HPC with R: The Basics

Drew Schmidt

Some chuckle at the thought of R as a high performance language. However, as the landscape of HPC changes to better accommodate data analysis applications and users, R is a natural choice due to its overwhelming popularity. Fortunately, many have already taken up the call to embed R in HPC environments.

In this tutorial, we will explore R's relationship to current HPC hardware, discuss basic performance analysis and benchmarking tools for R, describe some best practices for improving basic R code, and conclude with a discussion of R's parallel computing capabilities.

The tutorial is meant to give a broad overview of R's capabilities as an HPC language. We will not go into great depth on any one topic; instead, we will introduce packages and utilities which will be of general interest, and provide resources to enable the attendee to study any given topic further as their interests align. As such, the attendee should come away with a foundational understanding of R's capabilities as a tool for high performance computing. The material is therefore well-suited to professionals and service providers alike.

There will be no "hands-on" portion for the tutorial, but there will be "take home" exercises. To complete these or to optionally follow along with examples presented at the workshop, the attendee should have R installed on their laptop.

Presenter Bio

Drew Schmidt is a developer for the pbdR project for distributed computing with R. Currently a graduate student at the University of Tennessee, he holds an M.Sc. in mathematics from the University of Tennessee, and was formerly a research associate at the National Institute for Computational Sciences.