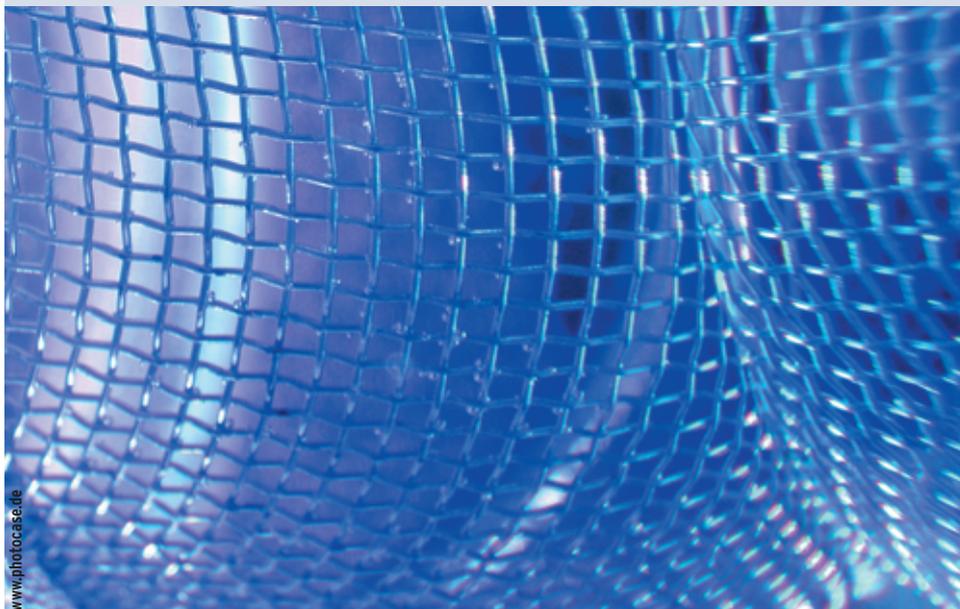


Vernetzte Welten

Die meisten aktiven Grid-Projekte benutzen das Globus-Toolkit in der Version 2. Es enthält ein ganzes Bündel Komponenten und vereinfacht so den Aufbau verteilter Rechenzentren. Harald Kornmayer



Wie in anderen Softwaregattungen erfreuen sich auch beim Grid Computing einige Technologien größerer Beliebtheit als andere. Primus inter pares ist heute das Globus-Toolkit [1]. Auf der Version 2 (GT 2) basieren die meisten namhaften Grid-Projekte, so das European Data Grid (siehe Artikel „Eurovision“) und auch die Wurzeln des LHC Computing Grid reichen bis zu GT 2.

Langfristig ist die Ablösung des GT 2 durch die Versionen 3 und 4 (siehe Artikel „Globus – The Next Generation“) zu erwarten. Dieser Schritt zu Web- und Gridservices wird einen kompletten Paradigmenwechsel bedeuten und daher nicht von heute auf morgen erfolgen.

Have it I-Way

Globus geht auf das Projekt I-Way zurück, das Information Wide Area Year [2]. Der Startschuss fiel im Dezember 1995 auf der Supercomputing '95 in San Diego. In Jahresfrist sollten sich Test-

plattformen an 17 Supercomputer-Zentren, Virtual-Reality-Laboratorien und Entwicklungszentren in den USA miteinander verbinden, um eine Reihe von verteilten Supercomputer-Anwendungen auszuführen. An I-Way arbeiteten auch zwei Veteranen des Grid Computing: Ian Foster vom Argonne National Laboratory und Carl Kesselman von der Universität South California.

Neuer Schub durch Globus

1996 erlaubten es neue Geldmittel, die Softwarekomponenten des I-Way-Testbeds weiterzuentwickeln. Dies gilt als Geburtsstunde des Globus-Toolkits. Anfänglich federführend waren neben Foster und Kesselman noch einige Wissenschaftler der Universität Chicago. Mittlerweile engagieren sich auch die Universität von Edinburgh und das schwedische Center for Parallel Computers. Das Projekt bekommt außerdem Unterstützung aus der Industrie: IBM und Micro-

soft zählen unter anderen zu den Sponsoren, die Anstöße für die weitere Entwicklung geben.

