Geophysical Research Abstracts,
Vol. 10, EGU2008-A-11281, 2008
SRef-ID: 1607-7962/gra/EGU2008-A-11281
EGU General Assembly 2008

Lighter-weight scientific workflows: balancing desktop ease with grid scale

D. L. Wang (1), C. S. Zender (2), and S. F. Jenks(1)

(1) Department of Electrical Engineering and Computer Science, University of California, Irvine, (2) Department of Earth System Science, University of California, Irvine

In response to the increasing proliferation of computational "farm" installations, grid technologies have emerged to better utilize and exploit the resulting gamut and glut of computational power. Workflow frameworks have emerged in order to standardize the specification, mapping, and execution of scientific workflows onto grids, and have seen success in mapping complex, unique, workflows in grid environments. However, between the large-scale institutional-sized data processing targeted by most grid frameworks, and the smaller user-sized personal desktop analysis, there exists a tremendous gap. "Medium-sized" computation borrows heavily from grid technologies' ability to exploit large, hetereogeneous computing resources, but trades away much of the extreme flexibility in order to better present an end-user experience minimally removed from desktop computing. The resultant ease allows even scientists' riskier data questions to be explored on a different scale.

We describe our medium-scale computational system, the Script Workflow Analysis for MultiProcessing (SWAMP) system, which is designed to utilize a common desktop paradigm– shell-scripted data analysis– to exploit and utilize computational power more commonly ascribed to large, high-performance systems. SWAMP trades off the full-expressivity and limitless flexibility of shell scripting, and implements a slightly simplified syntax and set of allowed executables. This provides a usability experience practically equivalent to desktop analysis in the geosciences domain, while transparently exploiting clustered, parallel computing. We describe the tradeoffs in implementing such a system, noting where the design goals of traditional grid frameworks are incompatible with desktop computing, and providing the rationale for our

chosen balance.