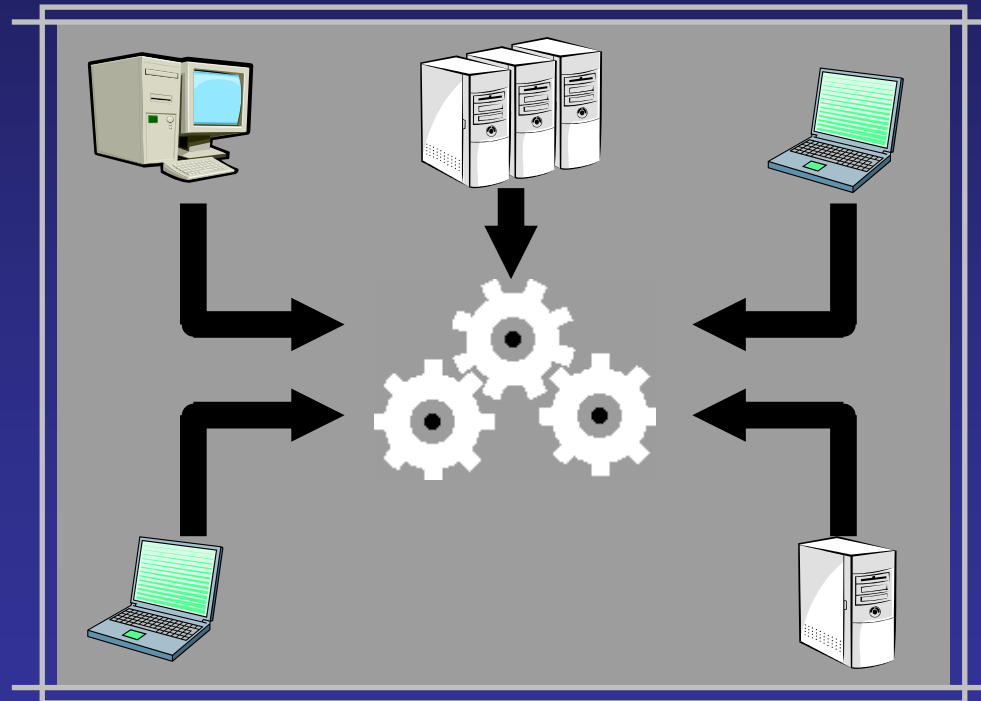


Seminar
Semantic Grid

Idee und Entwicklung des Grid Computing





Inhaltsverzeichnis

1. Was ist ein Grid?
2. Welche Grid Systeme gibt es?
3. Wer nutzt Grids?
4. Grid Architektur
5. Aktueller Stand / Entwicklung



1. Was ist ein Grid?

- existieren viele Millionen von Computern.
- Eine Menge davon sind über das Internet miteinander verbunden.
- Computer teilen Informationen in Form von Webseiten.
- In einem Grid enthaltenen Computer und Instrumente teilen nicht nur Informationen, sondern auch Rechenleistung und Ressourcen wie z.B. Speicherplatz, Datenbanken und Softwareanwendungen.



1. Was ist ein Grid?

- Grid = Gitter, Versorgungsnetz (Elektrizität), Gitternetz (auf Karten)
- „Grid“ in Analogie zu „Stromnetz“ (engl. „electric power grid“)
- Grid ist eine Infrastruktur, die Zugang zu Computerressourcen im high-end-Bereich ermöglicht



1. Was ist ein Grid?

- Unterschied Cluster \Leftrightarrow Grid

| Cluster | Grid |
|--|---|
| <ul style="list-style-type: none">• Verbund von mehreren Rechnern• Kommunikation über LAN• gemeinsamer lokaler Standort• zentraler Admin• Existenz des Clusters wegen Berechnungen• Cluster ist Hauptarbeit der Rechner | <ul style="list-style-type: none">• Verbund von mehreren Ressourcen• Kommunikation über WAN• verteilte Standorte• Jede Maschine eigener Admin• Existenz des Grids aus verschiedenen Gründen• Grid ist Nebenarbeit der Ressourcen |