Aufgaben des Toolkits

Das Globus-Toolkit vereinfacht den Aufbau und die Standardisierung von verteilten Rechenzentren. Dafür enthält es ein ganzes Bündel Komponenten. Die drei Werkzeuge Ressourcen-Management, Daten-Management und Informationsdienste sind besonders wichtig, da sie praktisch in jedem Grid zum Einsatz kommen (Abbildung 1).

Alle Komponenten des Globus-Toolkits greifen auf die Grid Security Infrastructure (GSI) zurück, die geprüfte Mechanismen bereitstellt, um weltweit verteilte Computer- und Daten-Grids sicher zu realisieren (siehe Artikel in diesem Schwerpunkt). Mit den Globus-Werkzeugen lässt sich eine Umgebung auf-

Globus-Installation

Das Globus-Toolkit steht für Linux mit Kernel 2.x, Solaris 8, Irix 6.5, Tru 64 und AIX 5.1 zur Verfügung. Die Linux-Version kommt nicht als RPM- oder DEB-Paket, sondern als Tar.gz. Aufgrund der Vielzahl der Komponenten ist die Installation auch nicht mit »configure; make; make install« möglich. Stattdessen kommt ein eigener Installer zum Einsatz, der die benötigten Quellpakete in der richtigen Reihenfolge entpackt und übersetzt. Auf dem Zielsystem muss zudem Perl 5.005 oder neuer vorhanden sein.

Alternativ stehen auf der Projekt-Site auch Binärpakete bereit, mit denen der Installer ebenfalls zurechtkommt. Die Schritte der Installationsvarianten beschreibt die Website [3] genau. Um das Grid zu nutzen, benötigt der Anwender noch ein passendes Zertifikat, worauf der Artikel über die Grid Security Infrastructure näher eingeht.

bauen, in der Anwender ihre Rechenaufträge in einem heterogenen und verteilten Rechnerumfeld gezielt und sicher ausführen, zum Beispiel mit Single-Signon, um die Autorisation und Authentifizierung der Benutzer abzuwickeln. Die Operatoren können zudem die Ressourcen überwachen und die Bewegung großer Datenmengen kontrollieren.

Bewusst nicht Teil von Globus ist ein Resource Broker, der gegebenen Ausführungskriterien einer Grid-Applikation den passenden Rechenknoten zuordnet. Dienstleistungen dieser Art stellt unter anderen das Software-Framework des European Data Grid bereit (siehe Artikel in diesem Schwerpunkt).

Ressourcen-Management

Der zentrale Globus Resource Allocation Manager (GRAM) stellt Dienste bereit, die Rechenaufträge verschicken und sie dabei gleichzeitig überwachen. Wenn ein lokaler Client einen Rechenauftrag an das Grid sendet, nimmt der Gatekeeper auf dem entfernten Rechner diesen an und reicht ihn an einen Jobmanager weiter, der den Auftrag startet und kontrolliert. Die Lebenszeit des Jobmanagers endet stets zusammen mit dem Rechenauftrag selbst.

Der Anwender auf Client-Seite verwendet den Befehl »globusrun«, um einen Job zu verschicken (**Abbildung 2**). Die Steuerungsbefehle des Rechenauftrags

gelangen per HTTP zum Grid. Beschreibungssprache ist die Resource Specification Language (RSL). Mit dieser Sprache legt der Anwender die Anforderungen an die zu verwendende Ressource fest, beispielsweise die erforderliche Größe des Hauptspeichers oder die Art des Betriebssystems.

Der Jobmanager alloziert die Ressourcen und führt den Rechenauftrag aus. Die Übermittlung von Daten in und aus dem entsprechenden Grid-Knoten gewährleistet die Komponente Global Access to Secondary Storage (GASS). Den GASS-Datentransfer kontrollieren der Gatekeeper, der Jobmanager oder der Anwender mit dem Globusrun-Befehl.

Daten-Management und Informationsdienste

Der Anwender greift mit Hilfe des Programms Gridftp auf Daten zu und verteilt sie im Grid. Es fußt auf FTP, erweitert um Funktionen wie Multistream-Transfer, Auto-Tuning und spezielle Globus-Sicherheitsmechanismen. Das Tool erlaubt den Transfer vom Client zum Grid-FTP-Server sowie die Datenübermittlung von einem entfernten Server auf einen anderen entfernten Server, ohne den Umweg über den kontrollierenden Rechner nehmen zu müssen (Third-Party File Transfer).

