Grid Computing

Luca Debiasi
0720045
ldebiasi@cosy.sbg.ac.at
Universität Salzburg

Siegmar Alber
0720046
salber@cosy.sbg.ac.at
Universität Salzburg

Abstract

In der heutigen Zeit gibt es Projekte, die sehr große und umfangreiche Datenmengen erzeugen, die natürlich auch ausgewertet und verarbeitet werden müssen. Dafür werden oft sogenannte „Supercomputer“ verwendet, die sehr teuer sind und Unmengen von Kosten für die Installation und Wartung verschlingen. Außerdem sind solche „Supercomputer“ oft nur auf eine Art von Berechnungen abgestimmt und ihre Rechenkapazität ist nur schlecht erweiterbar bzw. ausbaufähig. Hier greift das Konzept der GRIDs: Ihr Konzept besteht darin heterogene Ressourcen miteinander zu verbinden und die Summe dieser gemeinsam verfügbar zu machen. Somit ist es z.B. auch leicht möglich die Rechenleistung eines GRIDs auszubauen. Doch ein GRID stellt nicht nur Rechenleistung zur Verfügung, sondern ist noch für weitaus mehr Zwecke nutzbar.

1 Einleitung

Es gibt keine einheitliche Definition, was ein GRID genau ist. Laut [2] und [9] kann ein GRID jedoch wie folgt beschreiben:

Ein GRID kann als eine Form des verteilten Rechnens beschrieben werden und setzt sich dabei aus lose gekoppelten, heterogenen Einheiten zusammen. Diese Einheiten bestehen aus transparenten Ressourcen, die vernetzt werden. Hierbei stellen offene, standardisierte Protokolle und Schnittstellen nicht-triviale Dienstgüten (QoS) bereit.

Ressourcen bezeichnen dabei:

- CPU: Liefert CPU-Zeit für jegliche Art von Berechnungen
- Speicher: Kann sowohl Arbeits- als auch Massenspeicher sein
- Datenbank: Speichert Daten in Tabellen samt ihren Zusammenhänge
- Messgeräte: Können z.B. Sensoren sein
- Lizenzen: digitale Rechte eines Users
- Anwendungen: Programme jeglicher Art und Weise
- Web Services: Dienste im WWW

Ein Grid kann z.B. mit einem Stromnetz (eng. power grid) verglichen werden, da ihr Aufbau sehr ähnlich ist. Davon stammt auch der Name „GRID“.

Der Vergleich von GRIDs mit Cluster deckt einige Unterschiede auf: GRIDs besitzen eine losere Kopplung und können aus heterogenen Einheiten (Komponenten) aufgebaut sein. Cluster bestehen hingegen aus homogenen Einheiten (Komponenten), die eng gekoppelt sind. Dies bedeutet, dass Cluster Einheiten verbinden, die sich auf engstmöglichem Raum befinden. GRIDs hingegen unterliegen meist einer weiträumigen geographischen Zerstreuung. Ein weiterer wesentlicher Unterschied zwischen GRIDs und Cluster besteht darin, dass GRIDs nicht nur CPU-Leistung bzw. -Zeit zur Verfügung stellen, sondern gegebenenfalls auch andere Ressourcen (z.B. einen Sensor oder Speicher). Außerdem besitzen sie eine standardisierte Middleware, um unterschiedliche Hardware zu Abstrahieren und für universelle Zwecke zur Verfügung stellen zu können. Cluster hingegen sind homogen aufgebaut und benötigen dadurch keine Middleware, sind dadurch jedoch nicht so flexibel.

Im Vergleich zu SETI@Home & Co. weisen GRIDs im Wesentlichen den folgenden Unterschied auf: GRIDs können verschiedene Anwendungsgebiete (Multi-Purposse) haben, SETI@Home & Co. sind auf bestimmte Anwendungsgebiete ausgelegt und verwenden dadurch im Gegensatz zu GRIDs nur die CPU um Berechnungen durchzuführen.

2 Aufbau

Grids bestehen aus mehreren Komponenten. Nach [2] gehören dazu:

Netzwerk Es bildet die Infrastruktur des GRIDs und verbindet die Einheiten. Es gewährleistet außerdem eine sichere Kommunikation zwischen diesen.

Ressourcen Sie können unterschiedlichster Art sein und werden von (geographisch) unabhängigen Einheiten zur Verfügung gestellt.

