Extending R to the Enterprise

With TIBCO Spotfire and TERR
Lou Bajuk-Yorgan, Sr. Dir., Product Management, TIBCO
Extending R to the Enterprise

- TIBCO’s unique history with R
- Challenges of R for Enterprise applications
- TIBCO Enterprise Runtime for R (TERR)
- Applications of R/TERR in Analytic and Real Time Applications
- Learn more and try it yourself
TIBCO’s Unique History with R

- S language from Bell Labs
  - Forerunner to S+ & R
  - Developed by John Chambers starting in mid 70s
- Insightful/Statsci founded to commercial S as S+ (1987)
  - 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 on applying Advanced Analytics in Spotfire & TIBCO real-time applications
  - Embraced R language as part of Advanced Analytic Ecosystem
  - Clear customer-feedback on the Open Source R engines’s challenges for enterprise usage
Challenges with R in the Enterprise

• R provides tremendous benefits to statisticians
  – Enterprises are challenged to leverage that value

• Enterprises need:
  – R’s expansive flexibility to develop the most powerful and relevant predictive models
  – Better performance & reliability than open source R
    • Faster, more memory efficient, greater scalability
  – Fully supported environment
  – Agility to use R through the entire prototyping-to-production process

• Vendors & application developers need:
  – An embeddable R engine that is not open-source GPL licensed
  – Alternative to loose integrations, manual configuration and poor user experience
Introducing TERR

TIBCO Enterprise Runtime for R

- Unique, enterprise-grade statistics engine, architected from the ground up by TIBCO
- Based on TIBCO’s long history and expertise with S+
- R language compatibility
- Designed to provide embedded analytics
- A TIBCO licensed & supported product
  - Not GPL, not a repackaging of the Open source R engine
TERR: Enterprise-Grade Platform

- **Performance and Robustness**
  - TERR has better performance and memory management than open-source R
    - Handles much larger data sets in memory; designed and architected for 64-bit platforms
    - Linear, predictable performance as data set sizes increase
    - Architected for future extension to out-of-memory and in-database data

- Fully supported by TIBCO
- Wide range of deployment options to Spotfire and real-time applications

  Deploy from prototyping to production with R language analytics
Examples of TERR Performance

<table>
<thead>
<tr>
<th></th>
<th>OS R</th>
<th>TERR</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model Scoring on 5M rows</td>
<td>107.1 sec</td>
<td>17.5 sec</td>
<td>6.1 x</td>
</tr>
<tr>
<td>Model Fitting on 20M rows</td>
<td>84.2 sec</td>
<td>1 sec</td>
<td>84.2 x</td>
</tr>
</tbody>
</table>

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 users to try it themselves with free Developer’s Edition

Fitting and Scoring
Generalized Linear Models

Predictions using SVMs from the e1071 package
TERR Speeds Analytic Deployment for Competitive Advantage

• Save Time and Cost
• Respond to Opportunities and Threats more quickly

TERR enables customers to develop code in open source R, then deploy their R code on a commercially-supported and robust platform

• Eliminates need to rewrite and retest code in a different environment or using specialized R libraries
• Avoids limiting enterprises to a constrained set of analytic models with limited analytic power
• Wide range of deployment options to TERR embedded in Spotfire and real-time applications
  • Enables easy, seamless reuse and deployment
Developing TERR Analytics

• Develop R language scripts in TERR using Community-standard RStudio IDE
  – Script management, code folding, etc.
  – Debugger
  – Package Development
  – Integration with Source Control

• TERR Developer Edition
  – Full version of TERR engine for testing code prior to deployment
  – Free download for non-production use
  – Supported through Community site
Grid Computing with TERR

- TIBCO Cloud Compute Grid
  - 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
  - Massive scalability without upfront capital investment
TERR in Spotfire

- TERR powers predictive modeling & forecast tools in Spotfire
- Easily enhance Spotfire analyses and applications with R language scripts

- Write R code directly in Spotfire; TERR executes locally or on server
- Manage TERR analytics locally or in Server to reuse across community
- Part of Advanced Analytic Ecosystem (SAS®, MATLAB®, OS R, S+,...)

Deploy TERR-powered applications to the web
Simple Business User Application

Flexible R Scripting for Ad Hoc Analysis
Predicting the Business

Customer Churn:
• Retain your most profitable customers
• Increase upsell, decrease churn

Fraud Detection:
• Reduce losses due to fraudulent transactions

Supply Chain Optimization:
• Anticipate peaks and lulls
• Optimize distribution centers

HR Planning:
• Predict employee attrition and optimize retention
Using the model, TERR generates predictions by store and ranks stores in order of predicted campaign revenue lift. In Spotfire, controls and visualizations are used to find the campaign parameters that maximize campaign revenue lift.

