Introduction to R programming on Summit

Parallel R Programming on Summit

Youngseok Song[1]

[1] Department of Statistics
Colorado State University

SOARS, Spring 2018

Outline

1. Intro
2. About Summit
3. Setup
    - Install Packages
    - Batch
4. R Parallel Computing: Example
5. Discussion

Intro

My R Code is too slow!

When we need to speedup R code:

1. Optimizing Code
   - Vectorize
   - Pre-allocate Data Structure
   - Avoid loops, use apply family of functions
   - Give us some improvement, but not enough for some case.
2. Use `compiler` or Wrappers
   - `compiler` package: cmpfun()
   - Rcpp
   - Complied C, FORTRAN functions `.C()`, `.Fortran()`
3. **Parallal Computing**

Intro

Why Parallel Computing?

Q. Why do we consider parallel computing in R?

1. **Fast**
2. Easy to convert the code with apply family of functions
3. Cheap Computing resource

Examples:

- Size and Power simulation
- Run Multiple MCMC simultaneously
- Bootstrap, cross-validation, etc.

Summit is...

Hardware

- A new High Performance Computing system shared by **CSU** (22.5% of total), CU, Rocky Mountain Advanced Computing Consortium (RMACC).
- Full production on Feb 2017.
- About 400 TFLOPS (Rmax) (548.7 TFLOPS for World Top 500, Nov 2017).
- **380 Haswell** (General Computing, $380 \times 24 = 9,120$ cores, 128GB RAM/node), 10 GPU, 5 HiMem nodes.
- Storage: 2GB for Home, 250GB for Project.
- Condominium Computing Model

Summit is...

Softwares

- R (3.3.0/3.4.3), Python (2.7.11/3.5.1), Matlab (R2016b), etc.
- Intel, gcc, and pgi compilers
- MPI modules: opemmpi and Intel mpi
- git
- Can install custom software in a project directory

Summit is...

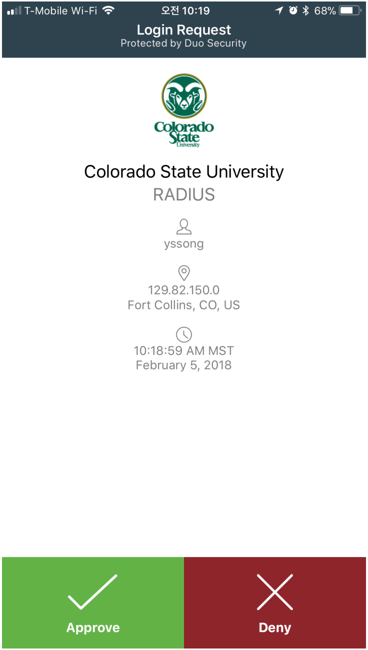
Support

- Located in CU, and Operated by CU IT staff.
- Reside on CU network and behind CU IT security infrastructure.
- CSU HPC staff provide support for CSU users.
- **50K Service Units** (SU) for new users, expires after 1 year.
- Need to apply project application for additional SU.
- Without SU, priority in the queue is very low.

Setup

Before run R

1. Get Summit accounts:
   - username will be the same as your eID.
   - `https://www.acns.colostate.edu/hpc/summit-get-started/`
2. Login to Summit
   - `ssh -l (username) login.rc.colorado.edu`
   - ex: ssh -l yssong@colostate.edu login.rc.colorado.edu
   - DUO Authentification
3. Login to scompile node (Until Janus @ CU be obsolete)
   - *ssh scompile*
4. **Load R module**
   - *ml R*



T-Mobile Wi-Fi 오전 10:19 68%

Login Request

Protected by Duo Security

Colorado State University

RADIUS

yssong

129.82.150.0
Fort Collins, CO, US

10:18:59 AM MST
February 5, 2018

Setup

Before run R

1. Get Summit accounts:
   - username will be the same as your eID.
   - `https://www.acns.colostate.edu/hpc/summit-get-started/`
2. Login to Summit
   - `ssh -l (username) login.rc.colorado.edu`
   - ex: ssh -l yssong@colostate.edu login.rc.colorado.edu
   - DUO Authentification
3. Login to scompile node (Until Janus @ CU be obsolete)
   - *ssh scompile*
4. **Load R module**
   - *ml R*

Setup

Before run R: Modules

Summit includes Lmod module systems:

Commands: Modules

- module avail
- module list
- module spider
- module load (or ml)

Setup

Before run R: Git Repository

Copy Github Repository

```
git clone https://github.com/EnigmaSong/Parallel_R_Summit
```

Including

- .bash_profile
- .R/MakeVars
- Batch file example
- Example R codes
- Wiki pages

Setup

R

R version 3.4.3 is installed (Checked Jan-25-2018).

