

Ein Überblick über Grid Computing Projekte

Oliver Hohlfeld
oliver@ohohlfeld.com

24. Februar 2006

1 Einleitung

Grid Computing ist eine vielversprechende Technologie zur Nutzung verteilter Ressourcen. Eine Ressource kann dabei ein Computer, ein Sensor, Speicherplatz, Programme oder Daten sein [13, S. 1]. Der Zugriff auf diese Ressourcen erfolgt über die Grid Infrastruktur sowie ein Netzwerk wie das Internet. Die Grid Infrastruktur ist als eine Art Middleware Schicht vorstellbar, die den Zugriff auf die Ressourcen verwaltet. Im Gegensatz zu Peer-to-Peer Systemen ist die Verteilung der Ressourcen durch Policies häufig limitiert und die Anwendungen sind oftmals datenintensiver [13, S. 3]. Des Weiteren sind Grid Ressourcen oft leistungsfähiger und besser vernetzt, als typische P2P-Ressourcen. Durch explizite Administration lässt sich im Weiteren auch ein gewünschter Grad an *Quality of Service* erreichen [13]. Die Ressourcen werden dabei in einer festen Gemeinschaft von Anwendern verteilt. Diese wird Virtuelle Organisation genannt. Das Grid Problem meint dabei den gemeinsamen Zugriff auf Ressourcen sowie das Lösen von Problemen in multi-institutionellen Virtuellen Organisationen [12, S. 2].

Dieser Aufsatz gibt einen Überblick über drei Projekte im Rahmen des *Grid Computing*. Anschließend werden noch zwei weitere Projekte kurz besprochen. Da momentan eine Vielzahl von Grid Projekten existiert, kann nur eine sehr geringe Auswahl im Rahmen dieses Aufsatzes vorgestellt werden. Ein guter Überblick über weitere Grid Projekte findet sich in [11]. Des Weiteren können die Projekte auch nicht in ihrer vollen Breite präsentiert werden, da dies den Rahmen dieser Arbeit sprengen würde.

2 Eurogrid

Im Folgenden Abschnitt wird das EUROGRID Projekt vorgestellt, welches ein europäisches Forschungsprojekt zur Erprobung der GRID Technologie darstellt. Das Projekt wurde in den Jahren 2000 bis 2004 von der Europäischen Kommission gefördert und war Teil des *Information Society Technologies Programme (IST)* [17]. Mittels EUROGRID wurden viele europäische Hochleistungsrechenzentren (High Performance Computing Centers) über ein Grid miteinander verbunden. Durch eine einheitliche Schnittstelle wurde den Partnern ein einfacherer Zugriff auf deren Ressourcen (Daten, Rechenleistung etc.) gewährt. Die Möglichkeiten, die sich durch diese neuartige Nutzung der Rechenzentren ergab, wurden an vier Teilprojekten demonstriert. Bei diesen handelt es sich um Meteo-GRID, Bio-GRID, CAE-GRID sowie HPC Research GRID [19]. Die ersten drei Projekte werden in den nächsten Abschnitten näher vorgestellt, wobei der Fokus hier auf dem Meteo-GRID Projekt liegt.

Neben dem Aufbau des Netzes gibt [17] noch weitere Ziele an, die während der Erprobungszeit erfüllt werden konnten. Zum einen soll EUROGRID die bestehende Infrastruktur des Internets verwenden sowie einen nahtlosen und sicheren Zugang zu den angeschlossenen Forschungszentren für dessen Nutzer gewähren. Des Weiteren wurden Softwarekomponenten, u.a. für eine schnelle Datenübertragung erstellt. Weitere Informationen finden sich in [17].

Eine Auswahl der beteiligten Institutionen ist in der nachfolgenden Auflistung gezeigt. Eine vollständige Auflistung findet sich in [18].

- Forschungszentrum Juelich, Germany
- Deutscher Wetterdienst, Germany
- Parallab - University of Bergen, Norway
- CNRS - IDRIS, France
- Warsaw University - ICM, Poland
- Victoria University of Manchester, UK

Die zugrundeliegende Infrastruktur ist das UNICORE (UNiform Interface to COmputing REsources) [4] Projekt. Hierbei handelt es sich um eine Grid Middleware, ähnlich dem Globus Toolkit, mit der auf einfache Art Grid Anwendungen aufgebaut werden können. Es erlaubt einen standardisierten Zugriff auf Rechenzentren. UNICORE wurde im Jahre 1997 ins Leben gerufen und vom deutschen Bundesministerium für Bildung und Forschung (BMBF) unterstützt [10, S. 1]. Die initiale Motivation bestand darin deutsche Hochleistungsrechenzentren auf eine einfache und zugleich sichere Art miteinander zu verbinden. Diese Motivation ähnelt der des EUROGRID Projekts. Vereinfacht könnte man daher EUROGRID als eine Erweiterung von UNICORE auf Europa ansehen.

