

Extending the Reach of R to the Enterprise

TIBCO Enterprise Runtime for R

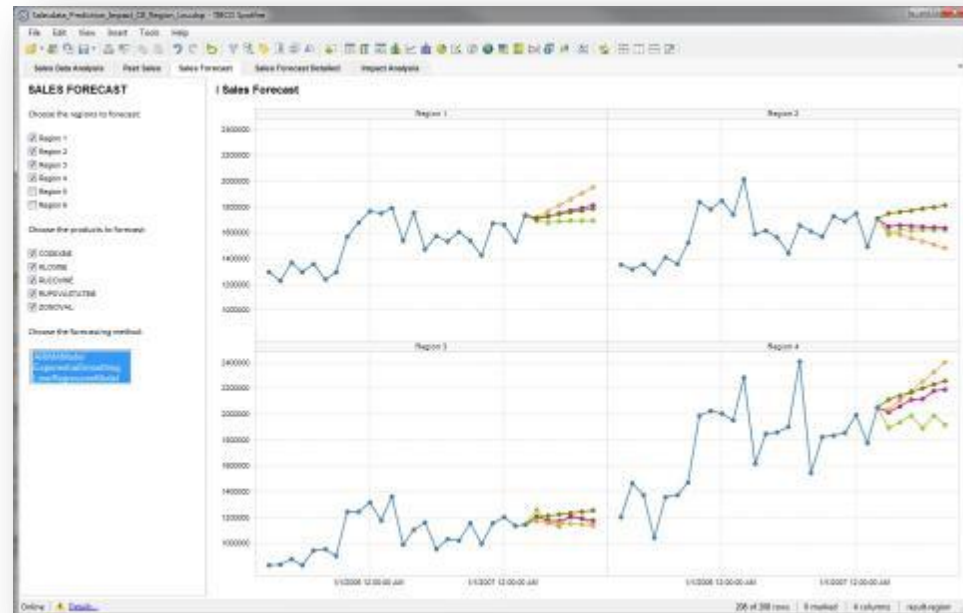
Lou Bajuk-Yorgan, Sr. Director, Product Management
lbajuk@tibco.com

- TIBCO, S+, R and me
- Why (and how) did we embrace R?
- Challenges of R for Enterprise applications
- TIBCO Enterprise Runtime for R (TERR)
- Benefits for organizations (and individuals) who use R
- Examples of TERR integration and performance
- Learn more and try it yourself

- John Chambers developed the S language at Bell Labs
 - Starting in the mid 70's
- Insightful (Mathsoft/ 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
 - *I started using S+ ~1993, joined Insightful in 1996*
- R: development begun by Ross Ihaka and Robert Gentleman at University of Auckland in 1995
- Insightful acquired by TIBCO in 2008
 - Spotfire (for Data Discovery and Visualization) acquired in 2007
- Focus shifted on integration of advanced analytics to enterprise applications
 - Step 1: Embrace R

- Focus of energy and innovation in the statistical community
 - Agility and power of the S/R language
 - Provides a freely available platform for anyone to use
 - Easily and develop and share new analytic methods as packages
 - Hugely successful
 - Many users, many contributed packages
 - Default way that people communicate in statistics
- R provides tremendous benefits to statisticians
 - But large enterprises are often challenged to leverage that value beyond the individual statistician

- TIBCO Spotfire Statistics Services
 - Integration of R & S+ into Spotfire (and other) applications
 - Easily provide targeted, relevant advanced analytics to larger community of users, combined with the interactive visualization of Spotfire
 - Statistician can ensure compliance & proper usage
 - Later integrated with SAS® and MATLAB® as well
- Contribute to the R community
 - Sponsor useR conferences, contribute to R Foundation
 - Contribute bug reports and fixes to R core
 - Contribute packages to community
- Well received—but our Enterprise customers need more



- R 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 built for enterprise usage and integration
 - Built as an academic tool for research and teaching
 - Some R Core members exploring different directions
 - Software vendors attempting to use R in ways it was never intended
- GPL has been great for statisticians, but limits enterprise innovation and investment
 - Viral open source licensing risks IP
 - Only attracts innovations people are willing to give away
 - Large vendors avoid tight integration due to open source concerns
- Free to acquire, but costly to maintain
 - Version incompatibilities, variable quality in packages
 - Lack of enterprise-level technical support
 - Difficult to validate