1. Interactive Session vs. **Batch Processing**
2. Install R packages from CRAN
3. Install R packages from other repositories (Bioconductor, github)

Setup

R: MakeVars

Some package needs to specify additional settings in installation.

Example: glmnet package

- Need to specify FORTRAN compiler.
- Add the following line in $\sim/.R/MakeVars$

Create MakeVars

vim $\sim/.R/MakeVars$

MakeVars

FC=ifort FCFLAGS=-fPIC

Setup

R: Library Path

Specify Library Path:

- **Use Project directory!**: Manage build, Long term Use, Storage, Share w/ other users, etc.
- Add R_LIBS in $HOME/.bash_profile

Example

```
export R_LIBS=”/projects/$USER/R/library”
```

- Don’t forget: source $\sim$ /*.bash_profile*
- Example (available at here)

Install Packages

Packages Installation

A few core packages for parallel R:

- Rmpi: the de facto standard in R parallel computing
- snow: Easier communication
- foreach: Loop statements for parallel computing
- doSNOW: Parallel backend for the %dopar% operator.
- doRNG: Reproducible parallel foreach loops
- etc.

Install Packages

Rmpi Installation

Installing Rmpi

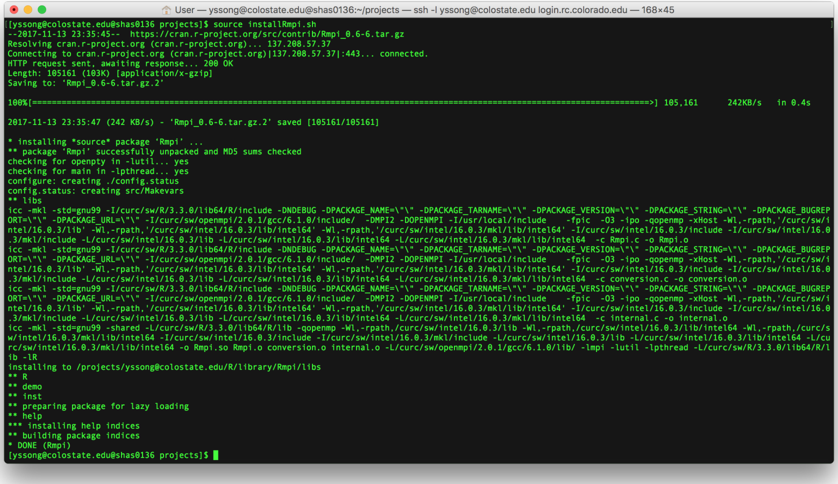
```
cd Parallel_R_Summit/Setting
source installRmpi.sh
```

