Extending the R Language
to the Enterprise with Spotfire and TERR

Lou Bajuk-Yorgan
Sr. Dir., Product Management
TIBCO Spotfire
l bajuk@tibco.com
@loubajuk
Extending the Reach of R to the Enterprise

- TIBCO, S+, and embracing R in Spotfire
- Challenges of R for Enterprise applications
- TIBCO Enterprise Runtime for R (TERR)
  - Benefits
  - Spotfire & other applications of TERR
  - TERR performance
- Learn more and try it yourself
Our Journey to TERR

• John Chambers developed the S language at Bell Labs
  – Starting in the mid 70’s
• Insightful (Statsci) founded to commercial S as S+ in 1987
  – The “plus”: statistical libraries, documentation, and support
  – Later focus on commercial users, ease of use, server integration
• R: development begun by Ross Ihaka and Robert Gentleman at University of Auckland in mid 90’s
• Insightful acquired by TIBCO in 2008
  – Spotfire (for Data Discovery and Visualization) acquired in 2007
• Focus shifted to applying Predictive Analytics in Spotfire
  – Step 1: Embrace R
Embracing R

- Spotfire Statistics Server
  - Integration of R & S+ into Spotfire applications
    - Later added SAS® & MATLAB®
  - Leverage the interactive visualizations of Spotfire

- Contribute to the R community

- Well received— but our Enterprise customers need more
  - R provides tremendous benefits to statisticians
  - But large enterprises are often challenged to leverage that value
Spotfire® Enterprise Challenges for Open Source R

• Core R engine struggles with Big Data
  – Customers don’t use R, or reimplement R code in specialized libraries or other languages
  – Lose agility & consistency, delay time to production, lose opportunities

• R was not intended for enterprise usage
  – Not intended for real-time/ high performance & reliability / big data
  – Built as an academic tool for research and teaching
  – Proprietary & Specialized packages still dependent on the core engine

• Free to acquire, but costly to maintain
  – Version incompatibilities, variable quality in packages
  – Lack of enterprise-level technical support
Spotfire® | TIBCO Enterprise Runtime for R (TERR)

- Enterprise-grade implementation of the R language
  - New architecture, developed from the ground up
  - Based on TIBCO’s long history and expertise with S+
  - Designed to be fully compatible with R language
  - Faster and more memory efficient than open source R
  - Fully supported by TIBCO

TERR extends the reach of R in the enterprise
• Develop code in open source R & Deploy on a commercially-supported and robust platform
  – Without the delay and cost of rewriting your code

• Higher performance & superior memory management

• Enables enterprise use of cutting-edge analytics from R
  – TERR is broadly compatible with contributed R packages (1700+)
  – Easily compare multiple analytic approaches, find the hidden insights & make the best decisions
  – Broadly leverage these insights across the organization
The old way of deploying a predictive model:

- Statistician builds predictive model in R
- Programmer translates model into a lower-level language for performance and integration reasons (which can consume weeks/months)
- Introduces limitations (only some types of models can be used)
- Model is deployed to production system to make real-time predictions

TERR eliminates Translation Step:

- Models can be built in open source R, then deployed to TERR (in the native R format) for high-performance real-time prediction
• **Open source R - GPL-licensed**
  – Software vendors are forbidden to embed or redistribute open source R as a part of any commercial closed-source software
  – Leads to poor customer experience for integration and installation

• **TERR - commercially licensed by TIBCO**
  – TIBCO embeds TERR Spotfire
  – Customers and partners can distribute and integrate TERR for embedding or redistribution in their products
Leveraging TERR

**TERR in Spotfire**
- Ad hoc tools and interactive applications powered by advanced analytics
  - Spotfire Analytics platform: interactive visualization & data discovery, easily build and share applications, broad data access, etc.

