

Deploying and Benchmarking R on a Large Shared Memory System

Pragneshkumar Patel¹, George Ostrouchov^{1,2}, Luke Tierney³

1. University of Tennessee

2. Oak Ridge National Laboratory

3. University of Iowa

*Contact author: pragnesh@utk.edu

Keywords: high performance computing, parallel computing, multicore architecture, multithreading

We describe our experience in deploying *R* on a large shared memory system, *Nautilus* (SGI UltraViolet), which has 1024 cores (Intel Nehalem EX processors), 4 terabytes of global shared memory, and 8 NVIDIA Tesla GPUs in a single system image. This system is part of RDAV, the University of Tennessee's Center for Remote Data Analysis and Visualization sponsored by the National Science Foundation as part of TeraGrid XD.

One of our goals on *Nautilus* is to provide implicit parallel computing for serial *R* codes. Implicit parallelism automatically exploits multiple cores without changes to user code. Another goal is to provide an environment where users can effectively explore explicit parallelism through the use of many *R* packages that have been developed recently for parallel computing.

For implicit parallelism, we report on Intel's Math Kernel Library (*MKL*) and on the **pnmath** package. *MKL* includes a high performance multithreaded *BLAS* implementation. The **pnmath** package provides multithreading to many *R* math functions for operating on large vectors. We report on benchmark runs for many core count and vector size combinations, which we use for optimal speedup calibration. We observe speedups in excess of 600 for some compute intensive **pnmath** functions when operating on large vectors.

In addition to reporting on specific *BLAS* and **pnmath** functions, we include **R Benchmark 2.5** and some of its modifications. If time permits, we also mention experiences with RDAV center customers and some packages for exploiting explicit parallelism.

References

Luke Tierney (2008). Implicit and Explicit Parallel Computing in R, *COMPSTAT* 2008, Part II, 43-51.

Markus Schmidberger, Martin Morgan, Dirk Eddelbuettel, Hao Yu, Luke Tierney, Ulrich Mansmann (2009). State of the Art in Parallel Computing with R. *Journal of Statistical Software*, August 2009, Volume 31, Issue 1.

Luke Tierney (2010). pnmath, <http://www.stat.uiowa.edu/~luke/R/experimental/>.

Intel Math Kernel Library,

http://software.intel.com/sites/products/documentation/hpc/composerxe/en-us/mklxe/mkl_userguide_lnx/index.htm.