

Bericht zum Grid Workflow Workshop - Oldenburg 2008

Stefan Gudenkauf
OFFIS - Institut für Informatik
E-Mail: stefan.gudenkauf@offis.de

Abstract

Am 03./04. März 2008 fand der erste Grid Workflow Workshop im OFFIS in Oldenburg statt. Über 30 Teilnehmer aus Wissenschaft und Wirtschaft tauschten sich im Rahmen der D-Grid-Initiative des BMBF (www.d-grid.de) über das Thema Workflows in Grid-Umgebungen aus, in denen verteilte Ressourcen standardisiert gemeinschaftlich genutzt werden. Workflows in Grid-Umgebungen können dabei als Arbeitsabläufe verstanden werden, die eine vordefinierte Abfolge von einzelnen Aktivitäten innerhalb eines Grids ausführen. Im Mittelpunkt des Workshops standen die Vorstellung und die Bündelung der Aktivitäten der einzelnen Projekte des D-Grid. Für einen Abgleich mit internationalen Initiativen sorgten die Vorträge der Keynote-Sprecher Prof. Dr. Wolfgang Gentzsch (D-Grid) und Dr.-Ing. habil. Thilo Kielmann (VU Amsterdam). Die Veranstaltung wurde durch Sun Microsystems unterstützt.

Der vorliegende Bericht fasst die Inhalte und wichtigsten Ergebnisse des Grid Workflow Workshop vom 03./04. März 2008 in Oldenburg zusammen. Die Präsentationen der Vortragenden und weitere Informationen sind auf der Internetseite https://bi.offis.de/bisgrid/tiki-index.php?page_ref_id=15 verfügbar.

1 Vorträge

Nach der Begrüßung durch W. Hasselbring (OFFIS) beschrieb T. Kielmann (Vrije Universiteit, Amsterdam) in der Keynote des ersten Tages, „The Grid on a Cloudy Day - Services, Abstractions, and Standards“, das Verhältnis von Grid Computing und Cloud Computing. Wichtige Punkte des Vortrags waren die Darstellung des Hype Cycle für Grid Computing, Beschreibungsversuche von Computing Clouds, Grids und Clouds in der Web-Infrastruktur, und die Bedeutung von Standardisierung. Ebenfalls wurde eine Übersicht über die Unterstützung Grid-relevanter Standards durch verschiedene Grid-Middlewares als „work in progress“ vorgestellt. Der zweite Vortrag, „Service Work-

flows in WSRF-Grids“ von T. Dörnemann (Philipps-Univ. Marburg), beschrieb die Aktivitäten der D-Grid Projekte Biz2Grid und InGrid. Dort werden Workflows als BPEL-basierte Orchestrationen von Diensten gesehen, die auf dem Web Service Resource Framework (WSRF) basieren. Hierzu wurden proprietäre Erweiterungen an der BPEL-Sprache und der verwendeten Workflow Engine ActiveBPEL vorgenommen. Die präsentierten Realisierungen und Werkzeuge befinden sich im produktiven Einsatz. A. Hoheisel (FIRST Fraunhofer) stellte in seinem Vortrag „Workflows in MediGRID und BauVOGrid“ die Handhabung von Workflows in den beiden im Vortragstitel genannten Projekten vor. Die inhaltlichen Schwerpunkte waren die Darstellung des sog. Grid Workflow Execution Service (GWES), die Kennzeichen von Grid Workflow-Beschreibungssprachen, die Eignung von Petrinetzen zur Workflow-Beschreibung und die Ressourcenabbildung (Scheduling). Die Workflow-Modellierung in den beiden Projekten erfolgt mittels Petrinetzen. Die Workflow-Ausführung wurde mit einer Live-Demo demonstriert. In ihrem Vortrag „Hierarchische Workflows in heterogenen Umgebungen“ hat G. Rümer (TU Chemnitz) hierarchische Workflows am Beispiel der Referenzarchitektur für E-Government (RAfEG) beschrieben. Neben der Vorstellung der Grundlagen standen die Beschreibung der Architektur von RAfEG, das Datenmanagement und Sicherheitsaspekte im Vordergrund. Im Vortrag „Workflows in UNICORE und dem A-Ware Projekt“ stellte B. Hagemeyer die Grid-Middleware UNICORE und das Projekt A-Ware vor. Wie in einigen D-Grid Projekten wird in A-Ware die Sprache BPEL zur Dienstorchestrierung verwendet. Interessant ist hier die Verwendung eines Enterprise Service Bus (ESB) zur Virtualisierung der Ressourcen. Ebenfalls interessant ist das erwähnte anstehende Release von UNICORE 6.1 für Mitte März 2008, das u.a. einen Rich Client umfassen wird und die Nutzung von domänenspezifischen Sprachen (DSLs) zur Workflow-Definition erlauben soll. Ch. Grimme (Univ. Dortmund) beschrieb mit dem Vortrag „Workflows in C3-Grid“ die Analyse, das Scheduling und die Bearbeitung von Workflows in C3-Grid. Als typische Workflows wurden Export und Ana-

