CIS 602: Provenance & Scientific Data Management

Scientific Workflow Provenance

Dr. David Koop
Reading Quiz
Reminders

• Reading Presentations
  - Volunteer for next class?
  - Look at web page, send me the top 5 topics you are interested in, in order
  - No reading response on the day you present
  - 15-20 minutes: summarize paper, present salient points, bring up highlights as well as concerns
  - We may stop to discuss/highlight certain aspects of the paper
  - You may present using the presentation software of your choice, please try to incorporate figures when appropriate

• Course Projects
  - Continue to think of ideas
Reading Responses

- Will work on feedback from last two submissions
- Summary:
  - Summarize the key points of the paper. Also, try to highlight unique points, not the background material covered before
- Critique:
  - No need to criticize the paper as a whole unless the central idea is something you wish to attack
  - Focus on a few points that bear some discussion
  - These will vary from person to person
- Questions:
  - Focus on the ideas not implementation or terms unless you are questioning results
Last Week

• Provenance
• Scientific Workflows
Scientific Workflows

```python
data = vtk.vtkStructuredPointsReader()
data.SetFileName('../examples/data/head.120.vtk')

contour = vtk.vtkContourFilter()
contour.SetInput(data.GetOutput())
contour.SetValue(0, 67)

mapper = vtk.vtkPolyDataMapper()
mapper.SetInput(contour.GetOutput())
mapper.ScalarVisibilityOff()

actor = vtk.vtkActor()
actor.SetMapper(mapper)

cam = vtk.vtkCamera()
cam.SetViewUp(0,0,-1)
cam.SetPosition(745,-453,369)
cam.SetFocalPoint(135,135,150)
cam.ComputeViewPlaneNormal()

ren = vtk.vtkRenderer()
ren.AddActor(actor)
ren.SetActiveCamera(cam)
ren.ResetCamera()
renwin = vtk.vtkRenderWindow()
renwin.AddRenderer(ren)

style = vtk.vtkInteractorStyleTrackballCamera()
iren = vtk.vtkRenderWindowInteractor()
iren.SetRenderWindow(renwin)
iren.SetInteractorStyle(style)
iren.Initialize()
iren.Start()
```
Scientific Workflows

• Differences from Business Workflows
• Workflow Engine
• Workflow Graph (Nodes=Actors/Modules/Activities and Edges=Connections/Channels)
• Data flows from input ports to output ports
• Parameters control how a module executes
• Workflows orchestrate different tools
• Workflows abstract computations and provide structure
Today: Provenance in Scientific Workflows

• Discussed Provenance and Scientific Workflows
• Now: how do the two work together?
A Framework for Collecting Provenance in Data-Centric Scientific Workflows

Y. L. Simmhan, B. Plale, D. Gannon, S. Marru

Presented by: Shrinivass Balasubramanian
Scientific Workflow Provenance

• From the paper:
  - “Provenance should identify the workflow, the services invoked during its execution, the methods/applications within the service that processed the invocation, the service invocation order and dependency, the time and status of the invocation, and the progress of the workflow, along with faults that may occur at various levels.” [Simmhan et. al]

• Structure of workflow = easier to capture/understand provenance

• Abstraction means less data needs to be captured
  - **Careful**: still may need/want to capture more than just the high-level view
Scientific Workflow Provenance

• Naming
  - Unique identifiers for all resources

• Overhead
  - Do not want extra work for users
  - Do not want lots of extra time spent recording provenance
Provenance Challenges

- Challenges to address provenance capture, interoperability, and provenance
- Web site: http://twiki.ipaw.info
- 1st Challenge: Understanding capabilities of different systems
- 2nd Challenge: Interoperability of provenance between systems
- 3rd Challenge: Evaluating the Open Provenance Model
- Different groups provided their solutions, allowed comparison and analysis and prompted general models like the Open Provenance Model and W3C PROV.