

Evolving R for Commercial Use

David Smith
useR! 2010

R is awesome

- Open Source, Free
- Language
- Graphics
- Statistics
- Cutting-edge methods
- Community
- No Limits

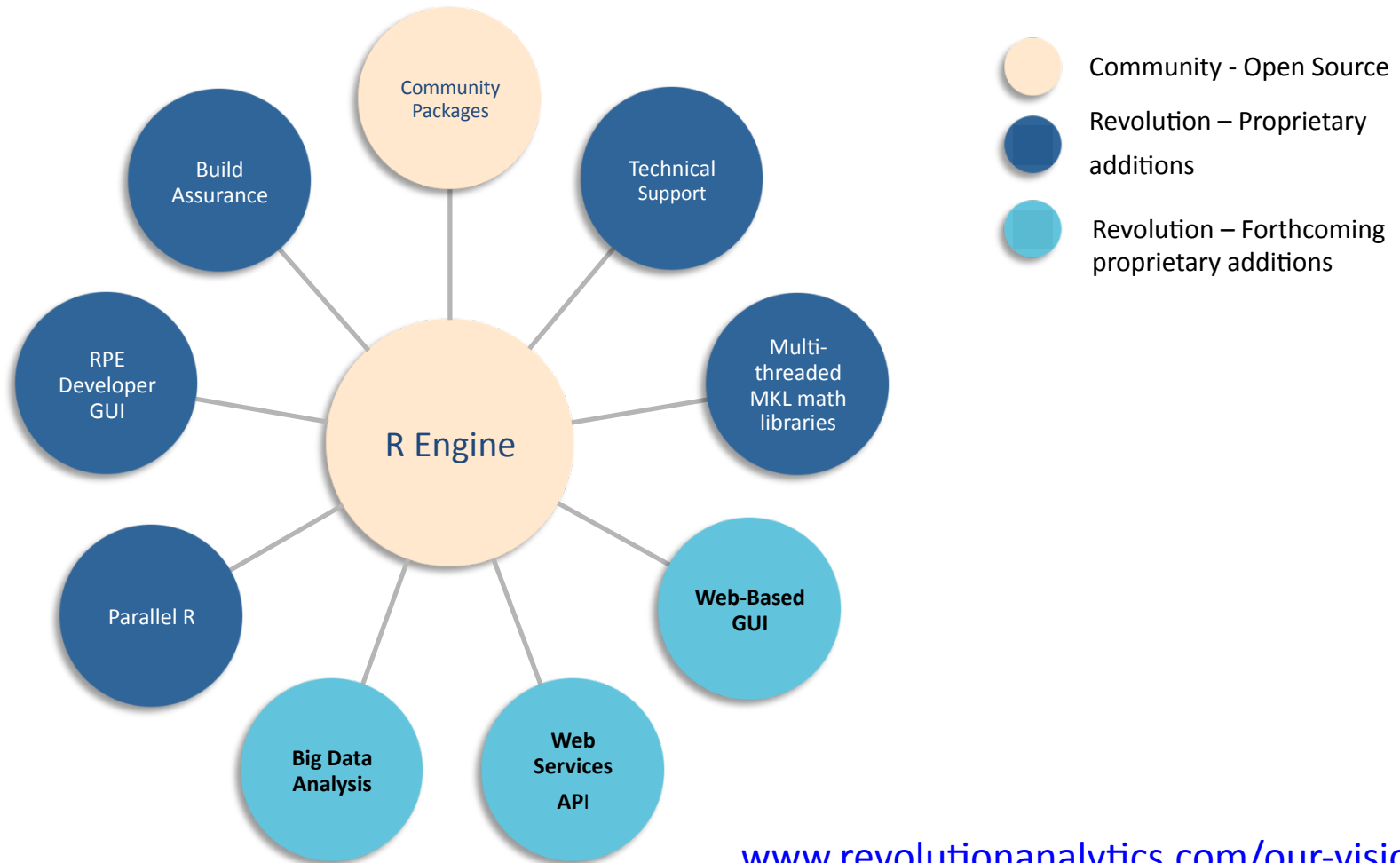


“R is the most powerful statistical computing language on the planet” – Norman Nie (CNET News, June 3 2010)

R at Work

- Windows (on the desktop)
- Developers (not necessarily statisticians)
- Managed by IT, not users
- Production applications *and* research
- Big data sets
- Deployed as part of a process

Revolution R Enterprise has Open-Source R Engine at the core



Open-Core Software Model

- Open-source “core” platform
- Bundled with proprietary add-ons that operate with core platform
 - Add-ons licensed/sold
 - Mark Radcliffe, OSI General Counsel:
 - <http://bit.ly/open-core>
 - revolutionanalytics.com/downloads/gpl-sources.php