**What If Analysis?**

- Number of Stores in Campaign (ranked by predicted revenue lift): 81
- Campaign Duration (days): 
- Fixed Cost per Store ($): 2039.90
- Cost per Extended Offer ($): 6.20
- Cost per Accepted Offer ($): 1.95
- Offer Acceptance Rate (%): 12.5
- Average gift_cards purchase When Offer is Accepted ($): 20.00

**Predicted Revenue Lift by Store**

**Predicted Cumulative Revenue Lift**

**Stores in Campaign**

<table>
<thead>
<tr>
<th>State</th>
<th>Store_City</th>
<th>Predicted Dollar Minus Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA</td>
<td>San Ramon</td>
<td>1586.14</td>
</tr>
<tr>
<td>CA</td>
<td>Palm Desert</td>
<td>1139.89</td>
</tr>
<tr>
<td>SC</td>
<td>Chapin</td>
<td>1005.21</td>
</tr>
<tr>
<td>AR</td>
<td>Little Rock</td>
<td>970.45</td>
</tr>
<tr>
<td>FL</td>
<td>Naples</td>
<td>925.77</td>
</tr>
<tr>
<td>OR</td>
<td>Bend</td>
<td>910.76</td>
</tr>
<tr>
<td>CA</td>
<td>Sacramento</td>
<td>877.06</td>
</tr>
<tr>
<td>KS</td>
<td>Overland Park</td>
<td>864.58</td>
</tr>
<tr>
<td>CA</td>
<td>Santa Monica</td>
<td>827.07</td>
</tr>
<tr>
<td>CA</td>
<td>Pismo Beach</td>
<td>823.79</td>
</tr>
<tr>
<td>CA</td>
<td>Santa Cruz</td>
<td>822.56</td>
</tr>
<tr>
<td>CA</td>
<td>Davis</td>
<td>814.38</td>
</tr>
<tr>
<td>CA</td>
<td>Roseville</td>
<td>806.90</td>
</tr>
<tr>
<td>CA</td>
<td>Silver Spring</td>
<td>805.24</td>
</tr>
<tr>
<td>MA</td>
<td>Dedham</td>
<td>805.00</td>
</tr>
<tr>
<td>CO</td>
<td>Colorado Springs</td>
<td>804.00</td>
</tr>
<tr>
<td>CA</td>
<td>San Rafael</td>
<td>790.90</td>
</tr>
<tr>
<td>CT</td>
<td>Darien</td>
<td>783.35</td>
</tr>
<tr>
<td>LA</td>
<td>Malibu</td>
<td>774.19</td>
</tr>
<tr>
<td>WI</td>
<td>Madison</td>
<td>750.62</td>
</tr>
<tr>
<td>NJ</td>
<td>Red Bank</td>
<td>750.66</td>
</tr>
<tr>
<td>TN</td>
<td>Franklin</td>
<td>745.57</td>
</tr>
</tbody>
</table>

**Campaign Summary**

- 81 Stores
- 268,174 Offers Extended
- 3,353 Offers Accepted
- $46,002 Revenue Lift
- Avg $568 Lift Per Store
TERR in TIBCO’s Complex Event Processing

• TERR powers real-time advanced analytics to “Fast Data” in TIBCO Streambase and TIBCO Business Events
  – When an event is identified, the CEP application can trigger:
    • An automated business process (e.g., extend a mobile offer to a customer; stop a fraudulent transaction in process)
    • A Spotfire application or dashboard, marrying real-time event data with historical contextual data, providing immediate root cause analytics (Spotfire Event Analytics)
Use Cases: TERR in CEP

Logistics Optimization

• Port Congestion Detection
  – Real time system triggers TERR
  – Analyzes port congestion
  – Recommends reduction of speed if no berths available

• Maritime Abnormality Detection
  – Based on Automatic Identification System info, TERR calculates likelihood of deviation from normal sailing routes
  – Alerts carrier & operator
Use Cases: TERR in CEP

Predictive Maintenance for Oil & Gas

• Oil & Gas Extraction
  – Maintenance downtime and equipment failures are costly
  – Engineers track sensor data to find leading indicators (e.g., temperature, vibration)

• Engineers usually use ad hoc rules on leading indicators
  – R/TERR used to develop predictive models for preventative maintenance
  – Deployed in real-time systems, alert when maintenance recommended
Use Cases: TERR in CEP

Customer Loyalty Analytics

• 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
Use Cases: TERR in CEP

Severe Weather Alerts Tracking for Facilities

- Alert Facilities Managers to Severe Weather Alerts
- Use TIBCO Streambase to monitor alerts in real time
  - TERR queries for current weather alerts
  - Uses sp & rgdal package to compare areas of alerts vs. facility locations
- On alert, update Spotfire visualization with mapped alerts, notify Facilities managers
TERR delivers production-grade R to enterprises

- Flexibility & accuracy of R language
- Enterprise-required performance
- Time-to-market agility

- Seamless, powerful integration in Spotfire for Data Discovery and Analytics Applications
- Event-driven, real-time analytics in TIBCO CEP
- Deployment in TIBCO products, customer & 3rd party applications
• **TERR Community** at TIBCOmmunity.com
  – Resources, Documentation, FAQs, Forums
  – More info at [spotfire.tibco.com/terr](http://spotfire.tibco.com/terr)

• **TERR Developer Edition**
  – Full version of TERR engine for developing & testing code prior to deployment
  – [Download via tap.tibco.com](http://tap.tibco.com)

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