Revolution R Enterprise

**Michele Chambers** – Chief Strategy Officer & VP Product Management @ Revolution Analytics

**Bill Franks** – Chief Analytics Officer @ Teradata
Agenda

- Emerging Big Data Analytic Patterns
- Teradata Overview
- Revolution R Enterprise
- Teradata + Revolution R Enterprise
- Q&A
Polling Question #1:

- Rate your current use/ knowledge of R:
  - Actively Using R Now
  - Are planning on using < 6 mo.
  - Evaluation R now
  - Not at all
Michele Chambers
Chief Strategy Officer & VP
Product Management

Revolution Analytics

@MCanalytics
1st Generation Predictive Analytics
Today’s Challenge: Accelerating Business Cadence

**Changing Business Environment**
- Fact Based Decisions Require More Data
- Need to Understand Tradeoffs and Best Course of Action
- Predictive Models Need to Continually Deliver Lift
- Reduced Shelf Life for Predictive Models

**Faster Time to Value**
- Reduce Analytic Cycle Time
- Build & Deploy Models Faster
- Eliminate Time Consuming Data Movements

**Rapid Customer Facing Decisions**
- Score More Frequently
- Need to Make Best Decision in Real Time
2nd Generation Predictive Analytics

- Big Data
- Machine Learning
- Quick to Fail
- Lift
Typical Challenges Our Customers Face

**Big Data**
- Many new data sources
- Data variety & velocity
- Fine grain control
- Data movement, memory limits

**Complex Computation**
- Experimentation
- Many small models
- Ensemble models
- Simulation

**Enterprise Readiness**
- Heterogeneous landscape
- Write once, deploy anywhere
- Skill shortage
- Production support

**Production Efficiency**
- Shorter model shelf life
- Volume of Models
- Long end-to-end cycle time
- Pace of decision accelerated
Big Data Big Analytics is different
Polling Questions #2

- What Analytics Challenges Are You Facing
- <Choose All That Apply>
  - Data Too Large for Current Solutions
  - Lack Timely Access To Needed Data
  - Slow Model Development or Execution
  - Inconsistent Solutions Across Platforms
  - Acceptability of Open Source
  - Cost of Solutions Too High
  - None of the Above.
Analytic Reference Architecture

- **Decision**
  - Analytic Applications

- **Integration**
  - Middleware

- **Analytics**
  - Analytic Tools
    - Hadoop
    - Data Warehouse
    - Other Data Sources
Architectural Approaches for Analytics

Beside Architecture

Decision
Integration
Analytics

Analytic Applications
Middleware
Analytic Tools
Local Data Mart

Data
Hadoop
Data Warehouse
Other Data Sources

Inside Architecture

Decision
Integration
Data & Analytics

Analytic Applications
Middleware
Analytic Tools

Hadoop
Data Warehouse
Other Data Sources
Pros & Cons of Approaches

**Beside Architecture**
- Analytic workflow tasks performed in a separate analytics environment outside of the source database
  - **Pros:** Segregates analytic workload
  - **Cons:** Doesn’t leverage powerful production for transformations, introduces scoring latencies

**Inside Architecture**
- Analytics workflow tasks performed inside the source database with embedded analytics
  - **Pros:** Eliminates data movement, reduces model latency, allows exploration of all data
  - **Cons:** IT governance on production, potential new skills

**Hybrid Architecture**
- Some analytic workflow tasks performed inside the source database & others performed in a separate analytics environment
  - **Pros:** Leverages strengths of each architecture
  - **Cons:** Maintain multiple environments
Building & Deploying Analytic Models

**Beside Architecture**
- Data Sources
- Analytics

**Inside Architecture**
- Data & Analytics
- Hadoop, DW, Other

**Hybrid Architecture**
- Data
- Analytic Tools
- Data Sources

Legend:
- 1 Data prep / marshaling
- 2 Model build
- 3 Model recode / PMML
- 4 Model deploy
- 5 Update data
The Teradata Unified Data Architecture™
Any User, Any Data, Any Analysis
Bill Franks
Chief Analytics Officer
Teradata
@billfranksga
Teradata Highlights

- 2,400+ Customers in 77 Countries
- 10,000+ Employees
- A Top 10 Public U.S. Software Company
- Ranked 2nd Best Company in America – Motley Fool
- World’s Most Ethical Companies – Ethisphere Institute
- Leader in Data Analytics
- Leader in Integrated Marketing Management

Source: Gartner Magic Quadrant for Data Warehouse Database Management Systems (February 2013)
Teradata is the global leader in analytic data platforms, business applications, and consulting services that help organizations become more competitive by increasing the value of their data and customer relationships.