R for Development

- Researchers prototyping
 - Point-and-click GUI
- Development teams building applications
 - Development environment
- Training
- Support
 - Someone to call for help



R Productivity Environment

The screenshot displays the RStudio IDE interface for R Enterprise for Windows (Administrator). The main editor window shows a script named 'WEBINAR(4-14-10)V3.R' with the following R code:

```
#  
# 5.1 Parallel Optimization dual core PC with doNWS  
#-----  
require('doNWS') # load the ParallelR packages  
s <- sleigh(workerCount=2)  
  
setSleigh(s)  
registerDoNWS(s) # use NetworkSpaces as the back-end  
M <- 100  
system.time (  
  SS <- foreach(j=5:(M-1), .combine=rbind, .packages=c('doNWS', 'ParallelR')) {  
    x <- rep(0,M)  
    for(k in min((j+2),M):M) {
```

The Revolution R Enterprise Console shows the execution of the script:

```
> setSleigh(s)  
> registerDoNWS(s) # use NetworkSpaces as the back-end  
> M <- 100  
> system.time (  
+ SS <- foreach(j=5:(M-1), .combine=rbind, .packages=c('doNWS', 'ParallelR')) {  
+   x <- rep(0,M)  
+   for(k in min((j+2),M):M) {  
+     R <- simpleRule(z$MSFT,j,k,9,Ra,Rb)  
+     Dt <- na.omit(R - Rb)
```

The Solution Explorer shows the project structure for 'WEBINAR (4-14-10)' (1 project), including 'RProject2' and several R scripts: 'debugger example.R', 'Script1.R', 'Select stock.R', 'WEBINAR (4-14-10) V2.R', and 'WEBINAR (4-14-10) V3.R'.

The Object Browser shows the environment structure, including 'Packages Installed' and 'Packages Loaded'. The 'op\$bg' object is highlighted, showing its type as 'character'.

The status bar at the bottom indicates 'Ready' and provides coordinates: 'Ln 106 Col 72 Ch 72 INS'.

IT: Fearing the worst, for you

- Installation (Upgrades)
- Virus checking
- Platform support (RHEL, 64-bit Windows)
- Multiple version control
- Support
 - One throat to choke!
- Contracts and licensing

R for Production Use

- Performance (Speed)
- Use computing resources
 - Clusters, Grids,
 - Cloud
- Scale to large data sets
- Validation

Intel MKL Benchmarks (Windows)

Computation	R 2.9.2	Revo R (1-core)	Revo R (4-core)	Speedup (4-core)
Linear Algebra¹				
Matrix Multiply	243 sec	22 sec	5.9 sec	41x
Cholesky Factorization	23 sec	3.8 sec	1.1 sec	21x
Singular Value Decomposition	62 sec	13 sec	4.9 sec	12.6x
Principal Components Analysis	237 sec	41 sec	15.6 sec	15.2x
Linear Discriminant Analysis	142 sec	49 sec	32.0 sec	4.4x
General R Benchmarks²				
R Benchmarks (Matrix Calc)	34 sec	6.6 sec	4.4 sec	7.7x
R Benchmarks (Matrix Functions)	20 sec	4.4 sec	2.1 sec	9.5x
R Benchmarks (Program Control)	4.7 sec	4 sec	4.2 sec	0x

1. <http://www.revolutionanalytics.com/why-revolution-r/benchmarks.php>

2. <http://r.research.att.com/benchmarks/>

Cloud Computing

- foreach replaces for loops
- Minimal code change required
- Parallel processing on CPUs on local machine, cluster, or cloud
- Significant speedups

```
# Birthday problem simulation run on 2.4 GHz Thinkpad T500 with dual core,  
# 64 bit cpu and 3 GB of RAM
```

```
birthday <- function(n) { # n is the number of people in the room  
  m <- 10000      # m s the number of rooms to simulate  
  x <- numeric(m)  
  for (i in 1:m) {  
    b <- sample(1:365,n,repl=T) # simulate birthdays for n people  
    x[i] <- n - length(unique(b))  
  }  
  mean(x) # average number of matches over m simulations  
}
```

