Service Oriented Science: Globus Software in Action

Half-day Tutorial Proposal for
Third International Conference on e-Social Science

Ian Foster and Lee Liming
University of Chicago and Argonne National Laboratory

Abstract

Science and engineering fields are experiencing a multi-decade transformative revolution. Early elements of this revolution have included telecommunication networks and the Internet, distributed computing, the Grid, and e-Science. The "cyberinfrastructure" concept has recently begun integrating these technological elements with the social components that are needed to deepen their reach. The Globus community has supported science and engineering activities that exemplify this transformation. Having worked closely with a number of such activities, we are now observing another stage of the transformation: service-oriented science, where the products of science and engineering expand from static artifacts (papers, data) to include more dynamic results (online models, designs, instruments). Already, some scientists and engineers are "publishing" their computational models and engineering designs as online services, with standard interfaces that allow peers to interact with, evaluate, and incorporate these models into their own workflow.

In this presentation, we will share experiences from our group's work with leading-edge service-oriented science communities. Along the way, we will explain Globus software, how it is being employed in scientific collaborations, and the significance of these capabilities for the scientists who make use of them.

This half-day tutorial includes a number of case studies from successful e-Science projects and highlights the lessons learned from those experiences. The tutorial will explore the science requirements, the technical approach taken in each solution, and the results. We will introduce reusable Grid technologies from the Globus Toolkit such as GridFTP, MyProxy, MDS, and GRAM, as well as the Swift scientific workflow system. Along the way, we will highlight the general lessons that these experiences provided and the implications for new e-Science projects.

General Description of Tutorial Content

This half-day tutorial will be presentation-based with no hands-on requirement.

The tutorial will introduce and explain the service-oriented science transformation and the key roles played by Web services, the WS-Resource Framework (WSRF), and Globus software. The first half is an overview of current work, including details about the technology and applications in specific science collaborations. The second half offers more detailed case studies of how Globus software was used to overcome challenges in e-Science projects. Each challenge is presented in the context of a specific science project

and the results are reviewed with a scientific perspective as well as an information technology perspective.

Although the solutions described in the case studies can be reproduced in other projects, the emphasis will be on the methodology that led to the solutions and how that methodology can be reproduced in new situations.

Targeted Audience

Project leaders, team leaders, managers, IT executives, product managers, principal investigators, researchers, program managers, scientists, engineers

Anyone who has recently or soon will accept a position of responsibility in a current or potential e-Science project or product development activity

Audience Prerequisites

Attendees should be familiar with the basic principals of information technology. For example: general computer, network, and Internet concepts, client/server systems, databases, current types of commercial IT products. Familiarity with basic software engineering processes is helpful.

Detailed Outline

I. Scaling e-Science Impact
- A. Finding answers to the "big questions"
- B. Science 2.0, or Service-Oriented Science
- C. Hosting and management

II. WSRF and Globus Toolkit 4.0
- A. Web services in a nutshell
- B. Virtual Organizations - Defining community
- C. Composing services
- D. Provisioning services
- E. The Globus community – dev.globus

III. Case Study 1 – Data Replication for the Laser Interferometer Gravitational Wave Observatory
- A. The Challenge – Project goals and relevant requirements
- B. Ingredients – Relevant Grid technologies
  1. GridFTP
  2. Replication Location Service (RLS)
  3. Metadata Catalog Service (MCS)
  4. PyGlobus
- C. The Solution – Integration and deployment strategy
- D. Results – Science capabilities and lessons learned

(continued on next page)

IV. Case Study 2 – Scientific Workflow for Computational Economics
    A. The Challenge – Project goals and requirements
    B. Ingredients – Relevant Grid technologies
        1. Karajan
        2. GRAM
        3. Swift
    C. The Solution – Integration and deployment strategy
    D. Results – Science capabilities and lessons learned
V. Conclusions - General Lessons and Patterns

Presenter Biographies

Ian Foster is Director of the Computation Institute at Argonne National Laboratory and the University of Chicago, where he is also the Arthur Holly Compton Distinguished Service Professor of Computer Science. His research deals with distributed, parallel, and data-intensive computing technologies; the applications of those technologies to scientific problems; and the mechanisms and policies needed to create and operate scalable scientific "cyberinfrastructures," or Grids as he likes to call them. Dr. Foster is a fellow of the American Association for the Advancement of Science and the British Computer Society. His awards include the British Computer Society's award for technical innovation, the Global Information Infrastructure (GII) Next Generation award, the British Computer Society's Lovelace Medal, R&D Magazine's Innovator of the Year, and DSc Honoris Causa from the University of Canterbury, New Zealand.

Lee Liming is a Technology Analyst in the Distributed Systems Laboratory at Argonne National Laboratory and the University of Chicago. The Distributed Systems Laboratory employs computer scientists, software architects, and software developers working together to develop solutions to the challenging problems associated with distributed collaboration in science and engineering. His team supports a wide range of Grid-related projects, from infrastructure-focused projects like TeraGrid, Open Science Grid, and Enabling Grids for e-Science across Europe (EGEE) to science-focused projects like the Earth Systems Grid and Cancer Biomedical Informatics Grid. This team also contributes a sizable portion of the code in the open source Globus Toolkit. Early Grid communities in which Lee participated include the NASA Information Power Grid, the ASCI DisCom program, the National Computational Science Alliance, and NEESgrid. He had a leadership role in the NSF Middleware Initiative's GRIDS Center. In 2006, Lee became an Area Director for Software Integration for the NSF's TeraGrid. In this role, he is responsible for applying recent advances in software R&D toward creating new production-quality scientific computing capabilities for the TeraGrid community.