- Data Warehousing
- Workload-specific Platforms
- Discovery Platforms
- Integrated Marketing Management
- Consulting/Support Services
- Partnerships with key analytic technology providers and system integrators
Why Is Teradata Different?

Server Based vs. In-Database Architectures

Desktop and Server Analytic Architecture

In-Database Analytic Architecture

Results

Sample Data

Yes

NO

Income>$40K

Debt<10% of Income Debt=0%

Good

Credit

Risks

Bad

Credit

Risks

Exponential Performance Improvement

In-DBS Request

Revolution Analytics

Teradata

Why Is Teradata Different?

Server Based vs. In-Database Architectures

Desktop and Server Analytic Architecture

In-Database Analytic Architecture

Results

Sample Data

Yes

NO

Income>$40K

Debt<10% of Income Debt=0%

Good

Credit

Risks

Bad

Credit

Risks

Exponential Performance Improvement

In-DBS Request

Revolution Analytics

Teradata
Raising The Bar With Teradata In-Database Analytics

1998

Teradata Warehouse Miner
Pioneer of in-DBS mining

2007

Partnership Centric
Collaborated with partners to enable in-DBS processing

2010

Data Labs
Simplified In-DBS self-service Analytics

2011

Aster Data Acquisition
Pioneer of big data analytics

2013

Expanding Portfolio
Premier of Revolution Analytics parallel R in-DBS
Teradata Data Lab

Self-Service Analytic Sandboxes With Governance

- Data Lab(s) inside system to easily join to production data
- Load experimental, untested data from external sources
- Rapid prototyping, exploratory and experimentation analysis
- An architecture design that enables governance
  - Works within your current data warehouse environment
- Teradata Data Lab portlets for IT and business analyst
Teradata Workload-Specific Platforms

*Fit Broad Range of Customer Needs*

<table>
<thead>
<tr>
<th>Workloads</th>
<th>Data Mart Appliance</th>
<th>Extreme Data Appliance</th>
<th>Data Warehouse Appliance</th>
<th>Active Enterprise Data Warehouse</th>
<th>Appliance for Hadoop</th>
<th>Aster Big Analytics Appliance</th>
<th>SAS High-Performance Analytics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test / Development</td>
<td>Analytical Archive,</td>
<td>Strategic Intelligence,</td>
<td>Strategic / Operational</td>
<td>Appliance for</td>
<td>Discovery Platform for Big</td>
<td>Dedicated appliance for SAS</td>
</tr>
<tr>
<td></td>
<td>or Smaller Data Marts</td>
<td>Deep Dive Analytics</td>
<td>Decision Support</td>
<td>Operational Intelligence,</td>
<td>Storing, Capturing</td>
<td>Data Analytics with</td>
<td>high-performance analytic</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>System, Fast Scan</td>
<td>Real-Time Update, Active</td>
<td>and Refining Data.</td>
<td>embedded SQL MapReduce for</td>
<td>model development</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>workloads</td>
<td>Hortonworks HDP 1.1</td>
<td>new data types and sources</td>
<td></td>
</tr>
</tbody>
</table>

*WE CAN COMPETE ON ANY WORKLOAD, AT ANY PRICE POINT, FOR ANY CUSTOMER*
Revolution R Enterprise

- What is Revolution R Enterprise?
- How well does Revolution R Enterprise play in the Big Data zoo?
Revolution R Enterprise

**RRE** is....

the only big data big analytics platform based on open source R

the defacto statistical computing language for modern analytics

- High Performance, Scalable Analytics
- Portable Across Enterprise Platforms
- Easier to Build & Deploy Analytics
R is open source and drives analytic innovation but...has some limitations for Enterprises