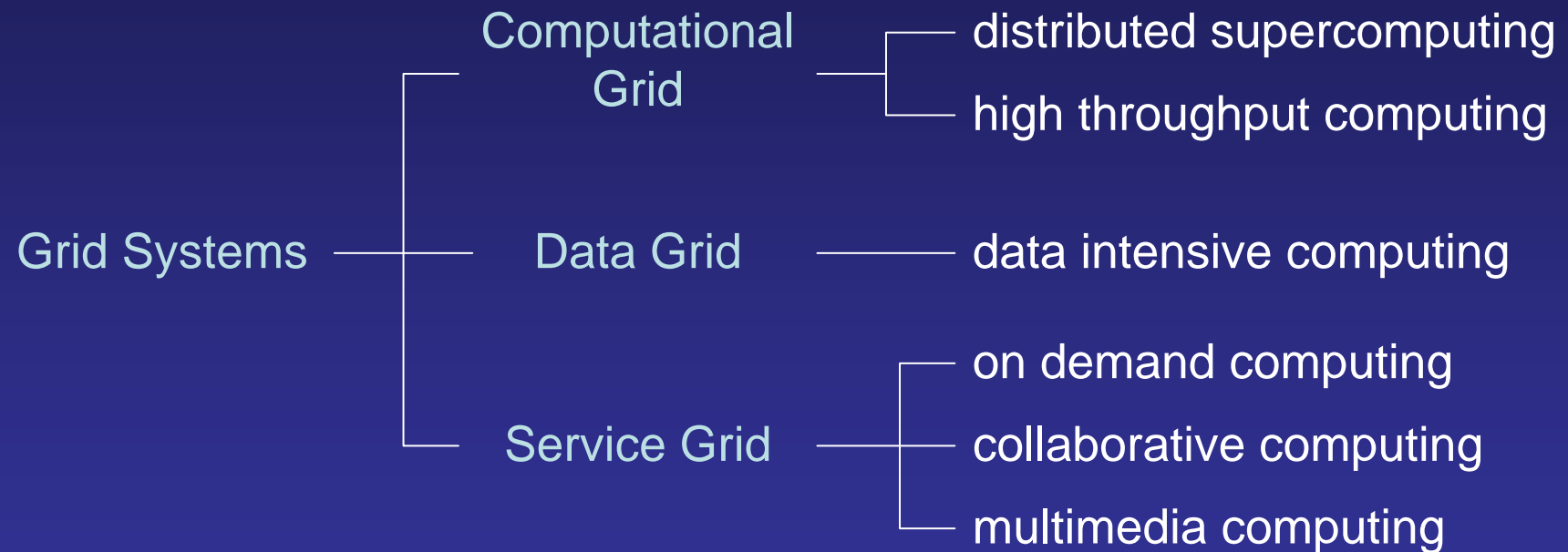


1. Was ist ein Grid?

- Grid Computing
 - Zusammenfassen der Rechenleistung innerhalb eines Netzwerkes
 - Jeder Computer gleichgestellte Einheit
 - geringere Kosten
 - höhere Kapazität und Rechenleistung als Supercomputer
 - gute Skalierung von Grid-Systemen
 - Hinzufügen von Rechnern zum Netz oder Zusammenfassen von Grids zu Über-Grids problemlos
- Entwicklung einer Infrastruktur von der *Globus Alliance*
(www.globus.org)



2. Welche Grid Systeme gibt es?

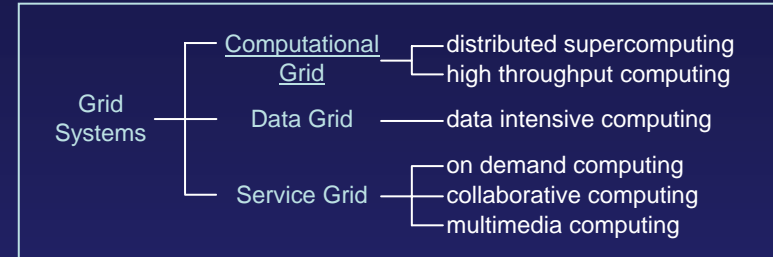




2. Welche Grid Systeme gibt es?

Computational Grid

Grid besitzt im Gesamten mehr Rechenkapazität für eine einzelne Applikation, als die Kapazität jedes einzelnen Rechners beträgt.

