REVOLUTION ANALYTICS

Revolution R Open

Denver R Users Group

6 Jan 2015

So why would we need another version of R?

- To provide support (need to control the build process)
- To provide better performance
- To solve some gaps in what's presently available

Agenda

- Company Background
- Revolution R Open Overview
- MRAN
- Demonstration

Agenda

- **Company Background**
- Revolution R Open Overview
- MRAN
- Demonstration

Revolution Analytics at a Glance

Who We Are

Only provider of commercial big data big analytics platform based on open source R statistical computing language

Our Software Delivers

Scalable Performance: Distributed & parallelized analytics
Cross Platform: Write once, deploy anywhere
Productivity: Easily build & deploy with latest modern analytics

Our Services Deliver

Knowledge: Our experts enable you to be experts
Time-to-Value: Our Quickstart program gives you a jumpstart
Guidance: Our customer support team is here to help you

Customers
200+ Global 2000

Global Presence
North America / EMEA / APAC

Global Industries Served

Financial Services
Digital Media
Government
Health & Life Sciences
High Tech
Manufacturing
Retail
Telco

REVOLUTION ANALYTICS

- Open Source development
  - Revolution R Open, RHadoop, ParallelR, DeployR Open, Reproducible R Toolkit
  - Project funding
- Community Support
  - User Group Sponsorship
  - Meetups
  - Events sponsorship
  - Revolutions Blog

Agenda

- Company Background
- **Revolution R Open Overview**
- MRAN
- Demonstration

The Revolution R Product Suite

Revolution R Open

New!

- Free and open source R distribution
- Enhanced and distributed by Revolution Analytics

Revolution R Plus

New!

- Open-source distribution of R, packages, and other components
- Enhanced, supported and indemnified by Revolution Analytics

Revolution R Enterprise

- Secure, Scalable and Supported Distribution of R
- With proprietary components created by Revolution Analytics

Revolution R Open is:

- Enhanced Open Source R distribution
  - Based on the latest Open Source R (3.1.2)
  - Built, tested and distributed by Revolution Analytics
- High-performance R language engine
  - Multi-threaded processing with Intel MKL
  - Up to 20x speed increase, without changing a line of code
- Compatible with all R-related software
  - CRAN packages, Rstudio, third-party R integrations, …
- Reproducible R Toolkit
  - Facilitate sharing, stability and traceability of R scripts
- MRAN website `mran.revolutionanalytics.com`
  - Enhanced documentation and learning resources
  - Discover almost 6000 free add-on R packages
- Open source (GPLv2 license)
  - 100% free to download, use and share

What affect does MKL have?

Revolutions

Learn more about using open source R for big data analysis, predictive modeling, data science and more from the staff of Revolution Analytics.

« R in Production: Controlling Runtime | Main | A first look at Distributed R »

October 22, 2014

How the MKL speeds up Revolution R Open

by Andrie de Vries

Last week we announced the availability of Revolution R Open, an enhanced distributi
One of the enhancements is the inclusion of high performance linear algebra libraries
specifically the Intel MKL. This library significantly speeds up many statistical calculati
the matrix algebra that forms the basis of many statistical algorithms.

Several years ago, David Smith wrote a blog post about multithreaded R, where he ex
benefits of the MKL, in particular on Windows machines.

In this post I explore whether anything has changed.

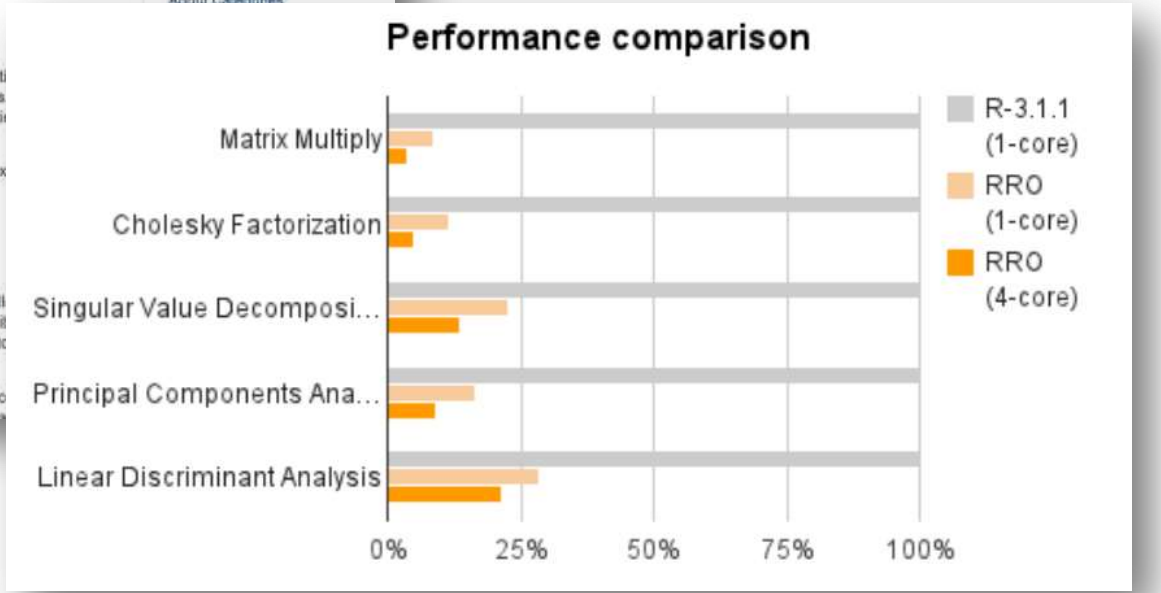
What is the MKL?

