SAS AND OPEN SOURCE

MATT MALCZEWSKI, SAS CANADA

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.

ACKNOWLEDGEMENTS

TAMARA DULL, SAS BEST PRACTICES
STEVE HOLDER, NATIONAL ANALYTICS LEAD, SAS CANADA

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

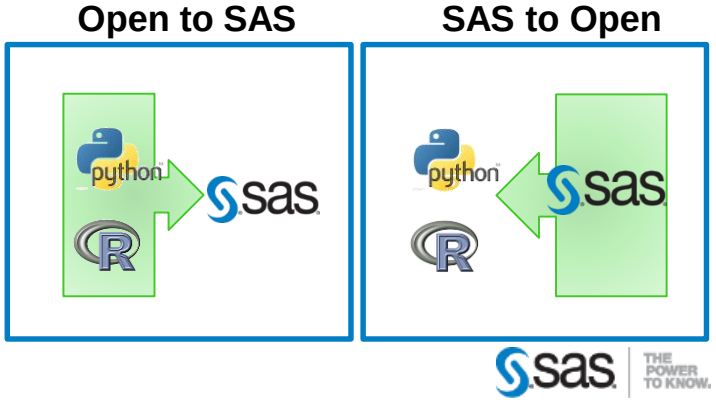
So why use SAS to extend open source?

SAS AS AN ENHANCEMENT

AND

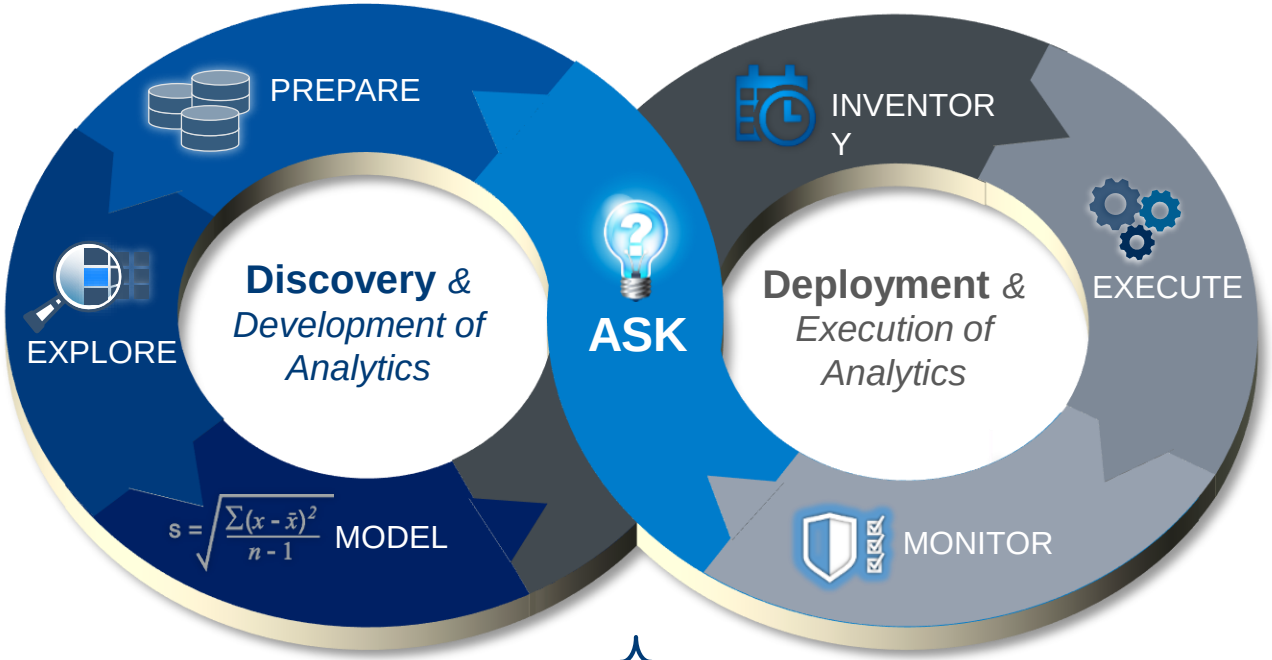
SAS can augment open source

- 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’