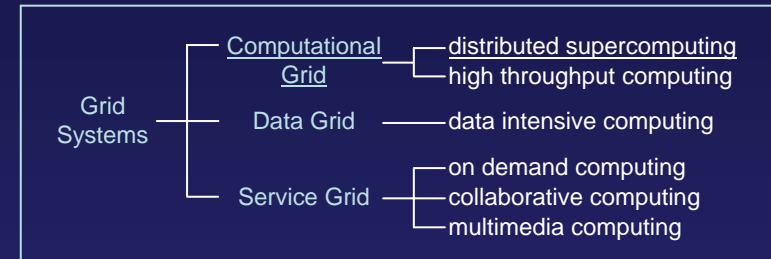




2. Welche Grid Systeme gibt es?

Computational Grid

- Distributed Supercomputing
 - „Verteiltes Supercomputing“
 - Verbund von Supercomputern um Berechnungszeit für Aufgabe zu verkürzen
 - Nutzung bei großem CPU-, Speicherbedarfs
 - Beispiel: Wettervorhersage

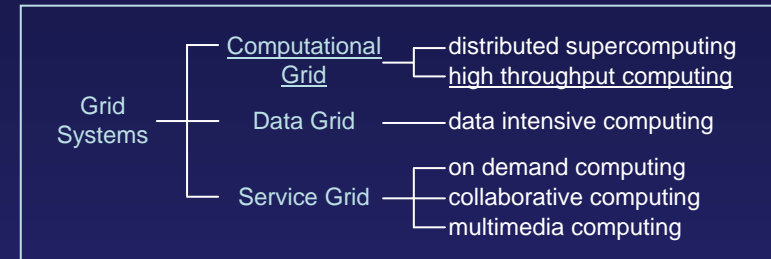




2. Welche Grid Systeme gibt es?

Computational Grid

- High-Throughput Computing
 - „Hoher Datendurchsatz“
 - Zusammenführung ungenutzter Ressourcen in schwach gekoppelten Systemen
 - Nutzung bei Berechnungen, die selbst für Supercomputer nicht in angemessener Zeit zu lösen wären
 - Bedingung: Problem muss in kleine Teilaufgaben unterteilt werden können
 - Beispiel: SETI@Home



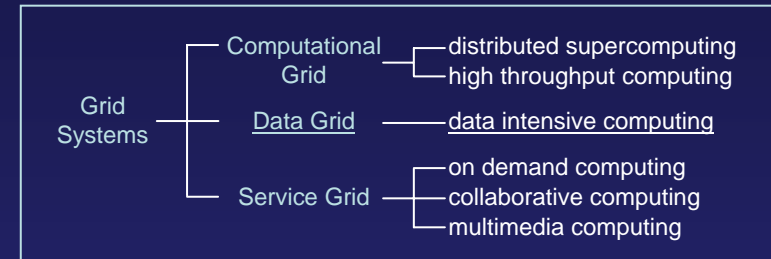


2. Welche Grid Systeme gibt es?

Data Grid

Systeme stellen neue Informationen aus vorhandenen Datenbeständen (z.B. digitale Bücherei) zusammen, die im WAN angeboten werden.

- Data Intensive Computing
 - „Datenintensiv“
 - Synthese von neuer Information aus geographisch verteilten Datenquellen und Datenbanken
 - Oft hohe Ansprüche an Rechenleistung als auch an Kommunikationsleistung





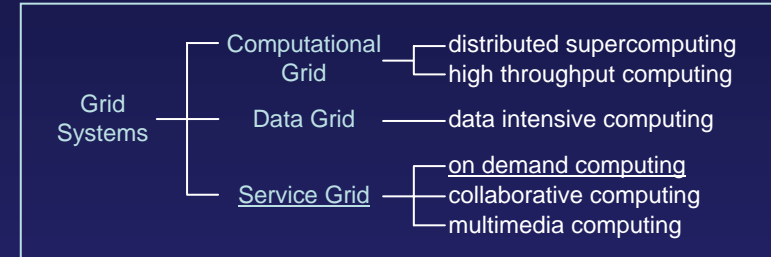
2. Welche Grid Systeme gibt es?

Service Grid

Dienste werden nicht von einem einzelnen Rechner angeboten.

- On-Demand Computing

- „Bei Bedarf“
- Dienstleistung
- Deckung kurzzeitiger Ressourcenbedürfnisse, die lokal nicht kosteneffektiv oder so bequem und schnell gedeckt werden können
- Beispiel
 - Wissenschaftler erhöht die Genauigkeit einer Simulation durch hinzufügen von weiteren Rechnern zur Berechnung der Simulation

