Grid Workflows: SODIUM and JOpera Projects

Cesare Pautasso
Department of Computer Science, ETH Zurich, Switzerland
pautasso@inf.ethz.ch – www.jopera.org
Annotate and normalize data

Extract raw spot intensities

Variability and error assessment

Determine parameters for error model

Significance assessment

Clustering

Determine Expression Pattern

Image processing

Data Preprocessing

Virtual Laboratories

in vitro | in silico

wet lab processing

MicroArray Scanner

Condition A

Condition B

Cell population

1 spot = 1 gene
 expresses level:

Green: A > B
Red:  A < B
Black: A = B

Expression level:

Green: A > B
Red:  A < B
Black: A = B
How to build a virtual laboratory?

Copy & Paste between different Websites

29 June 2006

Cesare Pautasso | www.jopera.org
How to build a virtual laboratory?

Copy & Paste
between different Websites

(Shell) Scripts
Tcsh, Bash, Makefiles, Python, Perl...

Programming
Java, C++, C#, Fortran...

Workflows
Graphical, Drag & Drop and Connect Environments
Grid Workflows

- High level model of a distributed computation
- Glue Grid services together
- Deal with heterogeneous environment
About JOpera

- **Modeling** virtual experiments behavior
  - Flow-based *composition language* (Visual & XML)
  - Development and Debugging tools for Eclipse
  - Composition of heterogeneous services

- **Execution** of the virtual experiments
  - Distributed engine (on a cluster of computers)
  - Autonomic platform (self-healing, self-tuning)
  - Extensibility (Eclipse plug-ins to provide custom service publishing and invocation adapters)
Drag, Drop and Connect
Run, Monitor, Steer and Debug
Dealing with heterogeneity in JOpera

- The JOpera composition language does not have to be changed when adding a new kind of service
A Growing User Community

- ETH Zurich
- NCSA
- Arjuna, UK
- SINTEF, Norway
- Locus, Norway
- Purdue University
- McGill University, Montreal
- Singapore Management University
- National University of Defence Technology, China
SODIUM

- Service Oriented Development in a Unified Framework
- EU IST-FP6 Project, 6 Partners
  - ETHZ, Switzerland
  - SINTEF, LOCUS, Norway
  - NKUA, ATC, Greece
  - MEDISYSTEMS, Romania

- UML Composition of Web, Grid and P2P Services
- Pilot applications in GIS, e-Health and emergency rescue services
Climate Modeling on TERAGRID

- Continuous processing of satellite feeds for climate modeling and weather forecasting
- JOpara a key part of the infrastructure to glue together the data and analysis services into Grid workflows
Cyberinfrastructure for e-Science at the National Center for Supercomputing Applications

- Grid Workflows important part of the Service Oriented Grid middleware stack
- JOpera Pilot Application: porting the data flow based “Data 2 Knowledge” toolkit to Eclipse
Why users like JOpera

- **High Level Workflow Language**
  - Data and Control Aspects (Graphical Representation)
  - Recursion, Iteration, Parallelism and Pipelining Constructs

- **Open and Extensible Component Model**
  - Run existing code without changes
  - Synchronous, Asynchronous, Streaming interaction
  - Web services support (Axis, WSIF)
  - Secure access to remote file systems and hosts (SSH, SCP)
  - Easy to integrate with existing schedulers (Condor already supported)
Why users like JOpera

- **High Level Workflow Language**
  - Data and Control Aspects (Graphical Representation)
  - Recursion, Iteration, Parallelism and Pipelining Constructs

- **Open and Extensible Component Model**
  - Run existing code without changes
  - Synchronous, Asynchronous, Streaming interaction
  - Web services support (Axis, WSIF)
  - Secure access to remote file systems and hosts (SSH, SCP)
  - Easy to integrate with existing schedulers (Condor already supported)

- **Strong Eclipse Foundation**
  - Platform Independent (Eclipse/Java)
  - Flexible, Extensible, Modular and Embeddable
JOpera Roadmap

- Eclipse 3.2
- Standalone JOpera Server
  - Remote Monitoring Client
- Streaming Support
  - Pipelining over RSS feeds (or other data stream sources)
- Lineage Tracking Perspective
  - Data Provenance Queries over Process Execution History
- Axis2 Upgrade (WSS, WSR)
Thank you for your feedback:
www.jopera.org
Free Download