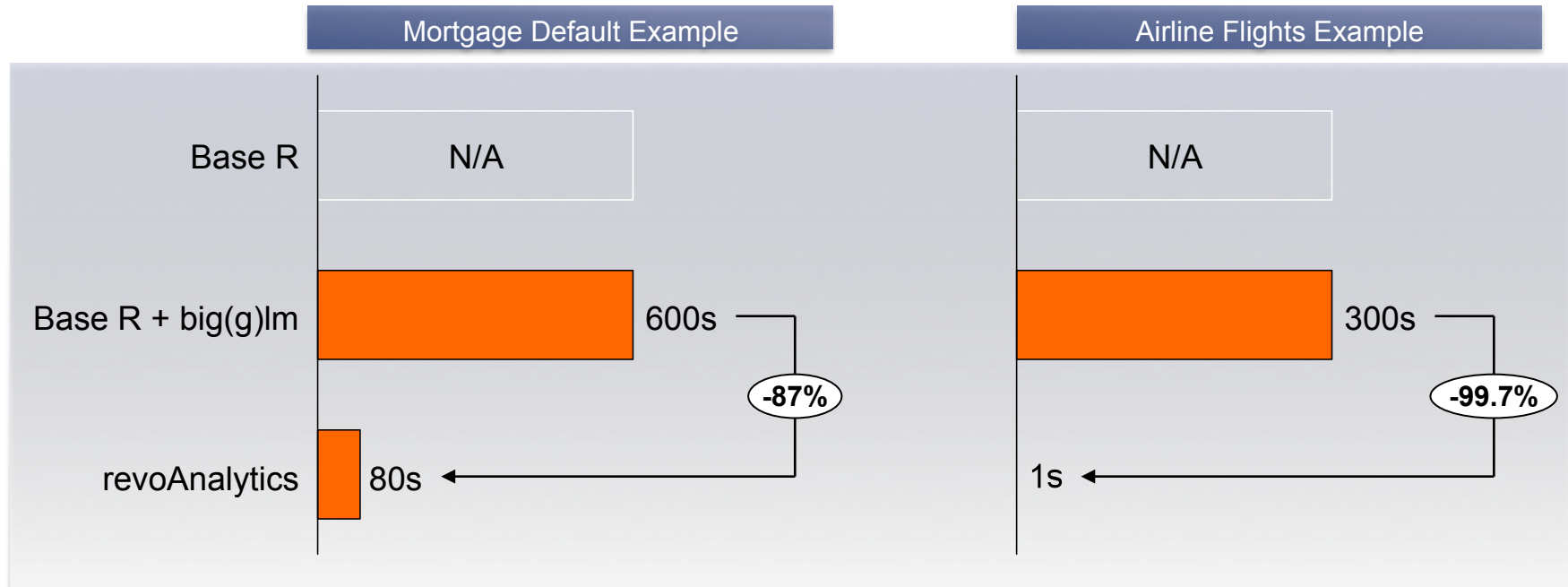
```
# run the loop sequentially  
system.time(for(j in 1:100) birthday(j))
```

```
# Results of sequential test run on 2.4 GHz Thinkpad T500  
# Elapsed: 50.94
```

```
# run the test with parallelR, two simultaneous workers  
library(nws)  
require("doNWS")  
s <- sleigh(workerCount=2)  
registerDoNWS(s)  
system.time(x <- foreach (j=1:100) %dopar% birthday(j))
```

```
# Results of parallel test  
# Elapsed: 28.75
```