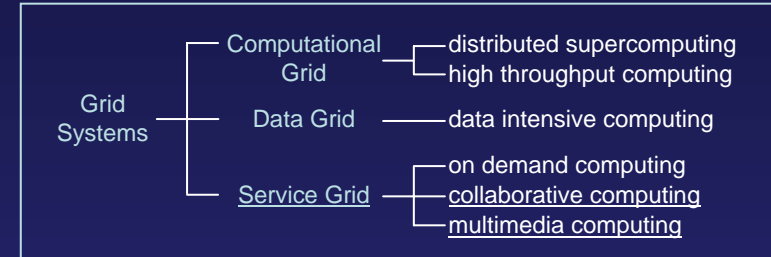




2. Welche Grid Systeme gibt es?

Service Grid

- Collaborative Computing
 - Verbindet Benutzer und Applikationen zu Arbeitsgruppen
 - Echtzeit- Interaktion zwischen Nutzern und Applikationen auf virtuellem Arbeitsfeld
- Multimedia Computing
 - Multimedia Grid stellt Infrastruktur für echtzeit Multimedia- Applikationen zu Verfügung





3. Wer nutzt Grids?

Wer nutzt Grids?

u.a.

- Regierungen
- Gesundheitsorganisationen
- Wissenschaftliche Einrichtungen



3. Wer nutzt Grids?

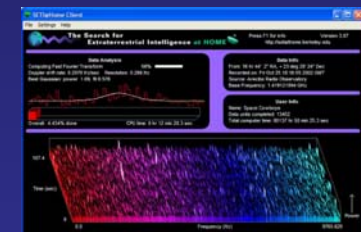
Wer nutzt Grids?

- Grid Computing Projekte:
 - Folding@Home <http://folding.stanford.edu/>
Simulation von Faltungsvorgängen bei Proteinen
 - Seti@Home <http://setiathome.ssl.berkeley.edu/>
Suche nach Außerirdischen
 - Internet Movie Project <http://www.imp.org/>
Modelling und Rendering eines kompletten Films
 - ZetaGrid <http://www.zetagrid.net/>
Verifizierung der Riemann'schen Hypothese (Mathematik)
- (...)

Projektübersicht: <http://www.rechenkraft.net>



Folding@Home



Seti@Home



3. Wer nutzt Grids?

Beispiel: CERN

- In Betriebnahme von Large Hadron Collider in 2007
- Datenaufkommen (10 Petabyte = 1.000.000 GB pro Jahr) nicht mehr von CERN und Partnern zu bewältigen
- $\frac{1}{2}$ aller Teilchenphysiker der Welt arbeiten regelmäßig an CERN-Rechnern



4. Grid Architektur

- Management System
- Donor Software
- Submission Software
- Distributed grid management
- Schedulers
- Communications
- Observation, management, measurement



4. Grid Architektur

- Management System
- Donor Software
- Submission Software
- Distributed grid management
- Schedulers
- Communications
- Observation, management, measurement

- Management System
 - Überwachen der für das Grid verfügbaren Ressourcen
 - Überwachen der Mitglieder
 - ⇒ Entscheidung über zugewiesene Grid Jobs

 - Messen der Verfügbaren Leistung und der Auslastung jedes nodes
 - ⇒ Jobplanung im Grid, Feststellung über Grid-Zustand, Statistik

 - „automatic computing“ / „recovery oriented computing“
 - ⇒ Fehlerbehebung im Grid



4. Grid Architektur

- Management System
- Donor Software
- Submission Software
- Distributed grid management
- Schedulers
- Communications
- Observation, management, measurement

- Donor Software
 - Jeder Rechner, der Ressourcen zur Verfügung stellt muss als Mitglied des Grids eingetragen sein
 - Autorisierung notwendig
- Submission Software
 - Verwendung jedes Rechners in Grid, um Aufträge zu senden
 - In einigen Systemen: „submission nodes“ / „submission clients“
 - separate Software auf Rechnern installiert



4. Grid Architektur

- Management System
- Donor Software
- Submission Software
- Distributed grid management
- Schedulers
- Communications
- Observation, management, measurement

- Distributed grid management
 - In größeren Grids: Organisation in hierarchischem Aufbau
 - Vorteil: Steigerung der Flexibilität des Grids
 - Einzelne Rechner in „Cluster“ zusammengefasst
 - Diese können wieder in Clustern zusammengefasst sein
 - Aufgaben werden von einem zentralen job scheduler an die einzelnen Cluster geschickt. Von dort aus bis zu den Rechnern weitergegeben.