<table>
<thead>
<tr>
<th>Big Data</th>
<th>In-memory bound</th>
<th>Hybrid memory &amp; disk scalability</th>
<th>Operates on bigger volumes &amp; factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed of Analysis</td>
<td>Single threaded</td>
<td>Parallel threading</td>
<td>Shrinks analysis time</td>
</tr>
<tr>
<td>Enterprise Readiness</td>
<td>Community support</td>
<td>Commercial support</td>
<td>Delivers full service production support</td>
</tr>
<tr>
<td>Analytic Breadth &amp; Depth</td>
<td>5000+ innovative analytic packages</td>
<td>Leverage open source packages plus Big Data ready packages</td>
<td>Supercharges R</td>
</tr>
<tr>
<td>Commercial Viability</td>
<td>Risk of deployment of open source</td>
<td>Commercial license</td>
<td>Eliminate risk with open source</td>
</tr>
</tbody>
</table>
Introducing Revolution R Enterprise (RRE)
*The Big Data Big Analytics Platform*

- Big Data Big Analytics Ready
  - Enterprise readiness
  - High performance analytics
  - Multi-platform architecture
  - Data source integration
  - Development tools
  - Deployment tools
The Platform Step by Step: 
*R Capabilities*

**R+CRAN**
- Open source R interpreter
  - UPDATED R 3.0.2
- Freely-available R algorithms
- Algorithms callable by RevoR
- Embeddable in R scripts
- 100% Compatible with existing R scripts, functions and packages

**RevoR**
- Performance enhanced R interpreter
- Based on open source R
- Adds high-performance math

**Available On:**
- Platform™ LSF™ Linux®
- Microsoft® HPC Clusters
- Windows® & Linux Servers
- Windows & Linux Workstations
- **NEW Teradata Database**
  - IBM® Netezza®
  - IBM BigInsights™
  - Cloudera Hadoop®
  - Hortonworks Hadoop
  - Intel® Hadoop
The Platform Step by Step: 
*Parallelization & Data Sourcing*

**ConnectR**
- High-speed & direct connectors

**Available for:**
- High-performance XDF
- SAS, SPSS, delimited & fixed format text data files
- Hadoop HDFS (text & XDF)
- Teradata Database & Aster
- EDWs and ADWs
- ODBC

**ScaleR**
- Ready-to-Use high-performance big data big analytics
- Fully-parallelized analytics
- Data prep & data distillation
- Descriptive statistics & statistical tests
- Correlation & covariance matrices
- Predictive Models – linear, logistic, GLM
- Machine learning
- Monte Carlo simulation
- **NEW** Tools for distributing customized algorithms across nodes

**DistributedR**
- Distributed computing framework
- Delivers portability across platforms

**Available on:**
- Windows Servers
- Red Hat and **NEW** SuSE Linux Servers
- IBM Platform LSF Linux
- Microsoft HPC Clusters
- **NEW** Teradata Database
- **NEW** Cloudera Hadoop
- **NEW** Hortonworks Hadoop
The Platform Step by Step: Tools & Deployment

**DevelopR**
- Integrated development environment for R
- Visual ‘step-into’ debugger

**Available on:**
- Windows

**DeployR**
- Web services software development kit for integration analytics via Java, JavaScript or .NET APIs
- Integrates R into application infrastructures

**Capabilities:**
- Invokes R Scripts from web services calls
- RESTful interface for easy integration
- Works with web & mobile apps, leading BI & Visualization tools and business rules engines
Write Once. **Deploy Anywhere.**

**DESIGNED FOR SCALE, PORTABILITY & PERFORMANCE**

- **Hadoop**
- **EDW**
- **Clustered Systems**
- **Workstations & Servers**
- **In the Cloud**

**In the Cloud**
- Amazon AWS

**Clustered Systems**
- IBM Platform LSF
- Microsoft HPC

**Workstations & Servers**
- Desktop Server

**EDW**
- NEW Teradata Database
- IBM Netezza

**In the Cloud**
- Hortonworks
- Cloudera

**Clustered Systems**
- IBM Platform LSF
- Microsoft HPC

**DevelopR**
- R+CRAN
- RevoR

**RRE**
- ScaleR
- DistributedR

**Write Once. Deploy Anywhere.**

**DESIGNED FOR SCALE, PORTABILITY & PERFORMANCE**
Revolution R Enterprise
*Propels Enterprises into the Future*

- **Analytics**
  - RRE
  - Hadoop
  - Other Data Sources

- **Integration**
  - Middleware

- **Decision**
  - Analytic Applications
One Big Data Big Analytics Platform

Multiple Architectures

Beside Architecture

Inside Architecture

Analytic Applications

Decision

Integration

Analytics

Data & Analytics

Data

DeployR

DevelopR

RevoR & DistributedR & ScaleR

ConnectR

Hadoop

TERADATA

Other Data Sources

DeployR

RevoR & DistributedR & ScaleR & DevelopR

Hadoop

TERADATA

Other Data Sources
Revolution R Enterprise Powers
*Write Once, Deploy Anywhere*

**Bottom Line: Save Time, Save Money, Get Insights Faster**
- Direct connectors access data without data movement
- Push down analyzing data without movement
- Use same R script on any platform without recoding
- Use right platform for the job!
The Power of Revolution R Enterprise

