Year 1 Internal Annual Presentation

Ketan MAHESHWARI

I3S-UNSA

November 17, 2008

Outline

1 Literature Review
- Taverna
- Triana
- Kepler

2 Enacting Cardiac Image Processing Application on MOTEUR

3 Work on Workflow Adapters

4 Work on gLite interface to Taverna v2

5 Future Work

Taverna

- SCUFL (Simple Conceptual Unified Flow Language) to express workflow.
- Advanced list processing operators.
- Not Grid Oriented.
- Good integration with WebService invocation.
- Sophisticated User Interface.
- Excellent community support via myexperiment.org.

Triana

- Pure java based scientific workflow system.
- Uses simple WSDL-like XML to express the workflow.
- Third party add-ons for grid and P2P computation.(eg. Grid Application Toolkit)
- Text and File processing capabilities of java are exploited.
- Data and file transfer capabilities for 'HTTP', 'FTP', 'GSIFTP' and 'file:///' protocols.
- No inter-task list processing.

Kepler

- Application level dataflow
- MoML (Modeling Markup Language) to express workflow.
- Separate paradigms for handling sequential (SDF) and parallel (PN) dataflow.
- "actor-director" metaphor for enactment.
- Dedicated provenance handling framework.
- Grid Support using RSL(Resource Specification Language)-based Globus job.
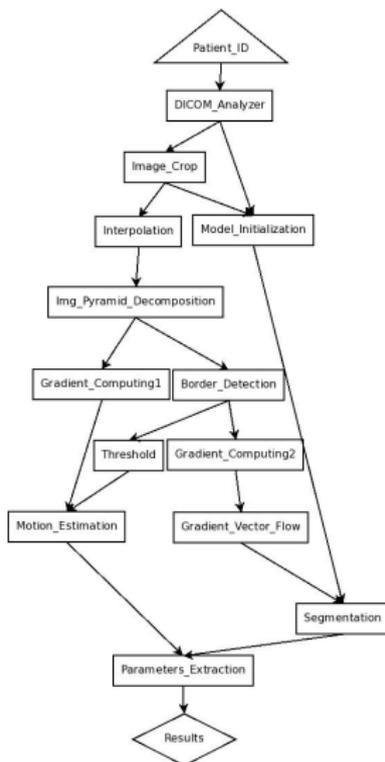
Findings

- Poor or No Support for Grid.
- Data Movement is mostly application-customized.
- Workflow engines are single strongest candidates to bridge applications and grids!
- Workflow interoperability does not exist.

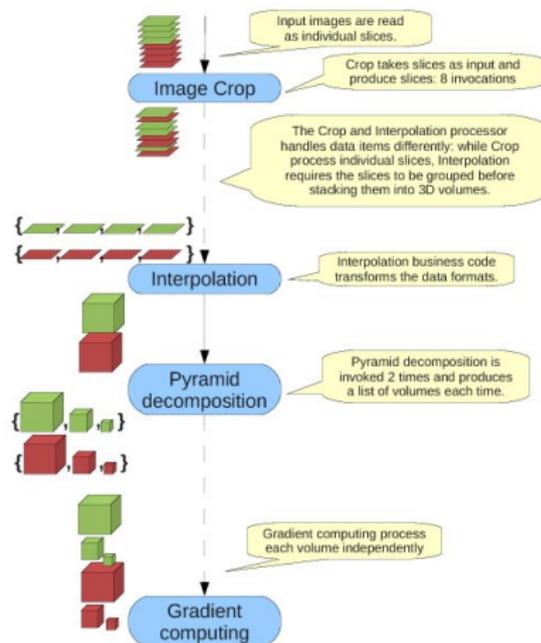# Outline

Cardiac Image Application

MOTEUR Enactment
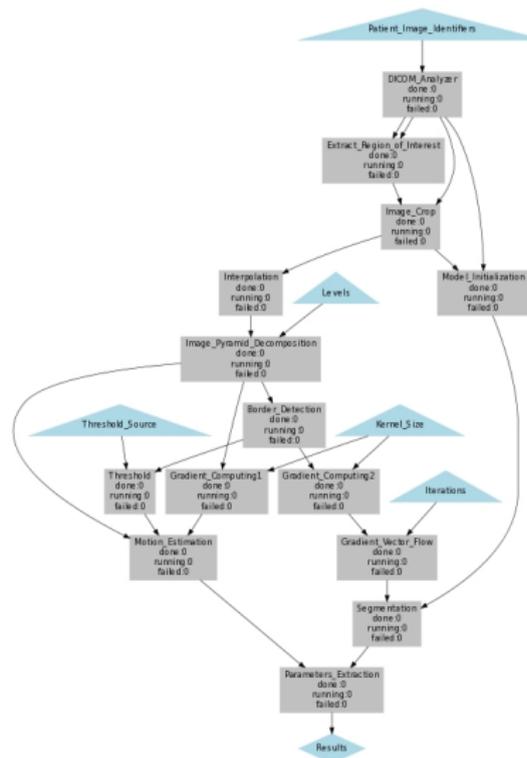
Limitations of Current Enactment