4. Grid Architektur

- Management System
- Donor Software
- Submission Software
- Distributed grid management
- Schedulers
- Communications
- Observation, management, measurement

- Schedulers
 - Ermittelt Rechner, der den von einem Nutzer eingesandten Auftrag abarbeiten kann
 - In den meisten Grid System enthalten
 - Einfachster Fall: Aufgaben werden im Round-Robin Verfahren verteilt
 - Einige scheduler beinhalten ein job priority System.
 - Manchmal auch policies enthalten
 - Auch scheduler können hierarchisch organisiert sein.
meta-scheduler → ... → cluster scheduler → ... → Rechner
 - Fortgeschrittene scheduler können Aufgaben überwachen und z.B. bei Verlust eine Aufgabe erneut an einen anderen Rechner schicken



4. Grid Architektur

- Management System
- Donor Software
- Submission Software
- Distributed grid management
- Schedulers
- Communications
- Observation, management, measurement

- Communications
 - Manchmal wird ein Auftrag in separate Unteraufträge aufgeteilt.
 - Wenn diese separaten Unteraufträge Informationen miteinander austauschen müssen, ist es notwendig andere Unteraufträge zu finden, eine Verbindung aufzubauen und die Informationen auszutauschen.
 - Open standard Message Passing Interface (MPI) wird oft in Grid Systemen für diese Kommunikation verwendet.



4. Grid Architektur

- Management System
- Donor Software
- Submission Software
- Distributed grid management
- Schedulers
- Communications
- Observation, management, measurement

- Observation, management, measurement
 - In donor Software ist normalerweise „load sensor“ enthalten
 - „load sensor“ misst Auslastung und Aktivität der einzelnen Rechner
 - Einige Grid Systeme stellen Möglichkeiten zur Verfügung eigene „load sensors“ zu implementieren, um auch andere Ressourcen zu messen.
 - Informationen sinnvoll für scheduling, aber auch für Analyse des Nutzungsmusters des Grids
 - Informationen können für Kontozwecke genutzt werden, als Basis für Ressourcenhandel.



5. Aktueller Stand / Entwicklung

Aktueller Stand

- Unterstützung von Grid Computing
 - Forschung (CERN, NASA)
 - IT Industrie (SUN, Microsoft, IBM, HP)
 - Regierungen (EU)
- Technisch-wissenschaftliche Anwendungen werden „gridfähig“ gemacht
 - z.B. Oracle



5. Aktueller Stand / Entwicklung

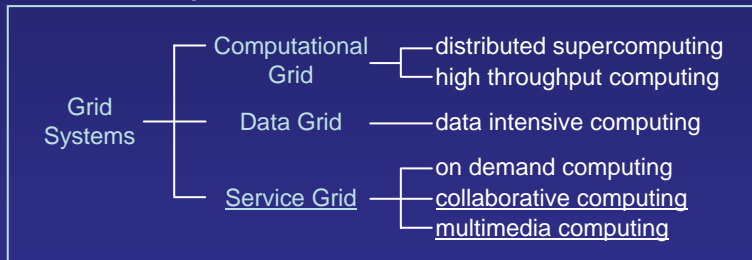
Entwicklung

- Kurzfristig
 - Unterstützung im Bereich Grid Computing wird zunehmen
- Mittelfristig
 - Integration von Grid Technologie in OS
- Langfristig
 - WWG (World Wide Grid)



Zusammenfassung

- Grid
 - Eine Infrastruktur, die Zugang zu Computerressourcen im high-end-Bereich ermöglicht
- Grid Systeme nach ihrem Aufbau und Aufgabe einteilen



- Wer nutzt Grid Systeme?
 - Regierung, Wissenschaftler, Organisationen, Seti@Home, CERN
- Grid Architektur
- Zukunft



Quellenverzeichnis

- I. Foster, C. Kesselman. *Computational Grids*. In: *The Grid: Blueprint for a New Computing Infrastructure*, Morgan-Kaufman, 1999. <http://www.globus.org/research/papers/chapter2.pdf>
- D. de Roure, M. A. Baker, N. R. Jennings, N. R. Shadbolt. *The evolution of the grid*. In: F. Berman, G. Fox, and T. Hey. *Grid Computing: Making the Global Infrastructure a Reality*. John Wiley and Sons Ltd., 2003. <http://www.ecs.soton.ac.uk/%7Enrj/download-files/evolution-of-grid.pdf>
- I. Foster, C. Kesselman, S. Tuecke. *The Anatomy of the Grid: Enabling Scalable Virtual Organizations*. *International J. Supercomputer Applications*, 15(3), 2001. <http://www.globus.org/research/papers/anatomy.pdf>
- K.Krauter, R.Buyya, M.Maheswaran. *A taxonomy and survey of grid resource management systems for distributed computing*. John Wiley and Sons Ltd., 2001.
- V.Berstis. *Fundamentals of Grid Computing*. IBM Corp. 2002. <http://www.ibm.com/redbooks>
- <http://de.wikipedia.org/>

