

Title

Introducing Workflow Integration of BioEnterprise Applications on a Java Platform

Instructors

Associate Prof Tan Tin Wee

Motivation

Too many bioinformaticists are doing too much repetitive tasks on systems and platforms that currently do not interoperate. Typically the bioinformaticist interfaces with other human operators, computers attached to data gathering devices for control and data input and data output (eg. DNA sequencers, microarrays), databases, and computational applications on compute servers. The manager of the bioinformatics team has to monitor and track the progress of all members of the team engaged in one or more of the above activities. The lack of an integrated architecture to handle integration of the above motivates the drive to search for such solutions.

Ignoring these issues typically results in delays in getting research results, low productivity, the data is prone to getting lost and error prone, the discovery process is often extended, effort duplication, questions during analysis are not asked because to ask them is too difficult due to lack of an architecture to handle integration. Gone are the days when database integration of heterogeneous distributed platforms was the sole challenge. The challenge today is integration of integrated databases with devices and computation and human operators and managers.

Objectives

1. To explain the need for automated workflow integration
2. To demonstrate examples of automated workflow integration
3. To coach participants in using a prototype system of automated workflow integration
 4. To describe with a few case studies of how automated bioinformatics workflows can be integrated in real-life situations using the prototype system.
5. To lay the framework for wide area network and Internet level of shared automated workflows beyond an intranet framework – the BioWorldWideWorkflow concept.

Expected Goals

1. Participants should be able to understand that routine sequence analysis and other bioinformatics activities carried out by manually is non-scalable and leads to regular staff burnout and boredom. All automatable processes and workflow should be maximised and human intervention minimised for workflow repetitive tasks. In this way, human intervention is reserved for key cognitive activities which cannot be machine-replaceable at the current state-of-the-art
2. Participants should be able to identify the key requirements and characteristics of a workflow integration system after experiencing a number of demos and examples.
3. Participants will be able to handle workflow integration in a hands-on approach using a Prototype based on a Java/RMI/XML/ platform

Intended Audience :

1. Bioinformaticists and Bioinformaticians doing hands-on research in support of genomics, proteomics, molecular biology research or bioinformatics resource support
2. Managers of such personnel.
3. Participants should have some experience in an operational environment that involves interaction among a team working towards life science projects using bioinformatics tools and databases.

Detailed Outline :

1. The Collaborative Environment in a Life Science lab or distributed project
2. Typical examples of Workflows
3. Analysis of the information flow in a workflow
4. How can the information flow be facilitated with a workflow integration system
5. Typical challenges of workflow integration in a few specific examples of real-life operations :
 - high throughput drug screening
 - microarray fabrication
 - analysis of microarray data
 - maintenance of databases
 - a simple genome sequencing and annotation workflow

6. Ingredients of a workflow integration system
 - Analogy with the World Wide Web and the Internet
7. The workflow browser
8. Resource discovery on the integrated workflow environment
9. Workflow submission and scheduling
10. Managing execution on remote systems
11. Introducing the KOOPPlatform as an example of an enterprise applications integrator and a workflow integration architecture
12. Setting up a Workflow editor and browser
13. Setting up a Resource meta server
14. Setting up a Workflow scheduler
15. Setting up a server daemon and executable repository
16. Interfacing with disparate software applications using a simple Unix runtime object
17. Writing a Java wrapper around a simple executable
18. Calling the executable through the workflow
19. Output of one process is the input of the next
20. Stitching together a workflow
21. Running the workflow
22. Example of Integration with grid computing software: using the globus grid
23. Hands on Practicing on a number of workflows and user parameterisation and customization
24. Designing your own workflow: a simple example
25. Summarising the basic requirements of a workflow environment in your own context: an exercise
26. Requirements of an ideal workflow integration system in a world wide context

Participants will receive:

1. Sample copy of the software for installation on their computer
2. Network access for their own personal computer in a local area network for the hands-on exercises only a limited number of notebooks will be available on-site. Windows and Unix platforms or emulated environments
3. Copy of notes and PowerPoint presentation
4. Online accounts for a limited trial period through the Internet for testing out workflow integration after they complete the tutorial.