

Globus - The Next Generation

Die Globus-Toolkits 3 und 4 verfolgen einen anderen Ansatz als das weit verbreitete GT 2. Beide bringen vorteilhafte Änderungen, der Umstieg bestehender Projekte wird aber ziemlich schwierig. Rüdiger Berlich, Marcel Kunze



www.photocase.de

Das Globus-Toolkit 2 (GT 2, siehe den Artikel „Vernetzte Welten“ in dieser Ausgabe) ist das Fundament für viele Grid-Projekte und Middleware wie das European Data Grid (siehe Artikel). Das Toolkit stammt aus den Laboren der Globus Alliance [1], einer Forschungseinrichtung, der auch Entwicklungsabteilungen großer Firmen wie IBM zuarbeiten. Die Globus Alliance fühlt sich dem Open-Source-Gedanken verpflichtet und vertreibt Globus unter der GTPL (Globus Toolkit Public License).

Besonders vorteilhaft ist das GT 2, wenn es zusammen mit Middleware wie der des European Data Grid für datenintensive Berechnungen ohne Austausch vieler Informationen zwischen den Rechenknoten eingesetzt wird. Bei vielen Anwendungen erwies es sich jedoch als schwerfällig. Besonders Geschäftsprozesse lassen sich mit Globus 2 nur schwer abbilden, denn es ähnelt zu sehr einem global verteilten Batch-System. Im Sinne einer offenen Architektur dis-

kutierten die Teilnehmer des 4. Global Grid Forum in Toronto im Februar 2002 bessere Ansätze [2].

Als neuer Grundgedanke kristallisierte sich die Vernetzung von Dienstleistungen heraus. Ähnlich der Informationsvernetzung im World Wide Web sollte eine neue Globus-Architektur auf kleinen geografisch verteilten Dienstprogrammen mit standardisierten Schnittstellen basieren. Vorteil: Die offene Entwicklung erlaubt es den Programmierern, ihre Software-Dienstleistungen der Allgemeinheit zur Verfügung zu stellen. Die Dezentralität und die zu erwartende Vernetzung der Dienste sollte ein schnelles Wachstum der Grid-Technik sichern.

Dienste sind Definitionssache

Das Globus-Toolkit 3 (GT 3) schlägt mit den Grid Services (GS) einen entsprechenden Standard vor, der sich im Rahmen der Open Grid Service Architecture

(OGSA) bewegt. Das Toolkit ist bereits seit Mitte 2003 erhältlich. Aktuell ist die Version 3.2. Ein Service ist darin als Protokoll sowie über das erwartete Verhalten beim Austausch vorgegebener Protokollmeldungen [3] definiert. Die Grid Services ähneln in ihrer Anlage damit solchen Webdiensten, wie sie beispielsweise Microsoft im Dotnet-Framework verwendet.

Webservices enthalten Techniken, die zugrunde liegende Dienstleistungen wie Programme beschreiben. Hinzu kommen Zugriffsmethoden oder Wege, über die sich Informationen zum Aufenthaltsort eines Dienstes beschaffen lassen (Discovery). Die Definition eines Webdienstes schreibt hingegen nicht vor, welche Programmiersprache ein Entwickler verwenden soll oder welche Systemsoftware vorhanden sein muss [4].

Kommunikation - gut eingeseift

Normalerweise tauschen sich Eigentümer und Nutzer eines Webservice über das Simple Object Access Protocol (SOAP) aus. Darin ergänzen sich in XML kodierte Kommandos und eine Menge an Daten (Payload), die zusammen mit den Befehlen durchs Netz gehen. Im Ergebnis stellt SOAP eine Variante des Remote Procedure Calls dar. Die Beschreibung eines Webservice erledigt der Anwender mit der Web Service Definition Language (WSDL, sprich Wisdel), die auch den Ansatz für Discovery-Mechanismen liefert. Wie bei SOAP ist WSDL in XML kodiert.