To best use the power available in the machines of today, Revolution R Open is install
default with the Intel Math Kernel Library (MKL), which provides BLAS and LAPACK li
functions used by R. Intel MKL makes it possible for so many common R operations to
the processing power available.

The MKL's default behavior is to use as many parallel threads as there are available c

Information

About this blog
Comments Policy
About Categories

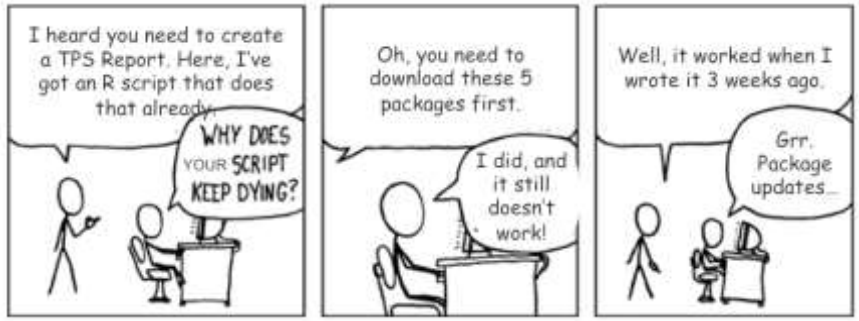


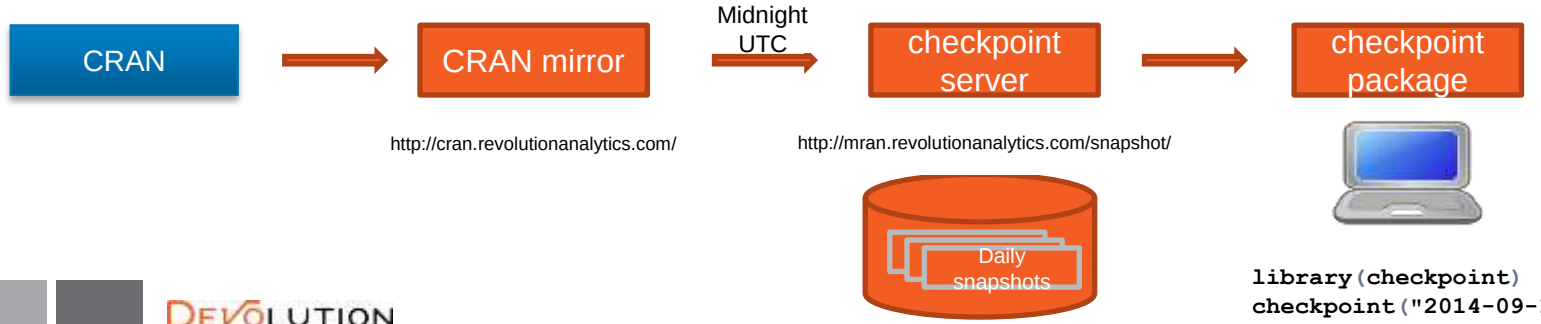
Source: http://blog.revolutionanalytics.com/2014/10/revolution-r-open-mkl.html

REVOLUTION ANALYTICS

Reproducible R Toolkit

- Provide control when working with R packages
  - 6000+ packages available, with new versions daily
- Facilitate sharing between collaborators
  - Ensure same versions of R and packages are being used
- Provide durability to R applications
  - Easily install correct package versions
- Ensure traceability of analytic results
  - Reliably reproduce results now and in the future
  - Trace versions of R packages being used