revoScaleR Performance

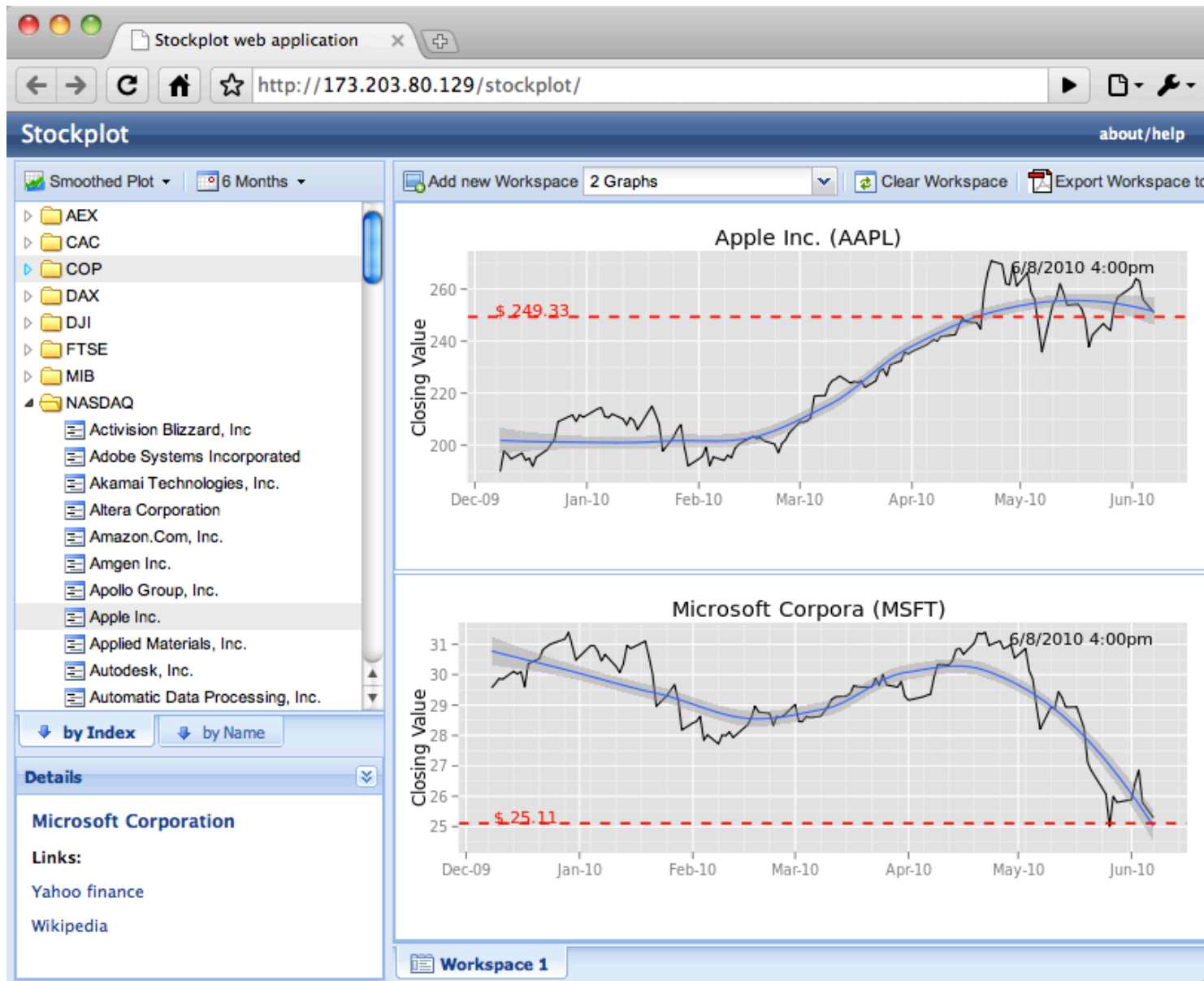


Dataset	10M rows & 6 variables	123M rows & 26 variables
Technique	Logistic regression	Linear regression
Machine	2-core laptop	8-core desktop
Alternative	Bigglm with all data in-memory	Biglm with sequential data chunking

Deployed Applications

- R as part of a process
 - Batch mode
 - Reporting
 - Interactive Applications
- Integration
 - With applications, data, and systems
 - Modern standards
 - Reliable (support many users, lots of data)
 - Users & Security
 - Maintenance

Web Services Integration



Community: Inside-R.org

inside-R

[Blogs](#)

[Language reference](#)

[Method Reference](#)

[Questions](#)

[Get R](#)

A Community Site for R – Sponsored by Revolution Analytics

Go



[Home](#) » [array {base}](#)

array {base}

Package: base

R Release: R 2.11.1

Multi-way Arrays

Description

Creates or tests for arrays.

Usage

```
array(data = NA, dim = length(data), dimnames = NULL)
  as.array(x, ...)
  is.array(x)
```

Arguments

data

a vector (including a list) giving data to fill the array.

About us

inside-R.org is a collection of resources about the R Project for the R Open Source Statistics Community. This site is sponsored by Revolution Analytics.

The site is currently in beta. Check back soon for more features – including user contributions, R Tips & Tricks, and more. Follow @inside_R on Twitter for updates about this site.

If you have questions or ideas for the site, please send us an email at community@inside-R.org.

Information about R

[What is R?](#)

[Why use R?](#)

[Resources for R](#)

Revolution R Enterprise

Production-Grade Statistical Analysis for Business

- ✓ High-performance R for multiprocessor systems
- ✓ Statistical Analysis of Terabyte-Class Data Sets
- ✓ Deploy R Applications via Web Services
- ✓ Easy-to-Use Graphical User Interface
- ✓ Parallel Programming on Clusters / Cloud
- ✓ Modern Integrated Development Environment
- ✓ Validation for use in regulated environments
- ✓ Telephone and email technical support
- ✓ Training and consulting services

Thank You!

David Smith

david@revolutionanalytics.com

blog.revolutionanalytics.com