Ein WSDL-Dokument definiert die Schnittstelle des Webdienstes und legt damit auch die Struktur der an ihn verschickten SOAP-Dokumente oder alter-

nativer Methoden wie HTTP-Get/Post und die genaue Abfolge von Mitteilungen und Antworten fest. Auch alle Grid Services erfüllen diese Eigenschaften. Darüber hinaus integrieren sie ein Lifetime-Management, wodurch sich ein Grid-Dienst dynamisch erzeugen oder beenden lässt und zusätzlich eine bestimmte Lebensdauer bekommt. So ist beispielsweise Load-Management möglich, indem der Anwender weitere Service-Instanzen nur bei Bedarf startet. Die Instanzen müssen nicht einmal am selben Ort laufen wie die Schwesterinstanz. Zudem besitzen Grid Services einen zeitlich definierten Status. Fällt ein Dienst aus, startet einfach eine neue Instanz und garantiert so ein fehlertolerantes Verhalten. Ein herkömmlicher Webdienst hingegen beginnt bei einem Neustart mit seiner Grundkonfiguration.

Start in der Fabrik

Der Start neuer Grid-Service-Instanzen erfolgt durch einen speziellen Grid Service, den die Grid-Welt Factory (Fabrik) nennt. **Abbildung 1** zeigt den Ablauf anhand eines Beispiels: Benötigt ein Benutzer in einer mit Grid Services ausgestatteten Welt Informationen zum Bakterium Escherischia Coli, sucht seine Suchapplikation zunächst im Registry-Katalog nach passenden Minern und bei einem Storage Service Provider nach einem Ort, an dem sich die Informationen speichern lassen. Eine Miner Factory erzeugt dann beim Compute Service Provider eine Miner-Instanz. Analog legt ein Prozess eine Datenbank bei einem Speicher-Dienstleister an, der die Ergebnisse aufbewahrt.

Die Miner kontaktieren nun verschiedene geografisch verteilte Datenbank-Services auf der Suche nach den gewünschten Informationen. Die Ergebnisse schicken sie an den Storage Service Provider, wodurch der Benutzer Zugriff

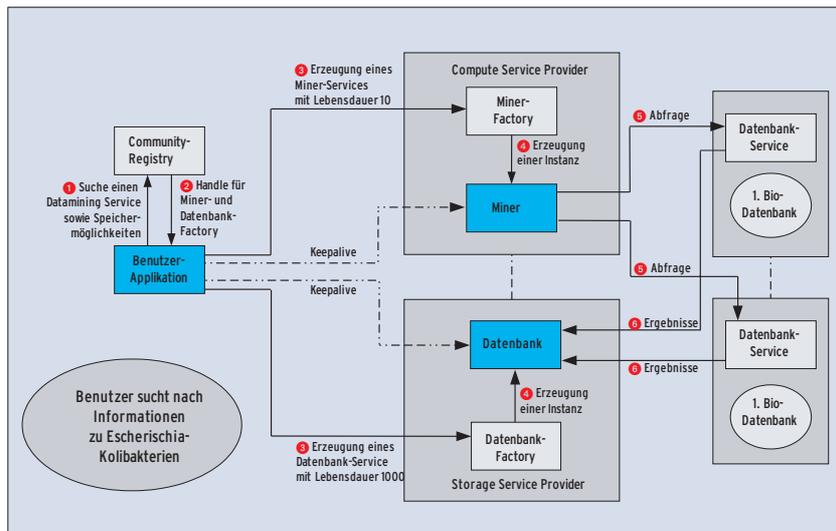


Abbildung 1: Bei der Suche nach Informationen über Escherischia-Kolibakterien greifen viele kleine Dienstleistungen koordiniert ineinander.

auf die Daten erlangt. Keepalive Messages sorgen dafür, dass Miner- und Datenbank-Service hinreichend lange erhalten bleiben, um die Suche durchzuführen. Da das System nicht weiß, wann der Benutzer die Ergebnisse abholt, ist die Lebensdauer der fertigen Informationsdatenbank länger als die des Miners. Hat der Anwender die Informationen heruntergeladen, beendet sich der dazu gehörige Dienst und gibt den Speicherplatz frei.