lyse von großen Mengen verteilter und strukturierter Klimadaten vorgestellt. Ebenfalls wurde die Architektur der realisierten Ausführungsumgebung gezeigt. Zur Beschreibung der sog. Jobs dient die Job Submission Description Language (JSDL). Im letzten Vortrag des ersten Tages, „Workflows“ in WISENT“, stellte J. Ploski (OFFIS) die typischen Abläufe zur Nahe-Echtzeit-Vorhersage der Windenergieerzeugung vor. Dies erfolgt mit dem sog. Weather Research and Forecasting Model (WRF). U.a. wurde die Ausführung des WRF-Modells im D-Grid mit Condor-G und DAGMan als Metascheduler beschrieben. Ausgehend von der Feststellung, dass die Anwender in der WISENT-Community Shell Scripts schreiben anstatt Workflows zu modellieren, schloß er seinen Vortrag mit der Betrachtung der kontroversen Fragen „Wer braucht Workflows?“ und „Sind Grid Workflow-Modellierungssprachen und -Tools heute nötig, Grundlage für zukünftige Grid-Frontends, oder lediglich ein technisches Implementierungsdetail?“.

Die Keynote des zweiten Tages wurde von W. Gentzsch (D-Grid) gehalten. Thema war ein Überblick über „Internationale Grid Workflow Projekte“, der als Short Communication vorgestellt wurde. Es wurde deutlich, dass Workflows auch im Fokus vieler internationaler Grid-Projekte stehen, mit jeweils verschiedenen anwendungsspezifischen Realisierungen. Der nächste Vortrag, „Workflows in GDI-Grid“ von Ch. Grimm (RRZN, Leibnitz Univ. Hannover), beschäftigte sich vornehmlich mit dem projektspezifischen Workflow-Szenario der Lärmverbreitungssimulation und mit der speziellen Anforderung von GDI-Grid, zwei verschiedene Grid-Infrastrukturen betrachten zu müssen (GDI-Grid, D-Grid), und dessen Auswirkungen (Scheduling in beiden Grids). In ihrem Vortrag „Workflows in BIS-Grid“ stellten S. Gudenkauf (OFFIS) und A. Höing (TU Berlin) Workflow-Kennzeichen und Prozess-Engineering im D-Grid II-Projekt BIS-Grid, sowie die Architektur der zu entwickelnden Workflow Engine vor. Dabei wurde betont, dass im Projekt Workflows industriellen Ursprungs adressiert werden, und dass besonderer Wert auf eine nachhaltige Lösung gelegt wird, die auf industriell anerkannten Standards basiert. Die Vortragsreihe des letzten Tages wurde durch H. Lemke (Sun Microsystems) mit dem Vortrag „BPEL-Workflows und Service-Enablement von Applikationen auf Basis von Sun-Produkten“ abgeschlossen. Dieser stellte mit Java CAPS eine Service-orientierte Suite vor, die zur Anwendungsintegration innerhalb eines Unternehmens verwendet werden kann und hierzu umfassende Workflow-Fähigkeiten bietet. Hervorgehoben wurde die zukünftige Version Java CAPS 6, die als Open Source-Produkt geplant ist. Im Anschluss an die einzelnen Vorträge schloß jeweils eine kurze Diskussionsrunde an. Die wichtigsten Ergebnisse der Vortragsdiskussionen waren:

- Eine Aktualisierung und Kommunikation der von T. Kielmann vorgestellten Übersicht über die Un-

terstützung Grid-relevanter Standards durch verschiedene Grid-Middlewares ist notwendig und sinnvoll.

- In den einzelnen D-Grid Projekten gibt es parallele Entwicklungen, z.B. im Bereich Scheduling, Workflow-Editor und Ablaufsteuerung. Obwohl diese in den verschiedenen Anwendungsbereichen der Communities begründet sind, sollte untersucht werden, inwieweit hier Vereinheitlichung, Interoperabilität und/oder Koordination möglich ist.
- Generell kann im D-Grid zwischen solchen Projekten unterscheiden werden, die im Sinne der *eScience* wissenschaftliche Workflows betrachten, und solchen, die betriebliche Workflows im Sinne von Geschäftsprozessen adressieren.

2 Panel-Diskussion

Im Anschluss an die Vorträge des zweiten Tages fand eine Panel-Diskussion unter Beteiligung des Plenums statt. Teilnehmer des Panels waren Ch. Grimm, A. Hoheisel und J. Ploski. U.a. wurden die Unterschiede von wissenschaftlichen und industriellen Workflows, die Frage nach der „richtigen“ Modellierungsnotation für Workflows, die Bedeutung der dynamischen Dienstsuche und von Transaktionen, sowie die größten Herausforderungen beim Grid Workflow-Scheduling bzw. Ressourcen-Scheduling diskutiert. Die wichtigsten Ergebnisse der Diskussion waren:

- Unterschiede zwischen wissenschaftlichen und industriellen Workflows beziehen sich u.a. auf die Beständigkeit der Workflows, auf den Fokus im Workflow-Lebenszyklus, auf die am Lebenszyklus Beteiligten, und auf menschliche Interaktionen.

So liegt der Fokus bei wissenschaftlichen Workflows auf Definition und Monitoring, wobei die an der Definition Beteiligten ebenfalls die später Ausführenden sind. Menschliche Interaktionen sind in ihnen eher implizit zwischen einzelnen automatisierten Aktivitäten vorhanden. Für wissenschaftliche Workflows ist der Begriff „Job“ weit verbreitet und eine strikte Trennung zwischen Workflow und Workflow-Instanz ist wenig ausgeprägt. Wissenschaftliche Workflows haben tendentiell explorativen Charakter und behandeln häufig den Transport und die Verarbeitung sehr großer Datenmengen. Hierzu erfordern sie z.B. den direkten Datentransfer von einer Datenquelle zu einem Transferziel, ohne dass die initiiierende Stelle beteiligt ist (Third Party Transfer).

Bei industriellen Workflows ist der Lebenszyklus differenzierter. Bei der Definition kann z.B. zwischen

Kontrollfluss und Datenfluss, sowie zwischen Prozesserfassung und -modellierung unterschieden werden. Ebenfalls werden die am Lebenszyklus Beteiligten häufig separat betrachtet, was die Definition von Rollenmodellen erlaubt. Es wird strikt zwischen Workflow-Beschreibung und -Instanz unterschieden. Anpassung, Optimierung und Monitoring von Workflows, sowie menschliche Interaktionen besitzen eine große Bedeutung, wobei letztere häufig als spezielle Aktivitäten aufgefasst werden. Die in industriellen Workflows betrachteten Daten sind eher klein, dienen häufig der Ablaufsteuerung und müssen selten aufwändig verarbeitet werden.