Using Revolution Analytics' Reproducibility Tools

- Scenario 1: Set up a consistent, company wide R environment
  - Have users download RRO
  - All users will get the base and recommended packages as of 12/1/14
  - For each project, R user run checkpoint to download a consistent set of packages that are appropriate for that project

- Scenario 2: With or w/o RRO share scripts synced to a snapshot
  - Have the user with whom you are sharing put your scripts in a separate project and download the checkpoint package
  - Have the user run `checkpoint("yyyy-mm-dd)` with a date appropriate for your project
  - Checkpoint will automatically download the correct version of the packages used in the scripts

Using checkpoint

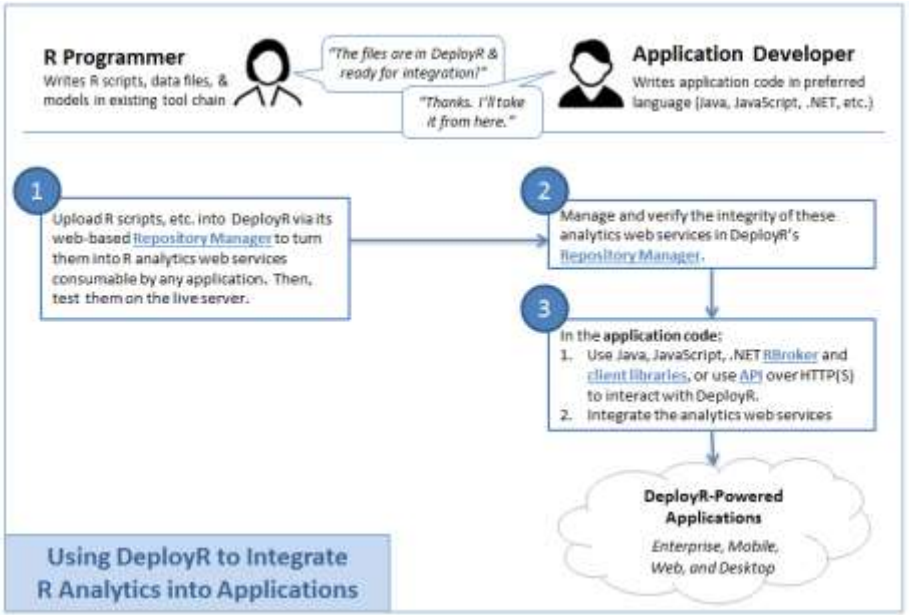
- Easy to use: add 2 lines to the top of each script

```
library(checkpoint)
checkpoint("2014-09-17")
```

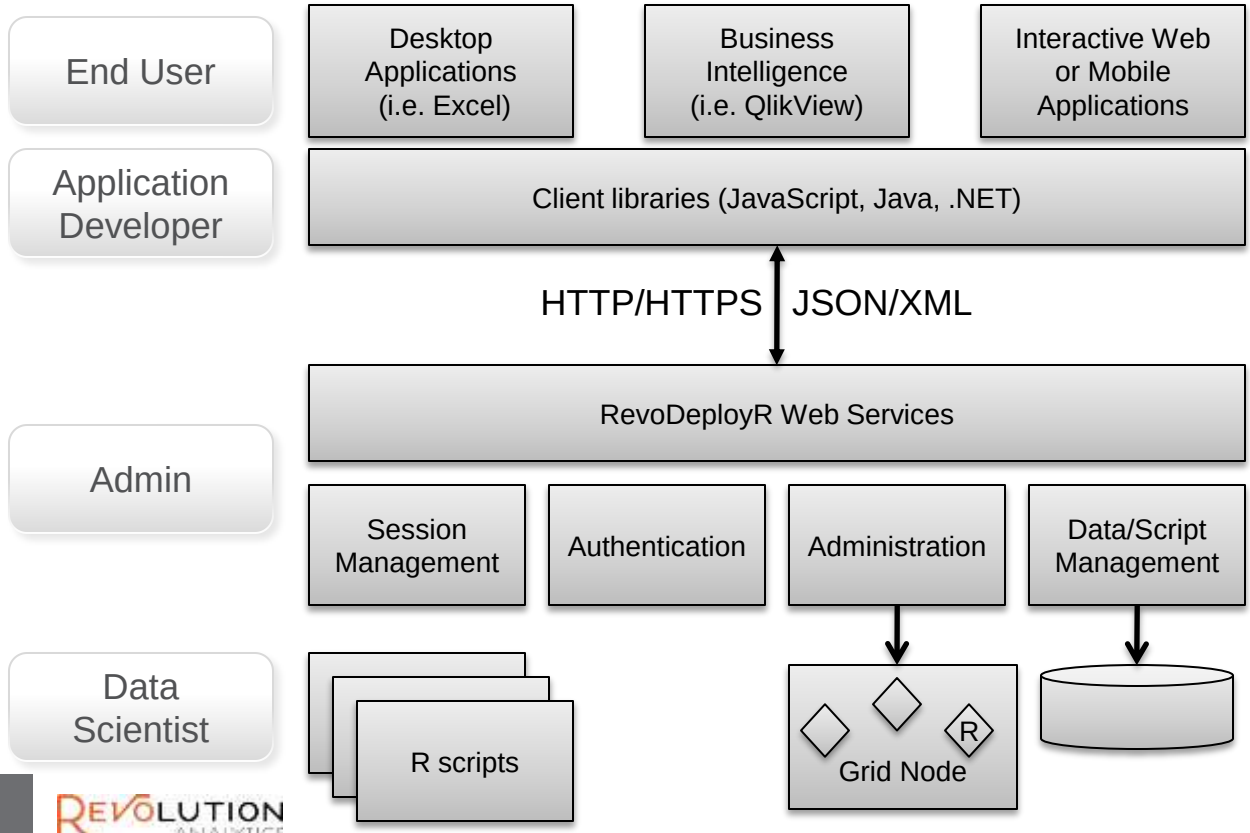
- For the package author:
  - Use package versions available on the chosen date
  - Installs packages local to this project
    - Allows different package versions to be used simultaneously
- For a script collaborator:
  - Automatically installs required packages
    - Detects required packages (no need to manually install!)
  - Uses same package versions as script author to ensure reproducibility

