Your Trial, Your Data

Visual Analytics – Register for Trial
• Smart data exploration with self-services analytics makes this product usable for anyone. Interactive reporting makes it collaborative. Scalability and governance make it fit the needs of your organization, no matter the size.

Visual Statistics – Register for Trial
• Multiple users can explore and visualize data, then interactively create and refine descriptive and predictive models. Distributed, in-memory processing reduces model development time so you can run complex analytic computations – and get precise results – in minutes.
WHY OPEN SOURCE?

Why the drive to open source?

- Cost effective – considering total cost of ownership
- Flexible – customers can “build anything”
- Immediate access & easy to get started
- Latest technology and latest algorithms
- Strong community and online support
- Many new data scientists learn in open source

So why use SAS to extend open source?
SAS can augment open source
• Increase productivity
• Leverage your assets, people and platforms
• Bring the power of SAS to open source
• Create deployable analytics
• Goal is to ‘embrace’ and ‘extend’
SAS embraces open source for Data Prep
• Open source and SAS work well for Discovery and Development
• SAS can extend open source
  • inventory, register and manage models
  • deploy and execute models in Hadoop and in database
  • enhance models and provide monitoring and reporting
THE ANALYTIC LIFECYCLE

Discovery & Development of Analytics

- Ability to connect to Hadoop
- Run natively in Hadoop
- Minimize data movement

Deployment & Execution of Analytics

- Optimized engine to access Hadoop
- Embedded engine so Hadoop can run SAS

Enterprise Wish List

How SAS Embraces…
HADOOP AS PROCESSING ENGINE

- Use Hadoop as the horsepower for analytics
- Run SAS in Hadoop - no data movement
- Expose Hadoop data to more people through a range of interfaces
- Predictive analytics and machine learning
- SAS for Model Deployment / Scoring
Enterprise Wish List

- A way for users to interact with Hadoop
- Ability to create analytic views and tables
- Ability to assess data quality

How SAS embraces…

- A business user interface to facilitate:
  - Querying Hadoop
  - Adding data
  - Profiling data
  - Cleansing data
  - Transforming data
- With no data movement
SELF SERVE ACCESS TO HADOOP

What directive do you want to perform?

- Browse Tables: Browse tables or open a table to see its contents.
- Query or Join Data In...: Query tables in a file or from multiple tables.
- Saved Directories: Open a previously saved directory for a new query or edit.
- Sort and De-Duplicate: Query and sort a file, and de-duplicate the data to an existing or new table.
- Run a SAS Program: Run a program that manipulates data in a database or Hadoop table.
- Load Data to Hadoop: Copy data from a source and load it into a Hadoop file.
- Load Data to SAS: Copy data from a file and load it into a SAS database or Hadoop table.
- Profile Data: Display analysis of the data in files.
- Saved Profile Reports: Display previously generated profile reports.

Business user UI

Create Trusted Data

EMBRACE

Profile Data
THE ANALYTIC LIFECYCLE

Discovery & Development of Analytics

- PREPARE DATA
- EXPLORE
- MODEL
- INVENTORY
- EXECUTE
- MONITOR

Enterprise Wish List

- Best possible analytics
- Flexibility of tools
- Productivity
- Greater insights = models
- Trusted models

Deployment & Execution of Analytics

How SAS Extends…

- A variety of options to develop models
- Allows data scientist to code in language of choice
- Ability to scale to any data volume
- Handle complex graphics
SAS FROM R • EXTEND
USE SAS TO INTEGRATE R

Why?
• Model comparison
• Leverage R for new algorithms
• Generate score code
• Deploy R models
PRODUCTIVITY

SAS Models (4)
- LASSO
- Random Forest
- Decision Trees
- Decision Forest

Gradient Boost
- Gradient Boost

Open Source (2)
- Open Source

Compare 7 models
Choose champion
Inventory Model
Generate score code
Deploy in database/Hadoop

What if you coded this?
TRAIN a random forest model on customer transaction data to predict which ones can be expected to be repeat customers.

```python
trainForest(data = "shoptrain", numTrees = 250, numVarsToTry = 20)
```