2.1 Meteo-GRID

Eine Anwendung des EUROGRID Projektes ist Meteo-GRID, das im Folgenden Abschnitt näher beschrieben wird. Die Ausführungen beziehen sich dabei auf [15].

Kurzfristige Wettervorhersagen (bis zu 48 Stunden) werden für viele Institutionen, beispielsweise in der Industrie, immer wichtiger. Berechnungen zur Wettervorhersage werden von vielen Wetterdiensten oft mehrmals täglich erstellt. Die Vorhersage beruht dabei auf numerischen Modellen, wie dem Lokal-Modell (LM)¹ des Deutschen Wetterdienstes (DWD). Da die Erstellung von Wettervorhersagen sehr zeitkritisch ist, kommen hierfür Grossrechner zum Einsatz.

Die Hauptidee hinter diesem Projekt ist die Bereitstellung eines Dienstes (Application Service Provider), der es angeschlossenen Nutzern erlaubt beliebige Wettervorhersagen auf Abruf zu erstellen. Dabei sind viele Parameter, wie die Vorhersagedauern oder die Auswahl des Zielgebietes, durch den Nutzer über eine graphische Benutzerschnittstelle (GUI) einstellbar. Die Berechnung selber findet auf einem am Grid angeschlossenen Rechenzentrum statt. Es handelt sich hier also im Wesentlichen um ein *Computational Grid*; die Nutzer nutzen die Rechenkapazitäten der angeschlossenen

¹Das Lokal-Modell (LM) des Deutschen Wetterdienst gehört zum Kern des dortigen numerischen Wettervorhersagesystems. Es überzieht ganz Europa mit einem engmaschigen Gitter, welches deutlich kleiner ist, als das des Global-Modells (GME). Hierdurch wird die nähere Betrachtung lokaler Einflussfaktoren ermöglicht. Im LM können sehr kleine Gittergrößen verwendet werden, welche jedoch durch die vorhandene Rechenleistung limitiert sind. (Vgl. hierzu [22])

Rechenzentren. Die Ergebnisse werden nach der Berechnung zur weiteren Aufbereitung (z.B. Visualisierung) an den Nutzer übertragen.

Als Projektbeteiligte werden in [20] die folgenden Rechenzentren genannt: CNRS-IDRIS (Frankreich), University of Manchester (UK), das Forschungszentrum Jülich und der DWD. Die Software zur Berechnung des Lokal-Modells wurde auf den Rechenzentren dieser Partner installiert.

Abschließend soll nun der konkrete Ablauf einer solchen Wettervorhersage beschrieben werden. Zunächst werden aus dem Geografisches Informationssystem (GIS) des DWD topographische Daten für das jeweilige Zielgebiet extrahiert. Der Datenbestand des GIS umfasst dabei ca. 7 GiB. Je nach den Einstellungen, die der Nutzer zu Beginn getroffen hat, umfassen die extrahierten Daten nun nur noch 1-5 MiB. Diese werden, zusammen mit weiteren Modelldaten (ca. 50 MiB), in die Berechnungsroutine gesteckt. Die Ausgabedaten einer 48-Stunden Vorhersage besitzen eine Größe von 1-20 GiB.

Die Kontrolle über die Verarbeitung erhält man über die UNICORE Software. Anfragen an ein UNICORE System werden als *Abstract Job Objects* (AJO) [10, S. 9] modelliert. Ein AJO dient zur Kommunikation mit dem Server und spezifiziert den nötigen Arbeitsablauf in einer rechnerunabhängigen Beschreibung. Zur Abbildung des Lokale-Modells auf UNICORE wurde ein LM-Plugin [7, S. 1] in die Architektur eingefügt. Hierbei handelt es sich um eine abstrakte Beschreibung des Prozesses zur Berechnung der numerischen Modelle zur Wettervorhersage, die in EUROGRID ausgeführt werden kann.

