

R, Rcpp and Parallel Computing

Notes from our Rcpp Experience

Dirk Eddelbuettel and JJ Allaire

Jan 26-27, 2015

Workshop for Distributed Computing in R

Intro

One View on Parallel Computing

The whole “let’s parallelize” thing is a huge waste of everybody’s time. There’s this huge body of “knowledge” that parallel is somehow more efficient, and that whole huge body is pure and utter garbage. Big caches are efficient. Parallel stupid small cores without caches are horrible unless you have a very specific load that is hugely regular (ie graphics). [...]

Give it up. The whole “parallel computing is the future” is a bunch of crock.

Linus Torvalds, Dec 2014

Another View on Big Data



Brian L. Troutwine

@bltroutwine



 Follow

Most 'big data' problems are solved with:

- * GNU parallel
- * a single beefy machine
- * cron
- * some C++
- * a RDBMS

Tell your friends.



RETWEETS

61

FAVORITES

71



7:48 PM - 26 Dec 2014

R

CRAN Task View on HPC

Lots of existing work to draw from

- Package snow by Tierney et al a trailblazer
- Package Rmpi equally important
- Packages multicore nee parallel help Windows (I)users
- Hundreds of applications
- It just works for *data parallel* tasks

<http://cran.r-project.org/web/views/HighPerformanceComputing.html>

Rcpp: Early Days

In the fairly early days of Rcpp, we also put out RInside as a simple C++ class wrapper around the R-embedding API.

It got one clever patch taking this (ie: R wrapped in C++ with its own `main()`) and sticking it into MPI.

Rcpp: More recently

Rcpp has gotten fairly easy to use.

OpenMP is easy to use and widely supported (on suitable OS / compiler combinations).

So we added support. Use not as wide-spread. Errors have commonality: calling back into R.

RcppParallel

Easy-to-use-wrapper around Intel TBB (and TinyThreads where no TBB)

Users still attempt to use R objects. . .