

Automatically Composed Workflows for Grid Environments

The researchers are addressing the problem of manually creating workflows to be run on a grid by devising Grid middleware using AI planning to automatically construct workflows from high-level requirements. Current Grid middleware allows users to discover and schedule Grid resources and locate data, but each of these steps must be done manually. By automatically generating these workflows, Grid resources become more usable, reliable, and cost effective while becoming less complex.

The AI planner uses information about the Grid environment and represented knowledge of the high-level requirements to intelligently create a workflow that maximizes its productivity (e.g. intelligent scheduling, grouping of tasks), and then executes the workflow on the grid. The planner was tested at the Laser Interferometer Gravitational Wave Observatory (LIGO) on their grid to detect pulsars. The initial implementation is aimed to become “the foundation of a system with which they can perform production-level analysis.”

Infrastructure for Science Portals

Scientists working in accord with the National Science Foundation’s (NSF) Partnerships for Advanced Computational Infrastructure (PACI) program have been working on extending the Web from a place to access multimedia into a framework of computational applications and their associated data. The Internet sites that compose this framework give users access to many advanced scientific tools using the computational power of remote supercomputers (or, more recently, groups of thousands of PC processors). The novelty in their approach is modeling these Web-based science workbenches on the portals popular in the commercial sector at the time the paper was written. This would integrate advanced Grid services, large-scale data analysis, etc. in a fashion where users could customize not only the content, but also the presentation and set of available tools.

A large limitation on the researchers was their quickly growing need to discover, analyze, mine, and publish data, for which access to tools that performed these tasks on Grid resources would prove to be a great success. By making Grid resources available to desktops, laptops, and even mobile devices in a visual and customizable manner, this would add “real-time tool access to the asynchronous technologies of e-mail and the Web.” One limitation to this approach, especially when dealing with Grid resources based in the private sector, is the visibility of the computation. By that I mean the results are returned, but not necessarily the method or process by which those results were attained. Without public scrutiny of this process, it’s impossible to know that it is not erroneous, biased, or skewed in some manner.