2.2 Bio-GRID

Bio-GRID [2] bietet Anwendungen für Biologen und Chemiker auf Basis der EUROGRID Infrastruktur. Das Projekt wird von der Universität Warschau geleitet. Die Hauptaufgabe besteht hierbei aus der Integration bestehender Anwendungen in EUROGRID. Die Anwender sollen somit einen einfachen Zugang zu den Rechenzentren erhalten.

Die integrierte Software stammt dabei im Wesentlichen aus den Bereichen der Quantenchemie sowie der Molekulardynamik. Viele Projekte befassen sich mit der Simulation, beispielsweise von bestimmten Eigenschaften von Molekülen. Des Weiteren werden auch Dienste angeboten, wie der Suche nach ähnlichen Fragmenten in Proteinketten.

2.3 CAE-GRID

Das CAE-GRID [16] unterstützt Ingenieure bei der Arbeit mit *Computer Aided Engineering* (CAE). Hierbei werden erstellte Prototypen mit numerischen Verfahren untersucht. Insbesondere die Simulation bestimmter Eigenschaften bzw. die Optimierung der Prototypen in Bezug auf bestimmte Fähigkeiten spielen hierbei eine Rolle. Die Lösung dieser Aufgaben ist sehr rechenintensiv. Im Rahmen des CAE-GRIDs werden zwei Anwendungen an die EUROGRID Infrastruktur angepasst. Die zwei führenden Institutionen in diesem Projekt sind T-Systems sowie das EADS Corporate Research Centre.

Zum Einen geht es um die *multi-physics optimisation*. Hierbei werden mehrere physikalische Modelle, beispielsweise die Kinetik und die Strömungsmechanik, betrachtet. Diese Verfahren können die notwendigen Prototypen sowie die Testkosten für einen Hersteller reduzieren [16].

Zum Anderen soll die Komplexität des Grids für den Ingenieur möglichst transparent sein. Über

eine einfache Benutzerschnittstelle erhält er die Kontrolle über die Ausführung seiner Berechnungen und kann diese ggf. auch wieder abbrechen.

3 AstroGrid

Das AstroGrid Projekt [1] stellt einen ersten Prototypen für ein Virtuelles Observatorium (VO) dar. Bei einem VO handelt es sich um eine Sammlung von interoperablen Datenarchiven mit astronomischen Daten. Die Daten stammen dabei aus einer Vielzahl von Institutionen, die über ein gemeinsames Interface miteinander verbunden sind. Auf diesen Daten können anschließend Untersuchungen ausgeführt werden. AstroGrid gehört zu den, im Jahre 2002, drei weltweit bedeutendsten VO-Programme und wird in Großbritannien durch das *e-science* Programm gefördert [14, S. 1]. Bei *e-science* handelt es sich um ein interdisziplinäres Programm zur Verbesserung der IT-Infrastruktur von wissenschaftlichen Einrichtungen in Großbritannien, wobei ein Fokus hierbei auf dem Einsatz von Grid Technologien liegt [14, S. 2]. Zum Astrogrid Konsortium gehören die folgenden Institutionen [14, S. 1]:

- University of Belfast
- University of Cambridge
- University of Edinburgh
- University of Leicester
- College London - MSSL
- University of Manchester - Jodrell Bank Observatory
- Rutherford Appleton Lab (RAL) of Central Laboratory of the Research Councils (CCLRC)

Im internationalen Kontext sind noch die beiden VO-Projekte AVO und US-NVO zu nennen. Bei *Astrophysical Virtual Observatory*² (AVO) handelt es sich um ein europäisches Projekt, das wie AstroGrid die Zielsetzung besitzt, Daten für die Forschung leichter zugänglich zu machen. Beim dritten Projekt handelt es sich um das US *National Virtual Observatory*³ (NVO), welches ebenfalls ein Grid Projekt darstellt. Hierbei sind die Projektziele jedoch nicht der Aufbau eines VO. NVO versucht vielmehr die Durchführbarkeit eines Globalen VOs zu erproben. Diese Projekte sind nunmehr in der *International Virtual Observatory Alliance* (IVOA) zusammengeschlossen.⁴ Die allgemeine Zielsetzung von AVO und AstroGrid ist dabei nicht gleich, doch beide überlappen [14, S. 2]. Bei AstroGrid handelt es sich jedoch eher um einen Prototypen, der erstmal die Technologie erproben soll. AVO hingegen strebt das Ziel einer vollständig implementierten Arbeitsumgebung an [14, S. 2].

