R: fast and big strategies

Adolfo Álvarez
Analyx

2015 Use R! Lightning talk
Introduction

- Size is limited by RAM. Measure size by `print(object.size(), units="Mb"), and `pryr::object_size()`

- Time is limited by your life. Measure time by `system.time(), rbenchmark::benchmark(), microbenchmark::microbenchmark(), lineprof::lineprof()`

---

**One R Tip a Day**
@RLangTip

Compare the speed of several R expressions with the benchmark function: [bit.ly/Sppip1 #rstats]

6:06 PM - 21 Oct 2013

عداد پیوسته: 5  بازخوانی: 16
1) Write efficient code

- Vectorize
- *apply family
- Use linear algebra
- Avoid loops

Adolfo Álvarez
@adolfoalvarez

In case you missed it: Vectorization, matrix operations and readability in #rstats adolfoalvarez.cl/vectorization-...
5:42 PM - 7 Aug 2014

Adolfo Álvarez

Vectorization, matrix operations, and readability
My thoughts about R and analytics.

View on web
2) Use modern data input and manipulation packages

- Related packages: data.table / fread, dplyr / readr
3) Connect R to most SQL and no-SQL databases

- Related packages: dplyr, RCassandra, rmongodb, RSQLite, RPostgreSQL, ROracle, RMySQL, monetdb.R, etc...
4) Include C++ code to reduce execution time.

- Related packages: Rcpp, RcppArmadillo, RcppEigen, RInside

Twitter post by David Smith:

In a comparison of high-performance computing methods in R by @tonyfischetti, Rcpp is the clear winner
blog.revolutionanalytics.com/2015/06/a-comp... #rstats
9:56 PM - 1 Jun 2015

Likes: 4
5) Connect R to different Basic Linear Algebra Subprograms routines.

- Blas, OpenBlas, Atlas, Intel MKL, ...
6) Try another open or proprietary implementations of the R language.

7) Work in parallel

- Related packages: snow, parallel, doParallel, doSNOW
8) Use the power of your GPU

- Related packages: gputools, gmatrix, Rth

*Revolution Analytics @RevolutionR*

Parallel programming with #GPUs and #R via the blog by Norman Matloff
bit.ly/1Hd88Fk
5:35 PM - 28 Jan 2015

🔗 2  ★ 2
9) Compile your code to make it faster

• Related packages: **compiler**

---

Lee Hawthorn
@lee_hawthorn

Want to speed up an #rstats function? Compile it to byte-code!
[ow.ly/DF0BE](http://ow.ly/DF0BE)

9:49 AM - 1 Nov 2014

icable 2
10) Use big data specific packages:

- `ff`, `bigmemory`, `pbdR`, `biglm`, `bigrf`, `RevoScaleR`, ...

Lisa Winter
@lisachwinter

programming with big data in R #pbdR #Rstats buff.ly/1rF4ENR #bigdata
2:00 PM - 17 Oct 2014

→ 9 ★ 8
11) Connect R to modern big data environments:

- RHadoop, SparkR

---

Gregory Piatetsky
@kdnuggets

#ApacheSpark 1.4 includes #SparkR - R package to analyze large datasets buff.ly/1f1kBjh #BigData #DataScience
4:14 PM - 10 Jun 2015
    🤳 41  ⭐ 26

---

Analyx®
Strategic Predictive Customer Insights
12) Set up R cloud instances.

- Digital Ocean, Amazon, Azure, etc.
- Related packages: cloudyr
Thank you and keep contact!

www.adolfoalvarez.cl
@adolfoalvarez

www.analyx.com