Performance & Scalability

- In-Database Execution: ScaleR
- In-Hadoop Execution: ScaleR
- Parallelized User Code: ScaleR
- Parallelized Algorithms: ScaleR
- Grid Processing: DistributedR
- Multi-Threaded Execution: DistributedR
- Memory Management: DistributedR
- Fast Math Libraries: RevoR
- R + CRAN: Open Source

- Moves computation to data
- Moves computation to data
- Leverage CRAN
- Labor saving power
- Maximizes computation
- Powerful divide & conquer
- Effective memory utilization
- 3-50X faster
- Leverage latest innovation
RRE in Teradata Database

Revolution R Enterprise:

Teradata Database

ODBC

Web Services

Desktop & Enterprise Applications
How Does It Work? RRE in Teradata Database

Teradata Database

AMPS

Database Nodes

Hybrid Storage

Parse Engine

External Stored Procedure

Revolution R Enterprise

Desktops & Servers

Desktop & Enterprise Applications

ODBC

Web Services

Table Operator

Table Operator

Table Operator

Table Operator

Table Operator

Table Operator
Data Prep, Distillation & Descriptive Analytics

R Data Step

- Data import – Delimited, Fixed, SAS, SPSS, OBDC
- Variable creation & transformation
- Recode variables
- Factor variables
- Missing value handling
- Sort
- Merge
- Split
- Aggregate by category (means, sums)

Descriptive Statistics

- Min / Max
- Mean
- Median (approx.)
- Quantiles (approx.)
- Standard Deviation
- Variance
- Correlation
- Covariance
- Sum of Squares (cross product matrix for set variables)
- Pairwise Cross tabs
- Risk Ratio & Odds Ratio
- Cross-Tabulation of Data (standard tables & long form)
- Marginal Summaries of Cross Tabulations

Statistical Tests

- Chi Square Test
- Kendall Rank Correlation
- Fisher’s Exact Test
- Student’s t-Test
- Subsample (observations & variables)
- Random Sampling
## Statistical Modeling

### Predictive Models
- Sum of Squares (cross product matrix for set variables)
- Multiple Linear Regression
- Generalized Linear Models (GLM) - All exponential family distributions: binomial, Gaussian, inverse Gaussian, Poisson, Tweedie. Standard link functions including: cauchit, identity, log, logit, probit. User defined distributions & link functions.
- Covariance & Correlation Matrices
- Logistic Regression
- Classification & Regression Trees
- Predictions/scoring for models
- Residuals for all models

### Data Visualization
- Histogram
- Line Plot
- Scatter Plot
- Lorenz Curve
- ROC Curves (actual data and predicted values)
- **NEW** Tree Visualization

### Variable Selection
- Stepwise Regression (linear, **NEW** logistic and **NEW** GLM)

### Simulation
- Monte Carlo

## Machine Learning

### Cluster Analysis
- **K-Means**

### Classification
- Decision Trees
- **NEW** Decision Forests

### Deployment
- **NEW** PMML Export
R + Revolution R Enterprise

*Unequaled Big Data Big Analytics*

- Deploy Analytics
  - Web, Mobile, Data Visualization, BI
- Big Data Distributed Analytics
- Open Source Analytics
- Performance Enhanced R
- R
- Performance Enhanced R
- Big Data Distributed Analytics
- Open Source Analytics
- R

---

40
Integrating Enterprise Apps

Revolution R Enterprise: Deploy R

Teradata Database

Web Services

Desktop & Enterprise Applications

ODBC

Database Nodes

Hybrid Storage
Decision Layer: Examples on-demand analytics

On-demand sales forecasting

Real-time social media sentiment analysis

Leveraging the power of R from Microsoft tools
RRE 7.0: Alteryx Integration

- Comparable to SAS Enterprise Miner, IBM SPSS Modeler (aka Clementine)
- Easy way to create analytic models with visual based analytic flows