Note: To install Rmpi

- Load R, openmpi module (Ver. 2.0.1)
- –no-test-load

Install Packages

Rmpi Installation


User — yssong@colostate.edu@shas0136:~/projects — ssh -l yssong@colostate.edu login.rc.colorado.edu — 168×45

```
[[yssong@colostate.edu@shas0136 projects]$ source installRmpi.sh
--2017-11-13 23:35:45--  https://cran.r-project.org/src/contrib/Rmpi_0.6-6.tar.gz
Resolving cran.r-project.org (cran.r-project.org)... 137.208.57.37
Connecting to cran.r-project.org (cran.r-project.org)|137.208.57.37|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 105161 (103K) [application/x-gzip]
Saving to: ‘Rmpi_0.6-6.tar.gz.2’

100%[=============================================================================================================================>] 105,161      242KB/s   in 0.4s

2017-11-13 23:35:47 (242 KB/s) - ‘Rmpi_0.6-6.tar.gz.2’ saved [105161/105161]

* installing *source* package ‘Rmpi’ ...
** package ‘Rmpi’ successfully unpacked and MD5 sums checked
checking for openpty in -lutil... yes
checking for main in -lpthread... yes
configure: creating ./config.status
config.status: creating src/Makevars
** libs
icc -mkl -std=gnu99 -I/curc/sw/R/3.3.0/lib64/R/include -DNDEBUG -DPACKAGE_NAME=\"\" -DPACKAGE_TARNAME=\"\" -DPACKAGE_VERSION=\"\" -DPACKAGE_STRING=\"\" -DPACKAGE_BUGREP
ORT=\"\" -DPACKAGE_URL=\"\" -I/curc/sw/openmpi/2.0.1/gcc/6.1.0/include/  -DMPI2 -DOPENMPI -I/usr/local/include    -fpic  -O3 -ipo -qopenmp -xHost -Wl,-rpath,'/curc/sw/i
ntel/16.0.3/lib' -Wl,-rpath,'/curc/sw/intel/16.0.3/lib/intel64' -Wl,-rpath,'/curc/sw/intel/16.0.3/mkl/lib/intel64' -I/curc/sw/intel/16.0.3/include -I/curc/sw/intel/16.0
.3/mkl/include -L/curc/sw/intel/16.0.3/lib -L/curc/sw/intel/16.0.3/lib/intel64 -L/curc/sw/intel/16.0.3/mkl/lib/intel64  -c Rmpi.c -o Rmpi.o
icc -mkl -std=gnu99 -I/curc/sw/R/3.3.0/lib64/R/include -DNDEBUG -DPACKAGE_NAME=\"\" -DPACKAGE_TARNAME=\"\" -DPACKAGE_VERSION=\"\" -DPACKAGE_STRING=\"\" -DPACKAGE_BUGREP
ORT=\"\" -DPACKAGE_URL=\"\" -I/curc/sw/openmpi/2.0.1/gcc/6.1.0/include/  -DMPI2 -DOPENMPI -I/usr/local/include    -fpic  -O3 -ipo -qopenmp -xHost -Wl,-rpath,'/curc/sw/i
ntel/16.0.3/lib' -Wl,-rpath,'/curc/sw/intel/16.0.3/lib/intel64' -Wl,-rpath,'/curc/sw/intel/16.0.3/mkl/lib/intel64' -I/curc/sw/intel/16.0.3/include -I/curc/sw/intel/16.0
.3/mkl/include -L/curc/sw/intel/16.0.3/lib -L/curc/sw/intel/16.0.3/lib/intel64 -L/curc/sw/intel/16.0.3/mkl/lib/intel64  -c conversion.c -o conversion.o
icc -mkl -std=gnu99 -I/curc/sw/R/3.3.0/lib64/R/include -DNDEBUG -DPACKAGE_NAME=\"\" -DPACKAGE_TARNAME=\"\" -DPACKAGE_VERSION=\"\" -DPACKAGE_STRING=\"\" -DPACKAGE_BUGREP
ORT=\"\" -DPACKAGE_URL=\"\" -I/curc/sw/openmpi/2.0.1/gcc/6.1.0/include/  -DMPI2 -DOPENMPI -I/usr/local/include    -fpic  -O3 -ipo -qopenmp -xHost -Wl,-rpath,'/curc/sw/i
ntel/16.0.3/lib' -Wl,-rpath,'/curc/sw/intel/16.0.3/lib/intel64' -Wl,-rpath,'/curc/sw/intel/16.0.3/mkl/lib/intel64' -I/curc/sw/intel/16.0.3/include -I/curc/sw/intel/16.0
.3/mkl/include -L/curc/sw/intel/16.0.3/lib -L/curc/sw/intel/16.0.3/lib/intel64 -L/curc/sw/intel/16.0.3/mkl/lib/intel64  -c internal.c -o internal.o
icc -mkl -std=gnu99 -shared -L/curc/sw/R/3.3.0/lib64/R/lib -qopenmp -Wl,-rpath,/curc/sw/intel/16.0.3/lib -Wl,-rpath,/curc/sw/intel/16.0.3/lib/intel64 -Wl,-rpath,/curc/s
w/intel/16.0.3/mkl/lib/intel64 -I/curc/sw/intel/16.0.3/include -I/curc/sw/intel/16.0.3/mkl/include -L/curc/sw/intel/16.0.3/lib -L/curc/sw/intel/16.0.3/lib/intel64 -L/cu
rc/sw/intel/16.0.3/mkl/lib/intel64 -o Rmpi.so Rmpi.o conversion.o internal.o -L/curc/sw/openmpi/2.0.1/gcc/6.1.0/lib/ -lmpi -lutil -lpthread -L/curc/sw/R/3.3.0/lib64/R/l
ib -lR
installing to /projects/yssong@colostate.edu/R/library/Rmpi/libs
** R
** demo
** inst
** preparing package for lazy loading
** help
*** installing help indices
** building package indices
* DONE (Rmpi)
[yssong@colostate.edu@shas0136 projects]$ 
```

Parallel R

Batch queueing

Summit uses `Slurm`: Queueing system
(Detail: `https://slurm.schedmd.com`)

Slurm Commands: Frequently Used

- sbatch job.txt
- scancel (job id)
- squeue $USER
- sreport -t hours cluster AccountUtilizationByUser start=2017-01-01 Users=$USER