- Expressibility
- Data Flow
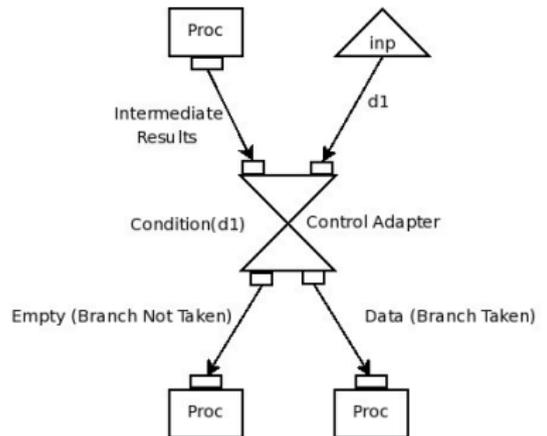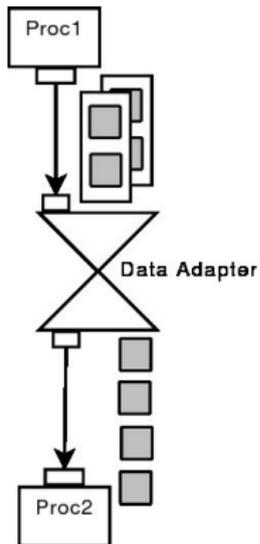- Notifications

[1]

[1] Image Source: J. Montagnat

Outline

1 Literature Review
- Taverna
- Triana
- Kepler

2 Enacting Cardiac Image Processing Application on MOTEUR

3 Work on Workflow Adapters

4 Work on gLite interface to Taverna v2
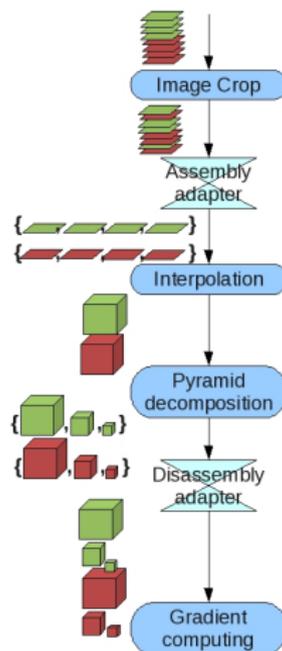
5 Future Work

Motivation

- Correct Implementation of complex dataflow
- Implementing control flow
- Intuitive dataflow
- Exploit parallelization

Data and Control Adapters

Cardiac Workflow with Adapters

2

---

[2] Image Source: J. Montagnat

Outline

Motivation

- Bring EGEE and Taverna communities closer.
- Taverna's sophisticated User Interface.
- Take inspiration from Taverna's list handling facilities.
- Mixed Mode Execution.

Screenshot

First Results

Patients=1, image slice sets=2 (1 fixed, 1 moving), total data size $\simeq$ 5MB;
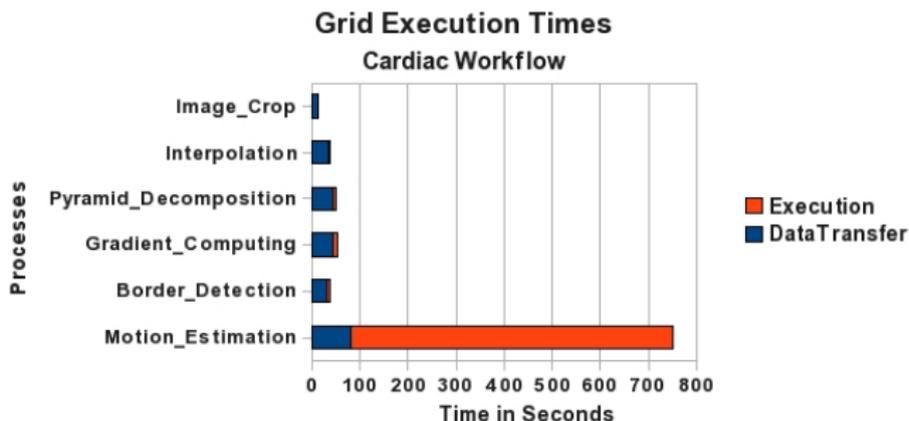
Current Limitations

- Needs a UI for initial data transfer.
- Polling at high frequency (I am looking into notification mechanism).
- Not yet asynchronous.

Conclusions

- Grid/Middlewares are <u>not</u> workflow oriented, they are batch oriented.
- Sequential execution only. Paradox??
- Plenty of manipulations required to make grid understand dataflow.

Outline

- Implementing Adapters.
- Functional vs Dataflow Languages and their interoperability.
- Enactment of Drug Discovery case.
- Implementing Grid Notification.
- Writing 300 pages of Thesis!!!

Thank You! Questions and Suggestions are Welcome!!