Setting Up a pbdR Environment

Installing MPI, R, and pbdR

Version 1.0

Prepared by the pbdR Core Team:

Drew Schmidt
Remote Data Analysis and Visualization Center,
University of Tennessee, Knoxville

Wei-Chen Chen
Computer Science and Mathematics Division,
Oak Ridge National Laboratory

Pragneskumar Patel
Remote Data Analysis and Visualization Center,
University of Tennessee, Knoxville

George Ostrouchov
Computer Science and Mathematics Division,
Oak Ridge National Laboratory

This publication was typeset using LaTeX.

Contents

1 Quick Introduction

In this guide, we will detail the necessary steps for how to set up a pbdR environment. What follows in the remaining sections is a very lengthy list of installation instructions; however, most users should find the process fairly straight-forward, and may not need (or want) all of the details we will provide unless something goes wrong. In any case, the short version for setting up a pbdR environment is to:

1. install R ; see http://cran.r-project.org/
2. install an MPI library; http://www.open-mpi.org/, or http://www.mpich.org/ for Windows
3. install the pbdR packages; see http://r-pbd.org/

Items 1 and 2 are interchangeable, and so if you already have R and/or an MPI library installed, then merely skip this/these step(s); there is no need to reinstall anything.

1.1 Installing R

This should be fairly painless. R has binary packages for every operating system you have heard of (and some you haven't), and the install should go fine. Of course, since R is open source, you are free to compile it yourself, should have have reason or need to do so. You can find both the source as well as binaries at the R project's main site: http://cran.r-project.org/.

Additionally, you may wish to customize your R build by compiling from source. For example, you may wish to link R with a high performance linear algebra library, such as MKL. See the *R Installation and Administration Manual* at http://cran.r-project.org/doc/manuals/R-admin.html for full details.

1.2 Installing MPI

For Linux and Mac users, we recommend installing OpenMPI, which is available from http://www.open-mpi.org/ in both binary and source formats.

1.3 Installing pbdR Packages

All released pbdR packages are available from http://cran.r-project.org/ which is the Comprehensive R Archive Network (CRAN). This is similar to the CPAN for perl or CTAN for LaTeX, although with many improvements and benefits over its competitors.

It is also possible to link pbdR with high performance linear algebra libraries, such as MKL. Figure 1

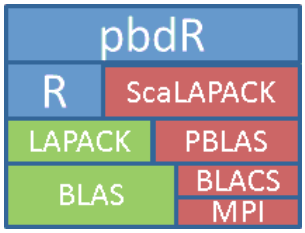

Figure 1: pbdR Relationships to Libraries

offers some insight into the package organization. See the pbdSLAP vignette for more details.

2 Mac OS X

Before starting, make sure you have installed XCode. You can find this in the Mac App Store.

terminal

2.1 Installing R

2.1.1 Installing from a Binary Package

2.1.2 Compiling from Source

You can find R sources from http://cran.r-project.org/sources.html

Generally, it should be enough to simply execute

```
./configure && make && make install
```

without problems.

2.2 Installing MPI

2.2.1 Installing from a Package Repository

2.2.2 Compiling from Source

If you want to install OpenMPI from source (I don't really recommend this unless this document is irrelevant to you in the first place), then the sources are available here: http://www.open-mpi.org/software/ompi/v1.6/ .

2.3 Installing pbdR

Installing pbdR should go smoothly. The simplest way to install the packages is from an R terminal, which will manage dependencies for you much like your distro's package manager. Additionally, our packages are available in the Fedora repositories.

2.3.1 Installing from CRAN

This is perhaps the simplest way to proceed, as R will handle any package dependency resolution for you. Simply start an R session (from the terminal[1], type R then press enter) and issue the command:

```
1 install.packages(<package>)
```

So for example, to install **pbdMPI**, you might execute:

```
1 install.packages(pbdMPI)
```

[1] Do *not* use the gui. See section 3.2 for details

2.3.2 Installing from the Shell

If you have downloaded a pbdR (or other R) package, then installing from the shell simply amounts to issuing the command:

```
R CMD INSTALL <package>
```

So for example, to install **pbdMPI**, you might execute:

```
R CMD INSTALL pbdMPI_0.1-6.tar.gz
```

2.3.3 Installing from Github

CRAN policy is such that updates to packages can not be made too frequently. For this reason, the development versions of our packages will have bugfixes and new features much more quickly than CRAN versions.

3 Installation Problems

During the course of installation, you may run into unrecoverable issues. The pbdR team does not support MPI libraries or R core, so if you have problems during that portion of the installation phase, we probably can not directly help you. However, there are still many great resources at your disposal, maintained by those individual projects.

3.1 R and MPI

If you have problems installing or customizing R, see the *R Installation and Administration Manual* at http://cran.r-project.org/doc/manuals/R-admin.html for help.

If you are having trouble installing an MPI library, you should see that library's official documentation. For OpenMPI, see http://www.open-mpi.org/community/help/ and for MPICH, see http://www.mpich.org/documentation/guides/ .

For the remainder, we will be addressing installation issues with pbdR packages.

3.2 pbdR

This is a quick list of potential problems you could encounter when installing pbdR packages. For additional troubleshooting or installation options, each package has a vignette which may offer additional useful information.

- **When compiling pbdMPI from source**, you may be required to pass a configure argument at compile time. So for example, if you have OpenMPI installed and were installing from the command line, then you would issue the command:

  ```
  R CMD INSTALL pbdMPI_0.1-6.tar.gz \
          --configure-args='--with-Rmpi-type=OPENMPI'
  ```

  or if installing from R:

```
1 install.packages("pbdMPI", configure.args='--with-Rmpi-type=
    OPENMPI')
```

See the **pbdMPI** vignette for more details.

- **If you are installing on a cluster** where you must install on the login node which can not execute `mpirun`, then pass the install option `--no-test-load`. So for example, if installing from the command line, then you would issue the command:

  ```
  R CMD INSTALL pbdMPI_0.1-6.tar.gz --no-test-load
  ```

  or if installing from R:

```
install.packages("pbdMPI", INSTALL_opts='--no-test-load')
```

- **If you are installing binaries on MAC OS X**, do not use the gui. You can install from source using the gui, or you can install binaries (or from source) using the terminal. But you can not install binaries using the gui. So if you want to install binaries, you should open Finder, then navigate to `Applications/Utilities/` and select `Terminal`. Next, type `R` and press enter. Now try to install the packages.

4 Running pbdR Scripts

This information is covered in *much* more detail in the pbdDEMO vignette, and should not be considered a substitute. However, there are two key points one needs to understand in order to use pbdR tools. Namely,

- pbdR codes are written in Single Program/Multiple Data style
- pbdR codes are executed in batch

For full details, see the pbdDEMO package vignette.

Below is a simple pbdR script. This will help you know if things are installed properly or not. To understand what the script is doing, or to learn how to do much more substantial things, you should see the pbdDEMO package vignette.

```
library(pbdMPI, quiet = TRUE)
init()

x <- comm.rank()

comm.print(x, all.rank = TRUE)

finalize()
```

To run the script, you must do so in batch (i.e., non-interactively). First save its contents to the file `my_script.r`, and then open a terminal. On Mac, you should execute the command:

```
mpirun -np 2 Rscript my_script.r
```