Middleware Die Middleware abstrahiert die Heterogenität der Ressourcen und verwaltet diese. Dies wird oft durch Virtualisierung umgesetzt.

Applikation(en) Applikationen bilden die Schnittstelle des GRIDs zum Anwender.

3 Geschichte

Laut [9] und [10] gab es bereits in den 60er Jahren erste Konzepte für die Verteilung rechenintensiver Aufgaben. Der Begriff „GRID“ existiert seit Mitte der 1990er Jahre. Die Projekte FAFNER (Factoring via Network-Enabled Recursion) und I-WAY (Information Wide Area Year) bilden den Anfang des GRID-Computings.

Das FAFNER Projekt wurde 1995 ins Leben gerufen von einem Konsortium aus mehreren Firmen in den USA und behandelte Primfaktorenzerlegung mittels RSA-130. I-WAY entstand ebenso wie FAFNER 1995. Es wurde von der University of Illinois initiiert und hatte das Ziel verschiedene Supercomputer unter Verwendung existierender Netzwerke zu vernetzen um ihre Ressourcen gemeinsam nutzen zu können. Dafür wurde eine Gigabit-Testumgebung erstellt.

4 Vorteile

Durch ein GRID verbundene Ressourcen können wesentliche besser ausgenutzt werden, da sie nicht nur für einen speziellen Zweck genutzt werden können, sondern universell einsetzbar sind und auch parallen von mehreren Anwendern genutzt werden können. Außerdem können GRIDs laufend ausgebaut werden und besitzen eine gute Skalierbarkeit. Sie verbinden geographisch unabhängige Ressourcen und dadurch können diese gegebenenfalls effizienter genutzt werden. GRIDs besitzen außerdem eine große Robustheit gegenüber „single points of failure“. Dies bedeutet dass einzelne Ressourcen ausfallen können und das GRID trotzdem noch lauffähig ist, da es diesen Ausfall kompensieren kann. Durch die Kombination von vielen Ressourcen wird es außerdem erst möglich, sehr aufwändige und ressourcen-intensive Projekte durchzuführen.

5 Nachteile

Ein erster Aspekt, der hier beachtet werden muss, ist die (Daten-)Sicherheit. Im wesentlichen versteht man darunter, ob man dem GRID seine Daten anvertrauen kann. Dadurch stellt sich auch die Frage, wer überhaupt garantieren kann, dass sensible Daten auch dementsprechend behandelt werden? Ein weiteres Problem könnte die Abhängigkeit vom GRID werden. Da es für den Endbenutzer nur mehr Thin Clients geben soll, die selbst nur eine sehr begrenzte Rechenleistung besitzen, könnte es bei einem Ausfall des GRIDs dadurch zu schwerwiegenden Folgen kommen (z.B. könnte die Gesamte Börse lahmgelegt werden, wenn das GRID einen Totalausfall erleidet).

6 Beispiele

6.1 Worldwide LHC Computing-Grid

Das Worldwide LHC Computing-Grid wird von der CERN verwaltet und ist Teil des EGEE (Enabling Grids for E-sciencE) Netzwerkes. Beim WLCG werden hunderte Cluster und große Anzahl von Sensoren vernetzt um Kollisionsberechnungen durchzuführen. Hierzu ist ein GRID notwendig, da ein dementsprechender Rechenaufwand von Nöten ist, um die sehr großen anfallenden Datenmengen auswerten zu können. Die Informationen zu diesem Abschnitt stammen aus [8].

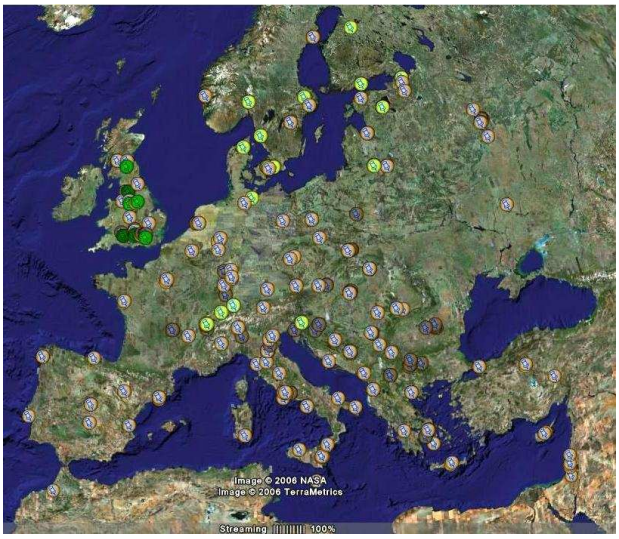

Abbildung 1. Bildquelle [8]. Wichtige Knoten des EGEE

6.2 Frauenhofer Resource Grid

Das Frauenhofer Resource Grid (FhRG) der Fraunhofer-Gesellschaft wird hauptsächlich für (interne) Forschung benutzt, steht jedoch auch für industrielle Nutzung zur Verfügung. Die Informationen zu diesem Abschnitt stammen von [3].

