MURPA Seminar: Friday 24 August 2012, 9am

Distributed Workflow-Driven Analysis of Large-Scale Biological Data using bioKepler

Speaker: Dr Ilkay Altintas
Venue: Seminar Room 135, Building 26, Clayton Campus, and Seminar Room H7.84, Caulfield Campus, Monash University

Abstract:
Next-generation DNA sequencing machines are generating a very large amount of sequence data with applications in many scientific challenges, placing unprecedented demands on traditional single-processor bioinformatics algorithms. Technologies like scientific workflows and data-intensive computing promise new capabilities to enable rapid analysis of next-generation sequence data. Scientific workflow systems promote scientific discovery by supporting the scientific workflow design and execution. They can be instrumented to optimize execution of workflow graphs across a variety of distributed computing and data resources while capturing provenance (execution history) of workflows and related data. The provenance contains information about how the execution took place and the products were derived, and it is crucial for enabling scientists to easily understand, reproduce, and verify scientific results. Based on this motivation and our previous experiences in bioinformatics and distributed scientific workflows, we are creating a Kepler Scientific Workflow System module, called "bioKepler", that facilitates the development of Kepler workflows for integrated execution of bioinformatics applications in distributed environments. This talk overviews scientific workflows with a focus on the open-source cross-disciplinary Kepler scientific workflow system, discusses the challenges related to next-generation sequencing data and explains the approaches taken in bioKepler to help with analysis of such data.

Biography:
Ilkay Altintas is the Director for the Scientific Workflow Automation Technologies Lab at the San Diego Supercomputer Center, UCSD where she also is the Deputy Coordinator for Research. She currently works on different aspects of scientific workflows in collaboration with various cross-disciplinary NSF, DOE and Moore Foundation projects. She is a co-initiator of and an active contributor to the open-source Kepler Scientific Workflow System, and the co-author of publications related to eScience at the intersection of scientific workflows, provenance, distributed computing, bioinformatics, networked observatory systems, conceptual data querying, and software modeling. Ilkay Altintas holds BS and MS degrees in Computer Engineering, both from Middle East Technical University in Turkey, and a PhD degree from FNWI, University of Amsterdam in The Netherlands. More info at:

http://users.sdsc.edu/~altintas/