Parallel Computing in R using NetWorkSpaces

N. Carriero[(1)], J. Lai[(1)], M. Schultz[(1)], S. Weston[(1)], G. Warnes[(2)]
[(1)]Scientific Computing Associates Inc
[(2)]Pfizer Research Laboratory

Statisticians often encounter computationally intensive problems which can be efficiently processed by splitting the calculations across several computers. The primary barrier to doing this splitting is the complexity of the software tools and programming models traditionally used for handling and managing this splitting.

NetWorkSpaces, developed by Scientific Computing Associates, Inc. (SCAI), provides a simple, but powerful[1], "globally shared namespace" programming model which is very similar to the namespace concept available in R and other interactive programming environments. NetWorkSpaces is implemented via an open-source server plus "adapter" packages for R and other interactive programming environments.

The R 'nws' package (available on CRAN) provides a simple and clean interface for using NetWorkSpaces for data analysis and programming tasks. In particular, it provides a simple mechanism ('sleigh') for splitting a single computation across a set of collaborating machines for processing. This code takes care of all of the details of launching jobs, managing the interaction between jobs, and retrieving all of the results, making it straighforward for even novice R programmers to utilize.

In this presentation present an overview of NWS in R, the supporting system NWS, and several prototypical applications.

[1]The NetWorkSpaces model is exceptionally flexible and powerful despite its simplicity. Unlike many competing approaches, it naturally allows automatic load balancing, dynamic addition and removal of workers, arbitrary communication patterns, and other advanced features.