3.1 Problemstellung

Der Sternenhimmel übte schon immer eine gewisse Faszination auf die Menschen aus. Man beschäftigte sich schon früh mit der Beobachtung und Dokumentation der Sterne. Doch das Wissen über diese

²Mehr Informationen über AVO finden sich unter <http://www.euro-vo.org/>.

³Siehe <http://www.us-vo.org/>.

⁴Die Seite der IVOA - <http://www.ivoa.net/pub/members/> - zählt noch weitere Mitglieder auf, auf die hier aber nicht weiter eingegangen wird.

Welt wächst ständig und damit auch das hierdurch produzierte Datenvolumen. Als Galileo Galilei mit seinem Teleskop den Himmel erforschte und viele für die damalige Zeit herausragende Entdeckungen machte, reichte noch etwas Papier für die Dokumentation des Gesehenen. Heutige Astronomen hingegen produzieren eine Vielzahl von Daten. Hierzu zählen Aufnahmen des Himmels in diversen Frequenzspektren, wie beispielsweise im Optischen-, Röntgen-, Gamma-, Infrarot oder im Radiobereich [5]. Dabei geht [5] davon aus, dass das Moorsche Gesetz auch für die Astronomie gilt. Ein Teleskop könnte hierbei im Jahre 2012 etwa 7 Petabytes Daten pro Jahr produzieren. Diese gilt es effizient zu speichern und verfügbar zu machen.

Im Allgemeinen handelt es sich bei AstroGrid um ein sogenanntes *Data Grid*. Es geht hierbei nicht um die Bereitstellung bestimmter Rechenkapazitäten, sondern um das Verwalten von möglichst großen Datenbeständen. Die Daten, die von Astronomen in diversen Instituten gesammelt wurden, sind oftmals in vielen lokalen Datenbanken gespeichert. Ein Zugriff auf diese Daten ist nicht allgemein möglich, es gilt hierbei immer die jeweiligen lokalen Gegebenheiten zu beachten. Im Wesentlichen geht es also darum die gesammelten Daten möglichst leicht für alle angeschlossenen Partner verfügbar zu machen. Dies schließt auch effiziente Suchmöglichkeiten und Auswertungstools ein. Durch das Grid erscheinen die verteilten Datenbanken als eine große.

Die im Grid gesammelten Daten können beispielsweise genutzt werden, um Forschung direkt auf diesen zu betreiben. Das Verwenden von Teleskopen ist meist teuer, da u.a. viele Teleskope zeitlich nicht direkt zur Verfügung stehen und meist eine längere Wartezeit erfordern [5]. Bei einige Untersuchungen könnte somit der Gang zum Teleskop sowie eigenständigen Datenerfassung durch Operationen auf den gesammelten Daten erspart werden.

3.2 Zielsetzung

Grundsätzlich soll das Grid die verteilten Datenbestände vereinen und Werkzeuge zur Auswertung bereitstellen. In [14] werden vier Visionen des Projekts angegeben:

- *The archive is the sky* - Wie bereits im vorherigen Abschnitt angedeutet, wird durch das *data grid* ein immenser Datenbestand zur Verfügung gestellt. Dieses virtuelle Abbild des Himmels kann immer öfter anstelle des tatsächlichen Himmels zur Forschung genutzt werden - Das Archiv ersetzt also den Himmel.
- *Everybody can be a power user* - Auswertungen von großen Datenbeständen sollen einfacher und transparenter für den Endanwender werden - sodass jeder zum *Power User* wird, der einfach komplexe Analysen durchführen kann.
- *Shift the results not the data* - Beim Umgang mit solch riesigen Datenbeständen können viele Probleme, z.B. durch ungenügende Bandbreite auftreten. Müsste der Client die Auswertung selber vornehmen, ist es durchaus möglich, dass er hierfür nur ungenügend Speicherplatz zur Verfügung hat. Daher sollten nur die jeweiligen Ergebnisse und keine vollständigen Datensätze verschickt werden.
- *A supercomputer on your desk* - Dienstleistungen des Grids können vom persönlichen Arbeitsplatzrechner (Desktop) in Anspruch genommen werden. Der Job wird einfach an das Grid geschickt. Die Ausführung ist dabei für den Anwender völlig transparent - er weiß nicht wo der Job erledigt wird. Jeder Nutzer verfügt somit über einen Supercomputer auf seinem Schreibtisch.