Batch File

For detail: See UC Boulder RC User Guide

Parallel R

Batch file Example

Example of Batch file

```
!/bin/bash

#SBATCH -J Test
#SBATCH -p shas
#SBATCH --qos normal
#SBATCH --nodes 5
#SBATCH --ntasks-per-node=24
#SBATCH -o log/log.out
#SBATCH --mail-type=END
#SBATCH --mail-user=yssong@colostate.edu

R_PROFILE=$PROJECTS/R/library/snow/RMPISNOWprofile;
export R_PROFILE
```

Parallel R

Batch file Example

Example of Batch file (continued)

```
ml R
ml gcc
ml openmpi/2.0.1

date
START=‘date +%s‘
mpirun Rscript −−no−save $PROJECTS/A/control.R
END=‘date +%s‘
date

ELAPSED=$(( $END − $START ))
echo "Elapsed␣time␣(hrs):
$(echo␣"scale=10; $ELAPSED/3600"␣|␣bc)"
```

R Parallel Computing: Example

π calculation

Code

- Code for Serial and Parallel run is from RMACC 2017 HPC symposium
- Draw 1 million samples from $X \sim unif(-1, 1)$ and $Y \sim unif(-1, 1)$
- $\pi \approx 4\frac{\#\text{ Samples in the Unit circle}}{\#\text{ Samples}}$
- Compare Serial Run, Parallel Run with `Parallel`, Parallel run with `Rmpi` and `snow` on Cray and Summit
- Use 10 cores for Parallel Run
- Available at github

R Parallel Computing: Example

π calculation

Result (Sec)

Type	Summit	CSU Cray	2013 Macbook Pro
Serial	9.76	25.21	12.7
Parallel	2.66	13.27	
`Rmpi w/ snow`	0.54	1.6	

R Parallel Computing: Example

Skewed Normal Power

Skew Normal

- Code from Josh's SOARS talk in 2015 (CRAYFISHING AT CSU)
- $f(x;\xi,\omega,\alpha) = \frac{2}{\omega}\phi(\frac{x-\xi}{\omega})\Phi\{\alpha(\frac{x-\xi}{\omega})\}$
- Let $X_1, ..., X_n \sim SN(0, 1, \alpha)$. Check the power of the test $H_0 : \alpha = 0$ vs. $H_a : \alpha \neq 0$ for $n \in \{100, 110, 120, ..., 1000\}$ and $\alpha \in \{1, 2, ..., 100\}$.
- 1000 reps for each scenario.
- Available at github

R Parallel Computing: Example

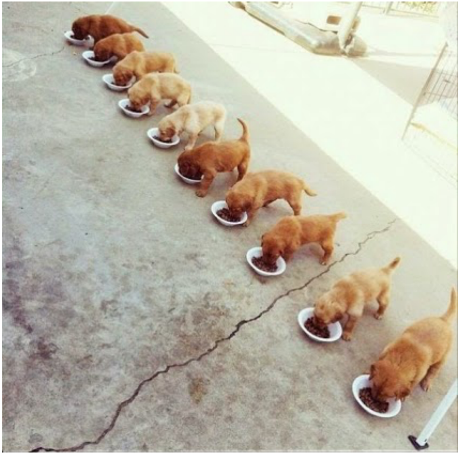
Skewed Normal Power

Result (Hours)

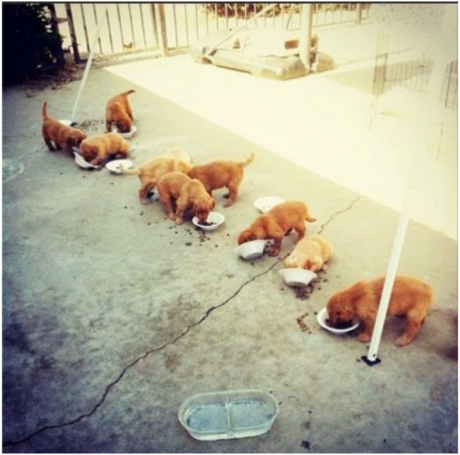
# Cores	Summit	CSU Cray
96	0.94	2.37
192	0.52	1.22
288	0.38	0.85

Discussion

Parallel Computing: Ideal and Real



Ideal



Real

Discussion

Want to know...

- Installing the latest version of `R` in personal Project directory
- GPU `R` computing: e.g., `tensolflow`

Further Resources

- **Github repository**: Parallel_R_Summit
- **Unix**
  - CS155
  - Internet resources (ex: [1])
- **R Parallel Computing**
  - CRAN Task View: High-Performance and Parallel Computing with R
- **Summit**
  - CSU HPC webpage
  - CU HPC webpage
  - RMACC HPC Symposium: Aug 15-17, 2017 (Aug 7-9, 2018)
- **More** `Rmpi` **Examples**
  - U Chicago