DeployR Open

- **Goal**: embed results from R scripts into real-time applications
- **Problem**:
  - Exposing arbitrary R functions is a security risk
  - Need to handle concurrent R sessions
- **Solution**: DeployR Open
  - Expose only validated R functions
  - Admin console to manage entry points
  - Sanitize inputs via Web Services API
  - Manage and monitor pool of R servers
- Ideal for prototyping integrations
  - For grid-scaling and enterprise authentication, use RRE DeployR



Analytic Application Architecture with DeployR



RHadoop and ParallelR

- Toolkits for data scientists and numerical analysts to create custom parallel and distributed algorithms
- ParallelR: parallel programming for multi-CPU servers and grids
- RHadoop: map-reduce programming in R language

- Mainly useful for “embarrassingly parallel” problems, where parallel components work with small amounts of data

- Big Data Predictive Analytics mostly not embarrassingly parallel
- PhD scientists at Revolution Analytics have pre-built 80+ “parallel external memory algorithms” for Revolution R Enterprise

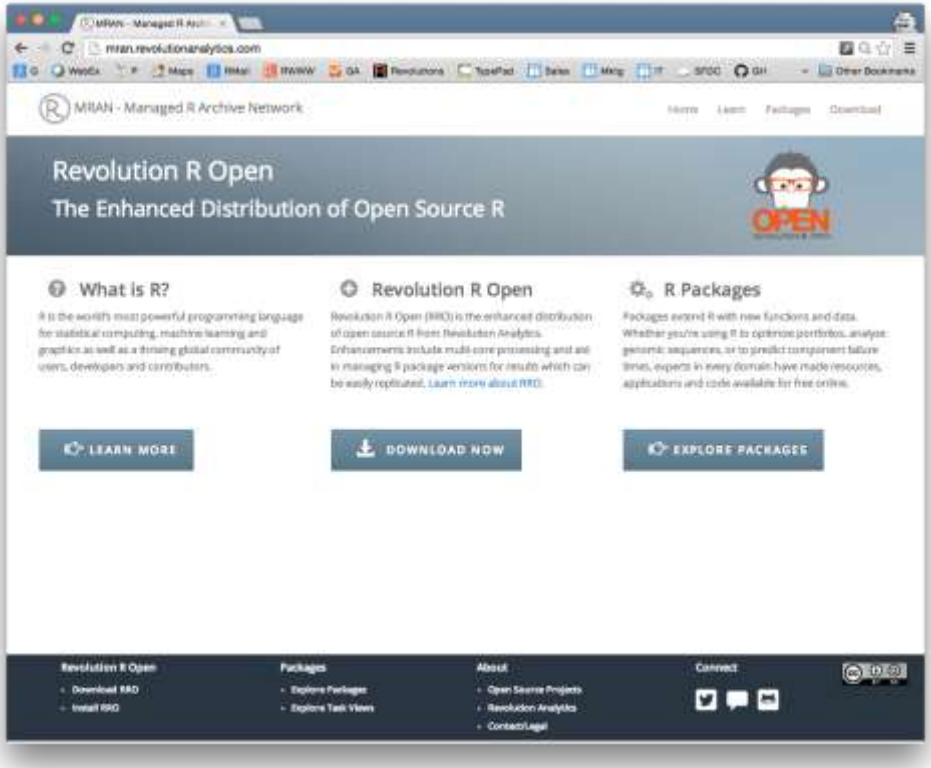
REVOLUTION ANALYTICS

	OPEN REVOLUTION R	PLUS REVOLUTION R	RRE REVOLUTION R ENTERPRISE
	Revolution R Open	**Revolution R Plus**	**Revolution R Enterprise**
R Language Engine with multi-core processing	Included	Supported	Supported