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>1.150</th>
<th>1.163</th>
<th>0.216</th>
</tr>
</thead>
<tbody>
<tr>
<td>240</td>
<td>2120</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>130950</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Loss Reduction Variable Importance**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number of Rules</th>
<th>Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITEM_FACTOR20</td>
<td>257</td>
<td>0.004025</td>
</tr>
<tr>
<td>ITEM_FACTOR18</td>
<td>208</td>
<td>0.004304</td>
</tr>
</tbody>
</table>
The Analytic Lifecycle

**Discovery & Development of Analytics**
- Prepare data
- Explore
- Model

**Deployment & Execution of Analytics**
- Inventory
- Execute
- Monitor

**Enterprise Wish List**
- Model management platform
- Inventory ALL models
- Know who’s working on what
- Ability to deploy models
- Auditable models

**How SAS Extends…**
- Central model management platform
- Repository for SAS models and open source (R, Python, PMML)
- Model history
- Version control
- Model and data lineage
- Model governance
MODEL INVENTORY

Model inventory and search

SAS and Open Source models

Model Metadata

EXTEND
Enterprise Wish List

- Deployable analytics
- Automation
- Faster time to model execution
- In Hadoop/database model execution

How SAS Extends…

- Model execution platform
- Execute models as database functions
- No language conversion
- Purpose built model execution engines
MODEL EXECUTION

Model Publishing and automation

Model Score Code Creation

In Hadoop/database deployment
How SAS Extends…

- Model performance platform to keep models “fresh”
- Compare multiple models at once
- Assess model accuracy (Lift, ROC, K-S)
- Champion/challenger modeling
- Model retraining including open source

Enterprise Wish List

- Best possible models
- Model tournaments
- Visibility into performance
- Easy retraining
- Champion/challenger modelling
MODEL PERFORMANCE

- Monitor data drift
- Retrain models
- Model performance reports
- Model comparisons
SAS VIYA

SUPPORTING CURRENT INDUSTRY TRENDS

Elastic
Scalable
Charge-back capable
Micro-services architecture

Easy installs
Backward compatible
Analytics lifecycle support
Advanced machine learning

Multi-threaded hyper-computing
Memory spillover
Integrated solutions
End-to-end

RESTful API’s
‘Any data, any platform’
Python, Java, Lua support
Plug n’ play

Copyright © 2016, SAS Institute Inc. All rights reserved.
SAS AND OPEN SOURCE

EMBRACE
open source by including it and leveraging it where we can

EXTEND
open source by improving its interoperability and utility for the enterprise
THANK YOU
MATT MALCZEWSKI
MATT.MALCZEWSKI@SAS.COM
FOR MORE INFORMATION

Empowering the SAS/IML user with the functionality of R

Documentation: IML User’s Guide - Calling Functions in the R Language

Video: Calling R Procedures from SAS/IML® Software
https://www.youtube.com/watch?v=rUaTTre24kl

Video: SAS/IML and R: Using Them Together
https://www.youtube.com/watch?v=nmRQ3MtGkG6A

Blogs: The DO Loop – R tags
http://blogs.sas.com/content/iml/tag/r/

Paper (p 14-17): Rediscovering SAS/IML® Software: Modern Data Analysis for the Practicing Statistician

Article: Versions of R that are supported by SAS/IML
http://blogs.sas.com/content/iml/2013/09/16/what-versions-of-r-are-supported-by-sas.html
FOR MORE INFORMATION - EXTENDING R

Video: Using R in SAS Enterprise Miner
https://www.youtube.com/watch?v=TbXo0xQCqDw

Blogs: Spectral Clustering in SAS® Enterprise Miner™ Using Open Source Integration Node
https://communities.sas.com/docs/DOC-8011

Blogs: How to execute a Python script in SAS® Enterprise Miner™
https://communities.sas.com/docs/DOC-10832

Blogs: Open Source Integration Using the Base SAS Java Object
https://communities.sas.com/docs/DOC-10746

Article: The Open Source Integration node installation cheat sheet
https://communities.sas.com/docs/DOC-9988

Usage Notes:
FOR MORE INFORMATION MATERIALS ON GITHUB

Sas integration and sample code
Integration with R, Python
https://github.com/sassoftware/enlighten-integration

Integration with Jupyter Notebook and Python
https://github.com/sassoftware/sas_kernel
https://github.com/sassoftware/saspy

Sample codes of SAS Machine Learning methods
https://github.com/sassoftware/enlighten-apply

SAS Enterprise Miner process flow diagrams
https://github.com/sassoftware/dm-flow