**TERR in Statistics Services**
- Distributed analytics
  - Managed pools of engines
  - Load balancing, queuing, failover, parallelization, etc.
  - High level APIs for loose integration, data i/o (C#, Java)
  - Central management of analytics, R packages

**Embeddable TERR Engine**
- Custom (tight) integration, batch, existing grids, etc.
  - Faster than R, more robust, better memory management, fully supported
  - Low level APIs for tight integration
  - Integrated into TIBCO products: CEP, Cloud Compute, ...
Easily provide targeted, relevant predictive analytics to business users to improve decision making

- Ensure compliance & proper usage
- Share best practices and consistent workflows
- Get the answer & do “What If?” analyses when needed
- Leverage investments in R, S+, SAS, MATLAB, …

**Powerful Predictive Analytics tools for Spotfire analysts**

- Integrated into Spotfire workflows
- Easily create, evaluate, and share Predictive Models
- Add Forecasts with a single click
- Easily leverage R scripts to enhance analyses

**Benefits of Predictive Analytics to a spectrum of users**
• TERR in Spotfire
  – Sales Forecasting (A simple application)
    • Calling TERR from Spotfire Expression Language
    • Data Functions: using the R language in Spotfire
  – Fraud Detection (A more complex application)
Deliver real-time predictions on whether to extend an offer to a given customer

Apply predictive models in real-time decision making
- Best marketing offer
- Customer churn
- Predictive Maintenance
- Yield optimization

Rapidly develop and iterate TERR models in production
- Respond to changing opportunities and threats
TIBCO Cloud Compute Grid

- **High performance computing on the cloud**
  - Available on TIBCO Cloud Marketplace

- **Robust DataSynapse GridServer architecture**
  - Used by Wall Street to manage 10K’s nodes
  - Java, .NET, and REST APIs (JSON)

- **Perfect for pure computational work**
  - Vastly easier to use for applications like Monte Carlo simulations than Map-Reduce
  - Run complex statistical models multiple orders of magnitude faster than open source R on a single computer
  - Unparalleled scalability without upfront capital investment

- **Easy to get started**
  - Uses your Amazon EC2 account
TERR for individual R users

- **Empower R users**
  - Enterprise platform for the deployment and integration of your work—without having to rewrite it!
- **Contribute to the R community**
  - Sponsor useR conferences, contribute to R Foundation
  - Contribute bug reports and propose fixes to R core
- **Contribute packages to CRAN**
  - As we port from S+ or develop for TERR
  - Develop in Open Source R, Deploy on TERR
- **TERR Developer Edition**
  - Full version of TERR engine for testing code prior to deployment
    - Compatible with RStudio & ESS Emacs
  - Free for non-production use
  - Supported through Community site
Examples of TERR Performance

• Predictions using SVMs from the e1071 package

Summary:
• Small to moderate size data sets
  • Many common operations
  • TERR: 2-10x as fast as OS R
• Larger data sets
  • Common operations (e.g., model scoring) or complex, real-world scripts
  • TERR: 10-100x as fast as OS R
• Individual tests can vary
  • Encourage you to try it yourself and share your results with us
Summary

• TIBCO makes benefits of Advanced Analytics available across the enterprise
  – Ad-hoc Predictive Analytics for Data Scientists
    • Spotfire & TERR
  – Guided Analytic Applications
    • Spotfire & Statistics Services predictive ecosystem, starring TERR
  – Event-driven, real-time analytics
    • Spotfire Event Analytics, TIBCO CEP & TERR

• TERR
  – Enables customers to develop code in open source R, then deploy their R code on commercially-supported and robust platforms—without recoding, saving time & money
  – Enables TIBCO, partners & customers products to tightly & efficiently embed & redistribute R language functionality
• **TERR Community** at TIBCOmmunity.com
  – Resources, FAQs, Forums
  – Details of R coverage
  – Product documentation & download
  – More info at spotfire.tibco.com/terr

• **TERR Developer Edition**
  – Full version of TERR engine for testing code prior to deployment
  – Supported through TIBCOmmunity, **download via tap.tibco.com**

• **TIBCO Cloud Compute Grid**

• Presentations: [http://www.slideshare.net/loubajukyorgan/presentations](http://www.slideshare.net/loubajukyorgan/presentations)

• We want your feedback and input!
  – Real world performance tests
  – Package & R coverage prioritization
  – Via TERR Community, or contact me lbajuk@tibco.com or @loubajuk