In [14, S. 6-7] sind die wissenschaftlichen Zielsetzungen des Projektes angegeben. Diese lauten im Einzelnen:

- Optimierung des Datenzugriffs im Bereich der Forschung. Die Daten sollen dabei einfacher und schneller abrufbar sein.
- Neue Daten sollen, egal von welcher Quelle, einfach eingefügt werden können. Des Weiteren sollen diese miteinander vergleichbar sein.
- Barrieren bei der Datenanalyse sollen möglichst entfernt werden
- Daten sollen leicht manipulierbar sein

Des Weiteren gibt [14, S. 7] einige praktische Zielsetzung an. Einige davon sind in der nachfolgenden Liste aufgeführt.

- Einheitliche Standards für Daten- und Metadatenformate sowie zum Datenaustausch
- Software Infrastruktur für Datendienste
- Aufbau eines Grids zur gemeinsamen Nutzung von Ressourcen
- Umsetzung eines Virtuellen Observatoriums, basierend auf wichtigen Datenbanken in Großbritannien
- Bereitstellung einer Benutzerschnittstelle

AstroGrid gehört weiterhin zu den frühen Anwendern der OGSA⁵-DAI (OGSA Database Access Infrastructure) Technologie [6, 14]. Hiermit können verschiedene Datenquellen (z.B. relationale- bzw. XML-Datenbanken) in ein Grid integriert werden.

4 Distributed Aircraft Maintenance Environment (DAME)

Das DAME (Distributed Aircraft Maintenance Environment) Projekt [8] stellt ein Fehlerdiagnostik⁶ und Prognose System für die zivile Luftfahrt dar. Es werden dabei beispielsweise Daten über Sensoren aufgezeichnet und mit Daten für den Idealfall verglichen. Des Weiteren stehen Simulationen zur verbesserten Diagnose zur Verfügung. Nicht immer tritt eine Fehlersituation direkt nach den ersten Anzeichen auf. Durch die Simulation wird versucht festzustellen, warum die Abweichung vom Normalfall entstanden sein könnte und welche Auswirkungen dies haben kann. Diese Teilspekte werden jeweils als *Grid Service* angeboten. In einem ersten Prototypen zur Demonstration der Technologie geht es dabei nur um Flugzeugtriebwerke.

DAME ist ein Teil des bereits angesprochenen *e-Science* Programms. Zu den Projektpartnern zählen u.a.:

- Rolls-Royce
- University of York

⁵Bei OGSA (Open Grid Services Architecture) handelt es sich um eine Webservice orientierte Grid Infrastruktur.

⁶Systeme zur Fehlerdiagnose werden nicht nur in der Luftfahrt eingesetzt. Sie finden vielmehr in vielen verschiedenen Gebieten wie der Medizin oder im Ingenieurwesen anwendung. Neben diesen Beispielen nennt [8] noch einige Eigenschaften solcher Systeme. Diese sind datenlastig; die Beobachtung und Auswertung von Sensordaten sowie Fachkenntnisse in der jeweiligen Domäne sind entscheidend für die Diagnose. Solche Systeme können geschäftskritisch sein und sind oft verteilt (distributed).

- University of Leeds
- University of Sheffield
- University of Oxford

Wobei Rolls-Royce beispielsweise Daten zur Simulation bereit gestellt hat. Die aufgeführten Universitäten stellen hingegen die wissenschaftlichen Partnerinstitute dar.

4.1 Grundlagen und Zielsetzung

Aufbauend auf [8, S. 2] sollen nun einige Grundlagen zum Thema Diagnose- und Überwachungssysteme aufgeführt werden. Die beiden hier aufgeführten Projekte stellen die Grundlage des DAME-Projekts dar.

Ein verbreiteter Ansatz besteht in COMPASS der Firma Rolls-Royce kombiniert mit Prognose Anwendungen. Hierbei handelt es sich um eine Anwendung, die in diskreten Zeitabständen Daten über die zu beobachteten Maschinen abfragt und analysiert. Diese Daten stammen dabei aus Sensoren an den jeweiligen Maschinen. Diese Ist-Daten werden mit den Ideal- (Soll-) Werten verglichen, um Abweichung vom normalen Betriebszustand festzustellen. Bei diesem Ansatz fallen nur kleine Datenmengen an, die bequem während dem Flug übertragen, bzw. am Flughafen ausgelesen werden können.