Abbildung 2. Bildquelle [3]. Rechenzentren des FHRG

7 Zukunft und Ziele

Das grundlegende Ziel von GRIDs sollte es sein, „Rechenleistung aus der Steckdose“ für jedermann bereitzustellen (angelehnt an das Stromnetz). Eine weitere Vision ist, dass Endbenutzergeräte auf dem Prinzip der Thin Clients basieren sollen: Hierbei gibt der Anwender eine Berechnung in Auftrag und die Thin Clients dienen nur der Visualisierung der Ergebnisse, die eigentliche Berechnung wird an das GRID abgegeben und von diesem durchgeführt. Dadurch soll es auch möglich werden komplexe Datenauswertungen überall verfügbar zu machen. Ein weiteres Konzept befasst sich mit dem Thema „Grid als Dienstleistung“. Hierbei sollte es in Zukunft gegen Bezahlung möglich sein, auf sehr große Rechenkapazitäten zugreifen zu können und diese für die Industrie oder gegebenenfalls auch für Privatanwender verfügbar und nutzbar zu machen. Durch all diese Konzepte soll ein weltweites Grid entstehen, das abertausende von Ressourcen vereint und gemeinsam nutzbar machen soll.

8 Zusammenfassung

Zusammenfassend kann man also sagen, dass ein Grid folgende 5 Grundideen vereint:

1. Ressourcen teilen (anderen zur Verfügung stellen)
2. Ressourcen gemeinsam nutzen (Ressourcen anderer für eigene Zwecke verwenden)
3. Sicherer Zugang und Datensicherheit
4. Räumliche Distanz verschwindet (durch die Vernetzung der Ressourcen)
5. Offene Standards (Austauschbare und laufend verbesserbare Schnittstellen)

Da die GRID-Technologie noch in den Kinderschuhen steckt kann man noch nicht viel darüber sagen, wie sie sich in Zukunft genau entwickeln wird, jedoch stecken viel Potential und interessante Ideen dahinter. In Zukunft könnte es so weit kommen dass ein weltweit bestehendes GRID, auf das von überall zugegriffen werden kann, für uns so selbstverständlich sein wird wie heutzutage das Internet.

Weitere Referenztypen

Artikel (in Journal)[7], Artikel (in Collection)[5], Buch[6], Proceeding[1], These[4]

Literatur

[1] ACME. *First international Grid conference, Tokyo, Japan*, 2015. Nur Demo-Referenz.

[2] Corentin Chevalier. Gridcafe. Webseite, 2008. `http://www.gridcafe.org/index.html` am 18.12.2008.

[3] Fraunhofer Resource Grid. Fraunhofer resource grid. Webseite, 2008. `http://www.fhrg.fhg.de/` am 18.12.2008.

[4] Alf Heinrich. *Grids als Dienstleistung*. PhD thesis, University of Hamburg, 2000. Nur Demo-Referenz.

[5] Ferdinand Klaus. Was ist ein grid? In *Grundlagen Grids*. Athesia, 2009. Nur Demo-Referenz.

[6] Helene Musterfrau. *Grids*. Athesia, 2nd edition, 2020. Nur Demo-Referenz.

[7] Franz Mustermann. Grids und deren auswirkungen. *Das Grid-Journal*, 5(7):100, 2010. Nur Demo-Referenz.

[8] CERN (Conseil Européen pour la Recherche Nucléaire; Europäische Organisation für Kernforschung). Welcome to the worldwide LHC computing grid (WLCG). Webseite, 2008. `http://lcg.web.cern.ch/` am 18.12.2008.

[9] Wikipedia. Wikipedia. Webseite, 2008. `http://de.wikipedia.org/wiki/Grid-Computing` am 18.12.2008.

[10] Wikipedia. Wikipedia. Webseite, 2008. `http://en.wikipedia.org/wiki/Grid_computing` am 18.12.2008.