Viel geliebte Akronyme

Wer sich im Internet auf die Grid-Pirsch begibt, wird schnell mit der Vorliebe der Grid-Community für Akronyme aus vielen Buchstaben konfrontiert. Die unterschiedlichen Techniken kombiniert mit dem Bedarf nach schneller Kommunikation fördern die Geburt solcher Abkürzungen. Zwei wichtige Akronyme im GT-3-Dunstkreis sind OGSI (Open Grid Service Infrastructure) und OGSA (Open Grid Services Architecture), die als Oberbegriffe für die ganze im Rahmen des GT-3 anvisierte Grid-Architektur gelten.

Aufbauend auf Konzepten der Grid- und Webservice-Communities definiert das OGSA unter anderem, wie Schnittstellen von Grid Services aussehen und sich verhalten. Standardmechanismen, um neue GS-Instanzen zu schaffen, zu benennen und zu publizieren, sind darin ebenso enthalten wie die Abhängigkeiten und die Art der Zusammenarbeit der Grid Services. Als Teil von OGSA spezifiziert OGSI dabei die Schnittstellen, die ein Grid Service vorweisen muss, um überhaupt in das OGSA-Framework zu passen [2], [4].

Konkurrenz belebt das Geschäft

Während die Weiterentwicklung des Webservice-Standards beim WWW Consortium W3C [5] liegt, kümmert sich das Global Grid Forum um die Grid Services. Industrievertreter bemängeln aber häufig, dass es keine langfristige Entwicklungsplanung fürs Grid Computing gäbe. Die bloße Existenz zweier denkbar ähnlicher Technologien wie Web und Grid Services hat nicht gerade dazu bei-

getragen, die weitere Entwicklung voraussagbarer zu machen.

Ein Web- oder Grid Services Framework dient in erster Linie der Industrie, da Globus 2 zwar ausreicht, um Aufgaben datenintensiver Forschungszweige wie etwa der Teilchenphysik zu bewältigen. Doch sind die Techniken nicht dazu geeignet, Grid Computing in der Breite zu verankern und der Wirtschaft und Privatwendern schmackhaft zu machen. Der Pressemitteilung einer Initiative von Experten [6] ist zu entnehmen, dass die Entscheidung für das Web Service Resource Framework (WSRF) unter starkem Einfluss der Industrie entstand. Solche Entscheidungsprozesse sind im Highend-Bereich nicht ungewöhnlich, denn Impulse der Industrie beschleunigen oft die Entwicklung – IBM hat Vergleichbares mit Linux vorgemacht.

Neue Initiative fürs Web

Die gemeinsame Initiative schlug vor, OGSI als den niedrigsten Layer der OGSA-Spezifikation durch einen neuen Standard zu ersetzen. OGSA solle direkt auf kommerziellen oder freien Webdiensten, dem Web Service Resource Framework (WSRF) aufbauen [7], statt den Umweg über die komplexeren Grid Services zu gehen. WSRF besteht aus sechs Spezifikationen für Webservices. Unter anderem ist darin festgehalten, wie sich das Grid-Service-spezifische Lebensdauer-Management mit Hilfe von



Abbildung 2: Charlie Catlett, der scheidende Chairman des Global Grid Forum, befürwortete auf dem 10. GGF in Berlin die Entscheidung, das Web Service Resource Framework (WSRF) im Globus-Toolkit 4 zu verwenden.

Webservices abbilden lässt. Die Grid-Computing-Welt, soweit sie auf Globus basiert, bewegt sich demnach mit großen Schritten auf die Vereinigung mit der Webservice-Community zu.

WSRF im Kommen

Die endgültige Abkehr von OSGI und den Schritt zu WSRF soll erst die Version 4 des Globus-Toolkits bringen. Eine Beta-version von GT 4 ist für das zweite Quartal 2004 angekündigt (Stand bei Redaktionsschluss). Die finale Version 4 ist für das dritte Quartal geplant und liegt dann wie alle Globus-Toolkits auf [1] zum Download bereit.