- Eine D-Grid-weite einheitliche Modellierungsnotation ist nicht sinnvoll. Argumente gegen eine solche Modellierungsnotation sind die zuvor aufgeführten unterschiedlichen Beschaffenheiten von Workflows als auch unterschiedliche Anforderungen aus den Projekten. So erfolgt die Workflow-Modellierung in BIS-Grid stufenweise mit der Business Process Modeling Notation (BPMN) zur Kontrollflussmodellierung durch die Anwender und späterer Datenflussmodellierung zur Erstellung ausführbarer BPEL-Workflows durch BPEL-Experten, während Workflows in MediGRID von den Anwendungsentwicklern direkt als Petrinetze modelliert werden.
- Statt einer D-Grid-weiten Modellierungsnotation besteht der Bedarf nach Interoperabilität der einzelnen Workflow-Lösungen. So können die Ergebnisse eines Projektes anderen interessierten Projekten zugänglich gemacht werden.

Im Rahmen der Panel-Diskussion wurde ebenfalls diskutiert, wie die gefundenen Ergebnisse für jetzige wie auch zukünftige D-Grid-Projekte sinnvoll verwertet werden können. Zustimmung fand der Vorschlag, eine D-Grid Mailing-Liste für Grid Workflows einzurichten. Weiterhin bestünde in der Schaffung eines übergeordneten D-Grid Projekts zum Thema Workflows eine weitere Verwertungsmöglichkeit. Hier könnten repräsentati-

ve Workflow-Beispiele aus D-Grid-Projekten ausgewählt werden, die verschiedene Anwendungsbereiche abdecken. Diese könnten dann mit unterschiedlichen Workflow-Notationen und Beschreibungssprachen beschrieben und gegenüber gestellt werden. Das Grid Workflow Forum (www.gridworkflow.org) bietet bereits einen Überblick über Workflow-Editoren, Sprachen und Laufzeitumgebungen und könnte hierfür eine geeignete Publikationsplattform darstellen. In einem nachfolgenden Schritt könnte untersucht werden, in wie weit die verschiedenen Beschreibungen interoperabel sind. Gegebenenfalls ist es mit diesen Untersuchungen möglich, einen generischen, modularen und erweiterbaren Workflow-Editor zu erstellen, der für die Benutzer verschiedene anwendungsspezifische Modellierungsmöglichkeiten anbietet und unterliegend ein formales Modell verwendet, dass z.B. Verifikation und Simulation ermöglicht. Dieses könnte dann in die verschiedenen Beschreibungsformate exportiert werden, die von den Ablaufumgebungen der einzelnen D-Grid-Projekten erfordert werden.

3 Zusammenfassung und zukünftige Planung

Insgesamt ist uns sowohl der hohe Stellenwert von Workflows in Grid-Umgebungen aufgefallen als auch die Heterogenität der verfügbaren Realisierungen. Konsens über die zu verwendenden Modellierungsmittel war nicht erkennbar. Die vorgestellten Lösungen schienen stattdessen stark vom Anwendungskontext abhängig zu sein. Obwohl es auch kritische Stimmen zur Notwendigkeit von Workflows gab, schienen Workflows sowohl im D-Grid als auch in internationalen Projekten als geeignetes Mittel zur Handhabung und Ausführung von Abläufen gesehen zu werden. Dabei schien die Frage der Interoperabilität eine wichtigere Rolle zu spielen als die Frage der Vereinheitlichung.

Der nächste Grid Workflow Workshop findet voraussichtlich 2009 in Berlin (TU Berlin/FIRST Fraunhofer) statt. Weitere Details werden rechtzeitig bekannt gegeben.