Der Monitoring and Discovery Service (MDS) bietet Zugriff auf statische und

dynamische Ressourceninformationen. Diese Daten helfen dabei, die Rechenaufträge im Grid besser zu verteilen. Der MDS setzt LDAP (Lightweight Directory Access Protocol) für die verteilte Pflege von Ressourceninformationen ein.

Die wesentlichen Bausteine dieser MDS-Globus-Komponente sind der Information Provider, der Grid Resource Informati-

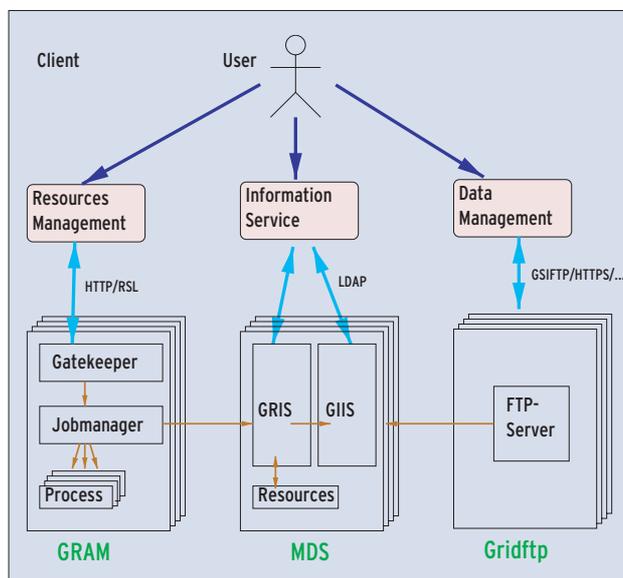


Abbildung 1: Die System-Sicht auf das Globus Toolkit. Die Werkzeuge Ressourcen-Management, Daten-Management und Informationsdienste sind besonders wichtig.

Rechenauftrag verschicken

- 1 Ein Client sendet den Auftrag in Form eines RSL-Dokuments per HTTP an die Server-Seite.
- 2 Der Server erzeugt für den Auftrag einen Jobmanager, der das RSL-Dokument dann entgegennimmt.
- 3 Der Jobmanager parst das RSL-Dokument und prüft die Anforderungen des Rechenauftrags.
- 4 Der Jobmanager verteilt den Rechenauftrag entsprechend den Anforderungen an die lokalen Ressourcen.
- 5 Der Jobmanager sendet einen eindeutigen Job-Identifizier an den Client.
- 6 Der Client erfragt mit Hilfe des eindeutigen Job-Identifizier den Status des Rechenauftrags oder bricht die Ausführung des Auftrags ab.

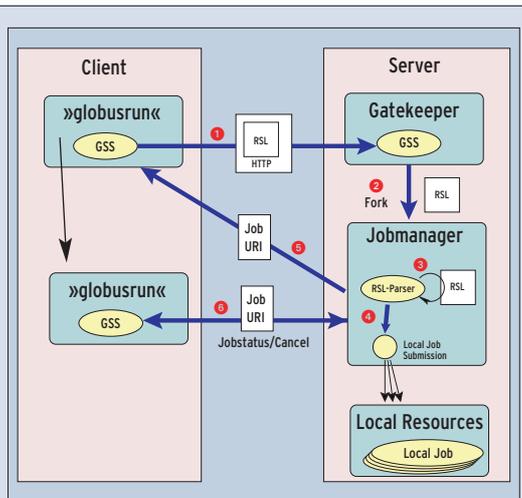


Abbildung 2: Der prinzipielle Ablauf beim Verschicken eines Rechenauftrags. Der Jobmanager erfüllt eine zentrale Aufgabe.

on Service (GRIS), der Grid Index Information Service (GIIS) sowie ein MDS-Client. Der MDS verwaltet neben Informationen über Computerressourcen wie Netzwerkschnittstellen, IP-Adressen oder Speichergröße auch Daten von Infrastrukturkomponenten wie den Namen eines Jobmanagers oder eines Rechenauftrags.

Ressourcen verwalten

Der Information Provider sammelt die Informationen der lokalen Ressourcen und leitet sie an GRIS weiter. Ressourcen in den Informationsdienst eines Grid einbringen ist nur über den Information Provider möglich. Der GRIS verwaltet lokale Ressourceninformationen und aktualisiert sie periodisch oder auf Anfrage. Damit das verteilte Grid auf diese Informationen zugreifen kann, registriert sich der GRIS bei einem GIIS.