Um die Effizienz dieses Systems zu verbessern besteht eine Möglichkeit darin mehr Daten zu erheben und diese genauer zu analysieren. Hierfür hat Rolls-Royce in Kooperation mit der Oxford Universität das QUOTE- System entwickelt. QUOTE führt eine kontinuierliche Datenanalyse durch. Die Daten selber sind dabei abgeleitet aus den physikalisch gemessenen Triebwerksschwingungen, kombiniert mit der momentanen Triebwerksleistung. Für die Analyse der Daten werden u.a. Neuronale Netze eingesetzt. Im Gegensatz zum COMPASS produziert QUOTE ein erheblich größeres Datenvolumen. Diese Daten werden von QUOTE jedoch nicht gespeichert.

Die Zielsetzung des DAME Projektes ist es nun beide Ansätze zum kombinieren. Durch eine längere Datenhaltung kann die Analyse verbessert werden. Somit könnten bisher unbekannt Anomalien, die mit bestimmten Ursachen korreliert sind, gefunden werden [8, S. 2].

Die Herausforderung besteht dabei in der effizienten Datenhaltung der während des Fluges gewonnen Daten. Ein Flug kann dabei ca. 1 GB an Daten produzieren. Pro Jahr können somit schnell mehrere TB erreicht werden. Die Daten sind dabei auf viele Quellen weltweit verteilt und müssen innerhalb des DAME Projektes jederzeit abrufbar sein. Des Weiteren besteht in DAME ein Bedarf an effizienten *Pattern Matching* Verfahren, die auf diesen großen Datenbeständen operieren können.

4.2 Aufbau des Grids

Basierend auf [8, S. 3-4] soll nun ein Einblick in die grobe Architektur des DAME Projekts gegeben werden. DAME verwendet für die Grid Infrastruktur das Globus Toolkit in Version 3 sowie OGSi bzw. OGSA.

Der *DAME Engine Data Service* übernimmt, als erste Instanz, den Download der QUOTE Daten aus einem Flugzeug und die Kontrolle zu einer Basisstation. Der Download erfolgt dabei über GridFTP. Die Verbindung zu den Datenbanken erfolgt über die Basisstation. Aus QUOTE gewonnene Daten

werden im *White Rose Grid*⁷ gespeichert. Die Übertragung zu diesem Grid erfolgt ebenfalls über GridFTP. Dieser Grid Service findet sich überall dort, wo ein Flugzeug landen kann.

Zur anschließenden Extraktion der Daten wird der *Data Mining Service* verwendet. Hierbei kommt eine Software-Implementierung des AURA Pattern Matching Systems zum Einsatz. Dieses System erlaubt es große Datenbestände recht zügig zu durchsuchen. Im Kontext des DAME Projekts liefert AURA Datensätze die zu einer Anomalie gehören. Bei ähnlichen Ereignissen kann das System Ratschläge zur Wartung geben. Es bedient sich hierbei einer Datenbank die als Grid Service (über OGSA-DAI) verfügbar ist.

Mit dem *Engine Modelling Service* der Universität Sheffield kann auf den Flugdaten auf den Zustand der Turbine geschlossen werden. Dieser Ansatz stellt eine Erweiterung der in COMPASS verwendeten Methode dar. Das Modell für ein Rolls-Royce Triebwerk ist ebenfalls als Grid Service verfügbar.

Durch den *Case-Based-Reasoning Service* wird eine Unterstützung im Diagnose-Prozess gegeben. Hierfür steht eine Wissensdatenbank zur Verfügung, die bei der Auswahl von Diagnose-Methoden oder anderem Vorgehen Unterstützung bietet. Diese Komponente wurde von der Universität Sheffield entwickelt. Wartungspersonal kann auf diesen Dienst über einen einfachen Webbrowser sowie mittels einer Internetverbindung zugreifen und Informationen zur Diagnose abrufen.

Der letzte Dienst ist der *Data Analysis Service*. Hierbei handelt es sich um ein, von der Universität Leeds, entwickeltes System zur detaillierten Datenanalyse. Betreut durch Experten sollen hierdurch die Ursachen für ein Fehlverhalten gefunden werden.