R analytics

Read any data source

Create analytic flow
RRE 7.0: Tableau Accelerator

- Makes analytics consumable to end users
- Tableau integration is via .NET (creates a Tableau document which is read by Tableau)
RRE 7.0: MS Excel Accelerator

- Makes analytics consumable to most pervasive BI tool
- Excel integration is .NET

R analytics

R visualization
Leveraging Teradata + Revolution R Enterprise
Teradata Database, DistributedR & ScaleR

- Scalable R-Based In-Database Transformation
  - High-Performance, Parallel R-Based Data Transformation
  - Augment Traditional ETL, ELT and SQL Methods
  - Free R Users To Load and Transform Data Independently

- Fast, In-database Exploration & Model Development
  - Vastly Accelerate Model Development, Validation and Re-Training
  - Eliminate Data Movement Penalties & Duplication Costs

- Embrace In-database Machine Learning
  - Next Generation Big Data Big Analytics
  - Scalable To Meet “All the Data All The Time” Requirements

- Integrate the Enterprise With In-Database Scoring
  - Make On-Demand Analytics a Reality For All Users
Polling Questions #3

- How would you consider applying Teradata + Revolution R Enterprise?
  - <click all that apply>
  - Operations
  - Risk Management
  - Customer Analytics
How is RRE Used?

Discovering Patterns with Big Data
- Customer segmentation
- Market basket analysis
- Social networking analysis
- Fraud detection
- Marketing attribution
- Sentiment analysis
- and much more

Building Models Efficiently
- Credit risk
- Customer churn
- Propensity to buy
- Market risk
- Operational risk
- and much more

Flexibly Deploying Models to Consumers
- Customer lifetime value
- Pricing optimization
- Recommendation engines
- and much more
Why Customers Buy Teradata & Revolution R Enterprise

- Powers R for the Enterprise
- Enabling Analytics Across the UDA
- Supercharge R with Teradata Scale
- Lower Cost & Risk of Big Analytics
Big Data Big Analytics Software
Innovation, performance and value for today's data architecture.

WHAT IS R?
Open Source Data Science

R IN BUSINESS
Your Competitive Edge

RRE SOFTWARE
Big Data Big Analytics

Top News, Coming Events and the Latest Information for You

- ALTERNATING HORIZONTAL AND VERTICAL ACCORDI NG TO CALLED AND SCALABLE PREDICTIVE ANALYTICS TO HADOOP...
- REVOLUTION SHIPS RRE7 TO MASH WITH HADOOP...
- DRIVING VALUE FROM BIG DATA WITH REVOLUTION R ENTERPRISE AND HADOOP...
- NEW REVOLUTION ANALYTICS RRE PLATFORM SUPERCHARGES THE NEXT GENERATION ENTERPRISE...
- WHAT'S NEW IN REVOLUTION R ENTERPRISE...
Revolution R Enterprise
BIG DATA BIG ANALYTICS WITH R

Discover the fastest, most cost effective, enterprise-class, big data big analytics software available. With high-performance advanced analytics, Revolution R Enterprise unlocks the insight hidden in Big Data with the power of the R language, the #1 choice of today’s data scientists.

LEARN MORE  GET RRE
Revolution Analytics

Revolution Analytics provides a scalable framework of software and services which enables organizations to take advantage of the powerful R language along with modern analytics innovations to make more precise data-driven assessments. The company brings high performance, productivity, and enterprise readiness to R, the most powerful statistics language in the world used by more than 2 million people. Revolution R Enterprise is designed to meet the production needs of large organizations in industries such as finance, life sciences, retail, manufacturing, and media. Revolution fosters the continued growth of the R community and sponsors the Inside-R community site, user groups worldwide, and offers free licenses of Revolution R Enterprise to academia.

Benefits of Partnership

Revolution Analytics enables Teradata clients to take advantage of the "Big Analytics" revolution with a modern analytics framework that tames the complexity of "Big Data." Revolution Analytics’ Revolution R Enterprise delivers performance, scalability, stability and governance along with the innovativeness of R.

Revolution R Enterprise

Revolution R Enterprise is production-grade analytics software built upon the powerful open source R statistics

Related Resources

Recommended
- Teradata and Revolution Analytics
- White Paper: Teradata and Revolution Analytics: For the Big Data Era, An Analytics Revolution
Contact Us

First Name:  

Last Name:  

Revolution Analytics on Twitter
Follow us on Twitter!
Any
Questions?
Thank you.