- Unique, enterprise-quality runtime for the R language
 - Fundamentally different from anything else on the market
 - Developed from the ground up, based on our long history and expertise with S+
 - Brings scalability & stability to agile R language
 - Faster, more robust and more memory-efficient than R
 - TIBCO IP: Not open source/GPL
 - Licensable for embedding and redistribution by partners
 - Brand new architecture, new foundation for future investment
 - Broad support for R functions and hundreds of CRAN packages
 - Ongoing effort to broaden our coverage of R
- Extends the Reach of R to the Enterprise
 - Develop in Open Source R, deploy on TERR
 - Agility of R
 - Rapidly iterate prototyping to production without recoding/retesting
 - Easily integrate predictive analytics consistently across organization
- Initial Focus: Integration with Spotfire platform

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
- Basic data i/o (text files, JDBC, etc.)
- Non GPL

- Not seeking to displace R from statistician's desktops
 - Enterprise platform for the deployment and integration of your work—without having to rewrite it!
- Contribute packages to CRAN
 - As we port from S+ or develop for TERR
 - Supports “Develop in OS R, Deploy on TERR”
 - Now available on CRAN
 - splusTimeSeries, splusTimeDate, sjdbc
 - tibbrConnector
- TERR Developer Edition
 - Full version of TERR engine for testing code prior to deployment
 - Console only—expect you to use your favorite R IDE
 - Perpetual License for free, non-production use
 - Supported through Community site

- TERR vs. R raw performance
- TERR in Spotfire—powering Predictive Analytics tools and applications for Data Discovery
- TERR and real-time event processing

One specific example

- Non-optimal, non-vectorized, real-world R script
- For loop with row by row processing

```
for (i in seq(1,length=nrow(df))) {  
  ...process each customer record..  
}
```

Results

- TERR is ~35x faster for 50K rows, 150x faster for 500K rows
- No code modification required

We are looking for more real-world performance tests!

- On average 2-10x faster than R in microtests

Predictive Modeling Tools

- Regression
- Classification
- Model Diagnostics

Build and shared Spotfire applications with embedded analytics

Power of predictive analytics in Spotfire

- For non-R programmer
- Leverage the interactive visualizations of Spotfire
- Powered by embedded TERR engine

Model Summary

The screenshot displays the Spotfire Model Summary window for a linear regression model. The model name is 'MyHousingModel' and the type is 'Linear Regression'. The model equation is $\text{SalePrice} \sim \text{BedroomCnt} + \text{BathroomCnt} + \text{FinishedSqFt} + \text{LotSizeSqFt}$. A 'Regression Modeling' dialog box is open, showing the response column 'SalePrice' and predictor columns 'PropertyID', 'Zipcode', 'ParcelNumber', 'BedroomCnt', 'BathroomCnt', 'FinishedSqFt', and 'LotSizeSqFt'. The formula expression is $\text{SalePrice} \sim \text{BedroomCnt} + \text{BathroomCnt} + \text{FinishedSqFt} + \text{LotSizeSqFt}$.

Name	Estimate	SSEError	t-value	p-value
(Intercept)	258.78	124.66	2.88	0.04
BedroomCnt	-123.65	50.23	-2.46	0.01
BathroomCnt	219.92	74.85	2.94	0.00
FinishedSqFt	0.11	0.08	1.37	0.17
LotSizeSqFt	0.00	0.00	-0.17	0.87

Diagnostic plots shown include:

- Residuals vs Fit:** A scatter plot of residuals against fitted values, showing a negative correlation.
- Normal Quantile-Quantile:** A Q-Q plot showing the distribution of standardized residuals against theoretical normal quantiles.
- Scale - Location:** A plot of the square root of absolute standardized residuals against fitted values.

- TIBCO Business Events
 - Complex Event Processing system
- Example integration with TERR
 - Deploy R models on TERR engine for real-time scoring in response to complex events
 - Random forest model, scoring online transactions for fraud
- Testing on a small group of servers
 - Total of 26 CPU cores
- Results
 - 2200 transactions scored per second, using random forest model in integrated TERR engine

- TIBCO, S+, R and me
- Why (and how) did we embrace R?
- Challenges of R for Enterprise applications
- TIBCO Enterprise Runtime for R (TERR)
- Benefits for organizations (and individuals) who use R
- Examples of TERR integration and performance
- Learn more and try it yourself

- useR presentation: Memory Management in TERR
 - By Michael Sannella, Friday 11:50
- [TERR Community](#) now available 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](#)
- 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