5 Weitere Grid Projekte

In diesem Abschnitt werden zwei Projekte kurz dargestellt. Die in diesen Projekten behandelten Themengebiete sind einigen der vorgestellten Projekten auf den ersten Blick ähnlich. Bei der Darstellung wird daher auf vorhergehende Projekte Bezug genommen.

5.1 EGSO

Beim *European Grid of Solar Observations* (EGSO) [21] handelt es sich um ein *data grid*. Ebenfalls, wie im AstroGrid Projekt, entsteht bei der Beobachtung der Sonne ein immenses Datenvolumen das verfügbar gemacht werden muss. Wissenschaftlern sollen diese Daten einfacher zugänglich gemacht werden, indem sie über das Grid verfügbar gemacht werden und EGSO Daten finden kann, die für die momentane Arbeit des Wissenschaftlers von Interesse sein können. Wie EUROGRID handelt es sich bei EGSO um ein europäisches Forschungsprojekt, dass durch das *Information Society Technologies Programme* (IST) gefördert wird.

Die grobe Architektur besteht dabei aus drei Rollen. Provider stellen alle Daten zur Verfügung, die in EGSO vorhanden sind. Hierbei werden Metadaten über Beobachtungen bzw. extrahierte Daten, die Daten selber sowie Dienste wie die Vorverarbeitung von Daten oder Wissensdatenbanken angeboten. Der Consumer stellt die Anforderungen des tatsächlichen Nutzers dar. Er kann Daten bzw. Metadaten auswählen, manipulieren und visualisieren. Als Zwischenschicht agiert der Broker.

⁷Bei dem *White Rose Grid* handelt es sich um ein *e-Science* Projekt der Universitäten Leeds, Sheffield und York. Für weitere Informationen sei auf www.wrgrid.org.uk verwiesen. Im Rahmen des DAME Projekts wird dies zur Datenhaltung der Flugdaten eingesetzt.

Er kennt die bereitgestellten Ressourcen sowie die Provider und kann zwischen diesen vermitteln (Resource location). Dies erleichtert die Implementierung des Consumers.

5.2 Earth System Grid

Im Gegensatz zu Meteo-GRID dient das *Earth System Grid* (ESG) [9] dazu das Wetter und dessen Veränderung im Laufe der Zeit zu verstehen. Es ist nicht für Wettervorhersagen ausgelegt. Zur Simulation der Modelle werden verteilte Rechenarchitekturen verwendet. Solche Simulationen produzieren ein Datenvolumen von etwa 100 TB. Erhöht man die Auflösung bzw. die Länge der Simulation, entstehen schnell einige Petabyte an Daten. Für den weiteren Fortschritt der Klimaforschung ist es bedeutend, diese Daten einer möglichst großen Anzahl an Wissenschaftlern zur Verfügung zu stellen. Daten aus den Simulationen sollen hierbei nicht nur leicht zugreifbar sein, neue Daten sollen auch leicht verfügbar gemacht werden können. Zu den Beteiligten gehören u.a. [3]:

- Oak Ridge National Labs (ORNL)
- University of Southern California Information Sciences Institute (USC/ISI)
- Los Alamos National Laboratory (LANL)
- Lawrence Berkeley National Laboratory (LBNL)
- Argonne National Labs (ANL)

ESG nutzt als Grundlage zum Aufbau des Grids das Globus Toolkit. Globus wird hierbei für die Authentifikation, den Zugriff auf Daten und Metadaten sowie die Verwaltung von Jobs genutzt. Zum Datenaustausch kommt u.a. das GridFTP Protokoll sowie OPeNDAP (Open-source Project for a Network Data Access Protocol) zum Einsatz. Der Metadatenkatalog kann über OGSA-DAI abgerufen werden.

6 Zusammenfassung

Die hier vorgestellten Grid Projekte zeigen, dass es sich beim *Grid Computing* nicht nur um die Bereitstellung einer möglichst großen Rechenkapazität handelt. Viele der hier präsentierten Projekte – die *data grids* – müssen mit einer großen Datenmenge umgehen können. Für sie ist die Rechenkapazität nicht elementar. Vielmehr ist die Speicherkapazität der verteilten Ressourcen von Bedeutung. Hierbei sind Datenvolumen von einige TerraByte bereits der Normalfall. Mengen von einigen Petabyte an Daten werden in absehbarer Zeit anfallen. Exabytes und noch größere Datenmengen sind durchaus vorstellbar und werden vermutlich in naher Zukunft vorliegen. Den Projekten ist daher nicht nur die effiziente Datenhaltung wichtig, sondern auch die effiziente Analyse und Suche innerhalb der Daten. Im Gegensatz dazu handelt es sich bei dem EUROGRID Projekt im Wesentlichen um ein *computational grid*, dass für die Berechnung bestimmter Probleme, wie beispielsweise einer Wettervorhersage, ausgelegt ist. Hier spielt die Datenmenge eine untergeordnete Rolle. Vielmehr soll die Rechenkapazität der verbundenen Rechenzentren möglichst gut zur Lösung der jeweiligen Probleme genutzt werden.

