



Der GridFTP-Client überwacht große Datenübertragungen

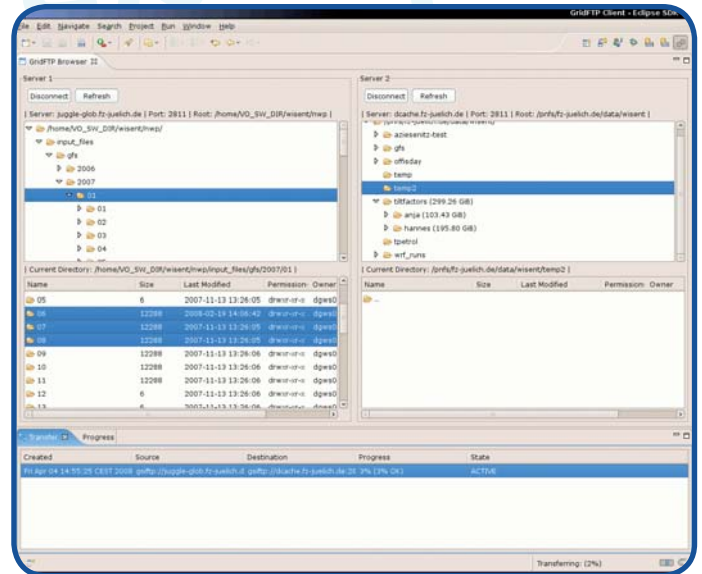
Jan Ploski, Jan.Ploski@offis.de, OFFIS

Das GridFTP-Protokoll unterstützt die effiziente Datenübertragung großer Datenmengen zwischen kompatiblen Servern. GridFTP-Server können entweder in Clustern installiert sein, die für die Ausführung von wissenschaftlichen Berechnungen zuständig sind, oder sie können den Zugriff auf dedizierte Speichereinheiten anbieten. Die letzteren können insbesondere für die mittelfristige Datensicherung gebraucht werden, wodurch es auch für kleinere wissenschaftlichen Einrichtungen möglich wird, große Datenmengen zu behalten, die für spätere Vergleiche oder Weiterverarbeitung benötigt werden.

Im D-Grid -Kontext wird vor allem die GridFTP-Implementierung von Globus Toolkit 4 als der Datenübermittlungsdienst in einzelnen Clustern eingesetzt. Vergleichsweise weniger bekannt, aber ebenso wichtig ist die GridFTP-Implementierung von dCache, der Middleware, die die dedizierten D-Grid-Speicherelemente in Jülich und Karlsruhe mit mehreren Terabytes an Bandkapazität zur Verfügung stellt.

Im Forschungsprojekt WISENT (Wissensnetz Energiemeteorologie) stellten wir fest, dass überraschenderweise noch keine Client-Software existierte, die einen bequemen, interaktiven und zuverlässigen Zugang zu den beiden Arten von GridFTP-Servern bot. Angetrieben durch Anwenderanforderungen entwickelten wir in OFFIS einen grafischen GridFTP-Client, der in seiner Benutzungsschnittstelle zwar einer gewöhnlichen FTP-Software ähnelt, aber zugleich die für große Datentransfers (z. B. mit >20000 Dateien, >500 GB Volumen) notwendigen Fehlertoleranzeigenschaften besitzt. Die Software wurde mit der dCache-Installation in Jülich getestet. Erfahrungen aus der Entwicklung wurden an dCache-Entwickler in Hamburg weitergereicht.

Die Weiterentwicklung unseres GridFTP-Clients wird als Open-Source-Projekt unter der Eclipse Public License erfolgen. Die erste Version für Eclipse-Nutzer ist installationsbereit; eine von Eclipse unabhängige Version ist geplant. Eine interaktive Präsentation, die die Software im Einsatz zeigt, und weitere Informationen finden Sie unter der Adresse <http://bi.offis.de/gridftp/>



Beschreibung der Abbildung: Die Bildschirmaufnahme zeigt den GridFTP-Client während der Übertragung von drei Datenverzeichnissen vom Juggle-D-Grid-Cluster zum dCache-Archiv in Jülich. Im linken Teil werden Inhalte des Quellserver dargestellt, im rechten Teil die des Zielservers. Im unteren Teil werden Statusinformationen zum Transfer wiedergeben.