THE ANALYTIC LIFECYCLE

Lots of Data
New Data
Experimentation
Fail Fast
Test & Learn
Interactive
Iterative
Innovation
Flexibility
Data Science



Regulated
Automated
Governed
Embed
Reliable
Decisions
Consistent
Documented
Actions
IT

THE ANALYTIC LIFECYCLE: SAS AND OPEN SOURCE

Discovery *& Development of Analytics* **Deployment** *& Execution of Analytics*

- 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

EMBRACE + ***EXTEND***

THE ANALYTIC LIFECYCLE

EMBRACE

Discovery *& Development of Analytics* **Deployment** *& Execution of Analytics*

Enterprise Wish List

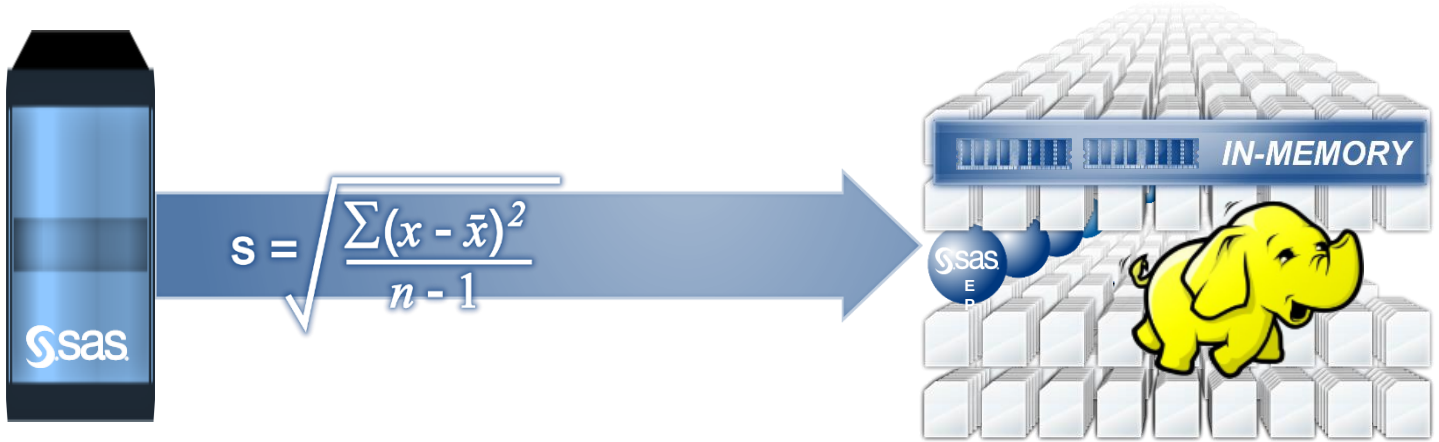
- Ability to connect to Hadoop
- Run natively in Hadoop
- Minimize data movement