Des Weiteren unterscheiden sich die Projekte im Wesentlichen nur in ihren Einsatzgebieten. Um die Infrastruktur für das Grid aufzubauen finden Standard Toolkits wie Globus Verwendung. Darauf aufbauend werden die jeweiligen, für die Projekte spezifischen, Anwendungen aufgesetzt. Jedes Projekt fordert einen eigenen Satz an Anwendungen, die im Grid verfügbar gemacht werden sollen.

Literatur

- [1] Astrogrid web page. <http://www.astrogrid.org>.
- [2] Bio grid web site. <http://biogrid.icm.edu.pl/>.
- [3] Esg website. <http://www.earthsystemgrid.org/>.
- [4] Unicore website. <http://www.unicore.org>.
- [5] What is the nvo? <http://www.us-vo.org/about.cfm>.
- [6] Projects using ogsa-dai. <http://www.ogsadai.org.uk/about/projects.php#astrogrid,2003>.
- [7] Eurogrid lm plug-in - installation and user guide. http://www.cscs.ch/~nnellari/LM_plug-in.pdf, 2004.
- [8] J. Austin, T. Jackson, and M. Jessop. Diagnostics and prognostics on the grid: the distributed aircraft maintenance environment project (dame). <http://www.cs.york.ac.uk/dame/DAMEAHMpaper2003formatted.pdf>, 2004.
- [9] D. Bernholdt, S. Bharathi, D. Brown, K. Chanchio, M. Chen, and A. C. et. al. The earth system grid: Supporting the next generation of climate modeling research. <http://www.globus.org/alliance/publications/papers/esgProcIEEE2005.pdf>.
- [10] D. W. Erwin. Unicore - a grid computing environment. Technical report, Forschungszentrum Jülich GmbH, Zentralinstitut für Mathematik (ZAM) D52425 Jülich, Germany, 2001.
- [11] I. Foster. Selected major grid application and deployment projects in science and engineering. <http://www-fp.mcs.anl.gov/~foster/grid-projects/>.
- [12] I. Foster. What is the grid? a three point checklist. http://www.univa.com/grid/pdfs/WhatIsTheGrid_ID.pdf.
- [13] I. Foster and A. Iamnitchi. On death, taxes, and the convergence of peer-to-peer and grid computing. http://people.cs.uchicago.edu/~anda/papers/foster_grid_vs_p2p.pdf.
- [14] A. Lawrence. Astrogrid : powering the virtual observatory. <http://wiki.astrogrid.org/pub/Astrogrid/ConferencePapers/astrogrid-SPIE.pdf>, 2002.
- [15] C.-J. Lenz and D. Majewski. World-wide local weather forecasts by meteo-grid computing. http://www.eurogrid.org/documents/MeteoGrid_DWD_QR_28.pdf, 2002.
- [16] D. Mallmann. Eurogrid project wp3: Cae grid. <http://www.eurogrid.org/wp3.html>.
- [17] D. Mallmann. Eurogrid project. <http://www.eurogrid.org/index.html>, 2004.
- [18] D. Mallmann. Eurogrid project participants. <http://www.eurogrid.org/participants.html>, 2004.
- [19] D. Mallmann. Eurogrid project workplan. <http://www.eurogrid.org/workplan.html>, 2004.
- [20] D. Mallmann. Eurogrid project wp2: Meteo grid. <http://www.eurogrid.org/wp2.html>, 2004.
- [21] K. Reardon, E. Antonucci, S. Giordano, and G. Severiono. Egso - the european grid of solar observations. <http://www.mssl.ucl.ac.uk/grid/egso/papers/Vulcano.pdf>.
- [22] D. Wetterdient. Lokal-modell (lm). <http://www.dwd.de/de/FundE/Analyse/Modellierung/lm.htm>.