Der GIIS fungiert also als Grid-weiter Informationsdienst. Dazu verwendet er einen hierarchischen Mechanismus ähnlich dem DNS (Domain Name System). Der Zugriff auf die Ressourceninformationen erfolgt mit einem MDS-Client, der sich des LDAP-Befehls »ldapsearch« bedient. Für die Suche nach Ressourceninformationen sind sowohl GRIS- als auch GIIS-Dienste verfügbar.

Schon mit diesen drei Globus-Komponenten ist es möglich, ein einfaches Grid aufzubauen. Eine eingebundene Certifi-

cate Authority (CA) gewährleistet die Sicherheit im Grid, für Testzwecke bringt das Globus-Toolkit eine einfache CA mit. Die Installation einer Testumgebung ist in IBMs Red Book „Introduction to Grid Computing with Globus“ [4] gut beschrieben.

Das Globus-Toolkit eignet sich nicht allein für die Installation vorhandener Dienste. Die Anbindung an ein Globus-basiertes Grid kann auch durch APIs geschehen. Solche gibt es für mehrere Programmiersprachen, unter anderem C, Java, Python, Corba und Perl. Mit ihrer Hilfe lassen sich Anwendungen für den Einsatz im Grid anpassen.

Die Gretchenfrage

Linux als wichtige Grid-Plattform hat Globus bei seiner Verbreitung erheblich geholfen. Freie Forschung benötigt Entwicklungsplattformen, die den ohnehin komplexen Projekten keine zusätzlichen Beschränkungen auferlegen. Globus 2 steht unter der Globus Toolkit Public License (GTPL), die allerdings nicht auf der Liste der von der Free Software Foundation als GPL-kompatibel anerkannten Lizenzen steht.

Jedoch erlaubt die GTPL eine fast unbeschränkte Nutzung der Software, einschließlich des Rechts zur Weiterverbreitung und zum Ändern der Software. Wie üblich müssen auch geänderte Pakete wieder unter der GTPL stehen. Es ist

deshalb nach Ansicht des Autors durchaus gerechtfertigt, Globus als Open Source zu bezeichnen.

Zusammenfassung

Das Globus-Toolkit ist als Basistechnologie für Grids verschiedener Topologie konzipiert. Die Einsatzmöglichkeiten erstrecken sich von kleinen lokalen Clustern bis hin zum weltweit verteilten Grid. Ein Beispiel ist das LHC Computing Grid, das ab 2007 Messdaten des Teilchenbeschleunigers LHC am Cern in Genf analysieren soll.

Dem Toolkit fehlen zurzeit noch einige den Komfort ver-

bessernde Dienste. Die Entwickler haben sie bewusst aus dem Globus-Toolkit ausgeklammert und in weiterführende Projekte verlagert, etwa in das European Data Grid. Wichtigste Erweiterung ist der Resource Broker, wie ihn das EDG-Projekt bereitstellt.

Bald werden bestehende Globus-Grid-Projekte den Versionswechsel zu 3 und 4 verkraften müssen. Die damit einhergehenden radikalen Änderungen fußen zwar in handfesten technischen Gründen. Jedoch wird diese Entwicklung sicherlich nicht die als babylonisch anzusehenden Verhältnisse des Grid-Umfelds beseitigen. (jre)

Infos

- [1] Globus Toolkit: [<http://www.globus.org>]
- [2] I-Way-Projekt: [www-unix.mcs.anl.gov/fl/publications/iway-ijsa.pdf]
- [3] GT-2-Software für Download und Anleitung: [<http://www.globus.org/gt2/install/>]
- [4] Aufbau einer einfachen Grid-Umgebung: [<http://www.redbooks.ibm.com>]

Der Autor

Seit 2003 arbeitet Dr. Harald Kornmayer im Rahmen eines europäischen Forschungsprojekts im Grid-Umfeld. Davor entwickelte er in der EDV-Abteilung einer deutschen Großbank ein Marktrisiko-Controlling-System in J2EE mit. Bevor die reine IT sein beruflicher Schwerpunkt wurde, hat er für mehrere Experimente der Astroteilchenphysik Analysensysteme entworfen.