Mitglieder der Grid- und Webservices-Gemeinde arbeiten mittlerweile gemeinsam an einer Standardisierung von WSRF. Dies geschieht im Rahmen eines technischen Komitees von OASIS (Organization for the Advancement of Structured Information Standards, [8]), eines weltweiten Konsortiums für die Entwicklung, Verbreitung und Zusammenführung von E-Business-Standards. In seinem Grußwort zum 10. Global Grid Forum lobte der scheidende GGF-Chairman Charlie Catlett (Abbildung 2) in einer Broschüre ausdrücklich die wachsenden Partnerschaften mit Vereinigungen wie OASIS, IETF und W3C.

Die Entscheidung für WSRF hat die Entwicklung von Globus im Speziellen und des Grid Computing im Allgemeinen nicht negativ beeinflusst, da sich lediglich der unterste Layer der Spezifikation ändert [9]. Alle darüber liegenden Dienstleistungen bleiben weitgehend unberührt und büßen auch nicht ihre Daseinsberechtigung ein. Im Gegenteil: Grid Computing profitiert durch Synergieeffekte bei der Zusammenarbeit mit der Web-Community.

Gewinner und Verlierer

Die Entscheidung, die Grid Services in die Version 3 zu integrieren, bringt für das Globus-Toolkit einen Technologieschub. Die offene Platt-

form vollzieht damit einen Paradigmenwechsel zur Version 2. Viele Experimente und Forschungsrichtungen haben deshalb noch nicht damit begonnen, sich auf die neue Technologie des GT 3 einzustellen. Die sich mit GT 4 vollziehende Substitution von OGSI zu WSRF wird nur noch wenig ändern.

Vor schwer wiegenden, kurzfristig zu lösenden Problemen stehen auch jene Arbeitsgruppen des Global Grid Forum und der Globus Community, die an Low-Level-Diensten des OGSI-Layers arbeiten. Bedenkt man aber den möglichen Nutzen eines gemeinsamen Vorgehens der Communities von Grid Computing und Webservices, erscheint das als ein geringerer Preis. Um mit dem Sänger Heinz Rudolf Kunze zu sprechen: „Eigene Wege sind schwer zu beschreiten, sie entstehen ja erst beim Gehen.“ (jre) ■

Infos

- [1] Globus: [<http://www.globus.org>]
- [2] GT-3-FAQ: [<http://www.globus.org/toolkit/gt3-faq.html>]
- [3] Anatomie eines Grid: [<http://www.globus.org/research/papers/anatomy.pdf>]
- [4] Physiologie eines Grid: [<http://www.globus.org/research/papers/ogsa.pdf>]
- [5] W3C: [<http://www.w3.org>]
- [6] IBM-Pressemitteilung zur WSRF-Initiative: [http://www.marketwire.com/mw/release_html_b1?release_id=61977]
- [7] WSRF: [<http://www.globus.org/wsrffaq.asp>]
- [8] OASIS: [<http://www.oasis-open.org>]
- [9] Entwicklung der GT-Versionen: [http://www.ggf.org/Meetings/ggf10/GGF10%20Presentations_b.htm]

Die Autoren

Dr. Rüdiger Berlich studierte Physik an der Universität Bochum. Von 1998 bis 2001 arbeitete er bei Suse, unter anderem als Technical Manager (Support) der US- und als Geschäftsführer der britischen Niederlassung. Im Januar 2004 promovierte er an der Uni Bochum im Bereich der Teilchenphysik und des verteilten Rechnens.

Dr. Marcel Kunze promovierte im Bereich Physik an der Universität Karlsruhe. Von 1990 bis 2001 war er als Assistent und Dozent an der Ruhr Universität Bochum tätig. 2002 wechselte er zum Forschungszentrum Karlsruhe und leitet heute die Abteilung Grid Computing and E-Science im Institut für Wissenschaftliches Rechnen.