R Reproducibility Toolkit & MRAN	Included	Supported	Supported
ParallelR: Parallel Programming Toolkit		Supported	Supported
RHadoop: R interface to Hadoop MapReduce		Supported	Supported
DeployR Open: Web Services API		Supported	Supported
RRE DeployR – Multi-server, enterprise authentication			Licensed & Supported
RRE ScaleR – Big Data toolkit and PEMAs for R			Licensed & Supported
RRE DistributedR – EDW, Grids, Hadoop			Licensed & Supported
AdviseR Technical Support		**Included**	**Included**
Open Source Assurance		**Included**	**Included**
Revolution Analytics Services (Consulting / Training)	**Available**	**Available**	**Available**

Agenda

- Company Background
- Revolution R Open Overview
- **MRAN**
- Demonstration

MRAN: The Managed R Archive Network

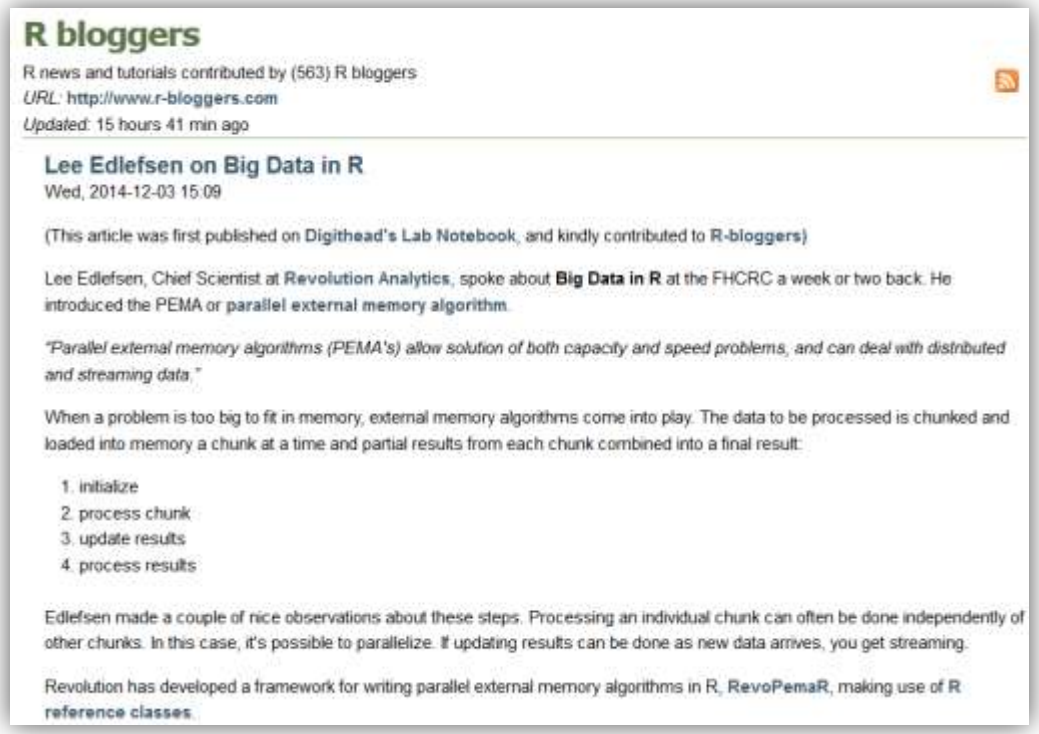


- Download RRO
- Learn about R and RRO
- Daily CRAN snapshots
- Explore Packages
  - and dependencies
- Explore Task Views

Demonstration

1) MRAN – guided tour
2) RRO vs OSR – performance difference
3) Packages
   - version.compare
   - checkpoint
4) DeployR

Some other things we're working on…



REVOLUTION ANALYTICS

Agenda

- Company Background
- Revolution R Open Overview
- MRAN
- **Demonstration**

REVOLUTION ANALYTICS

Thank You

michael.helbraun@revolutionanalytics.com
bill.jacobs@revolutionanalytics.com

blog.revolutionanalytics.com