How SAS Embraces…

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

HADOOP AS PROCESSING ENGINE

EMBRACE



- 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

THE ANALYTIC LIFECYCLE

EMBRACE

Discovery *& Development of Analytics*

Deployment *& Execution of Analytics*

Enterprise Wish List

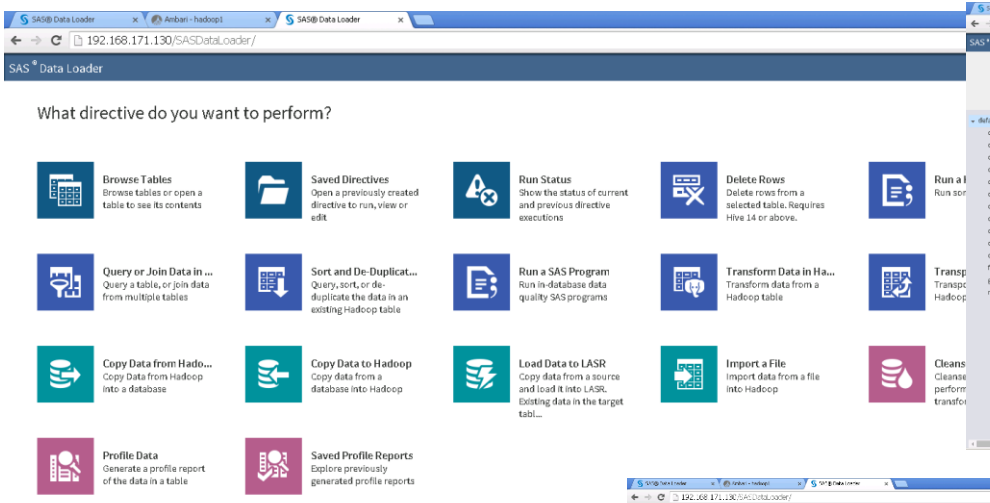
- A way for users to interact with Hadoop
- Ability to create analytic views and tables
- Ability to assess data quality

How SAS embraces…

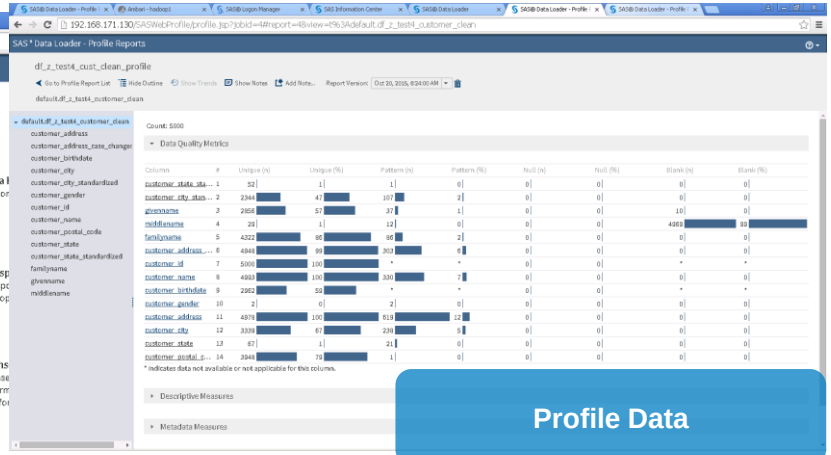
- A business user interface to facilitate:
  - Querying Hadoop
  - Adding data
  - Profiling data
  - Cleansing data
  - Transforming data
- With no data movement

SELF SERVE ACCESS TO HADOOP

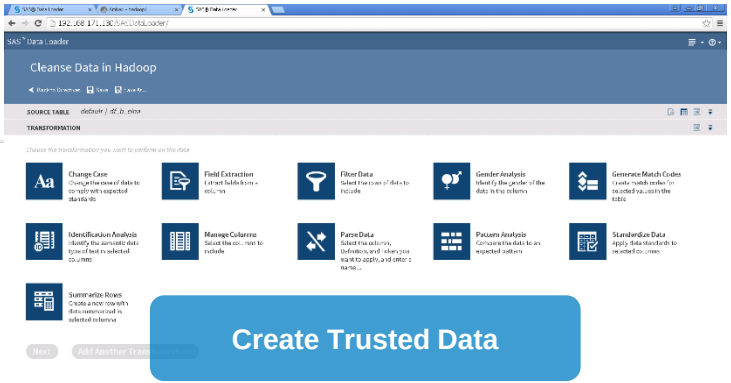
EMBRACE



Business user UI



Profile Data



Create Trusted Data

THE ANALYTIC LIFECYCLE

EXTEND

Discovery *& Development of Analytics* **Deployment** *& Execution of Analytics*

Enterprise Wish List

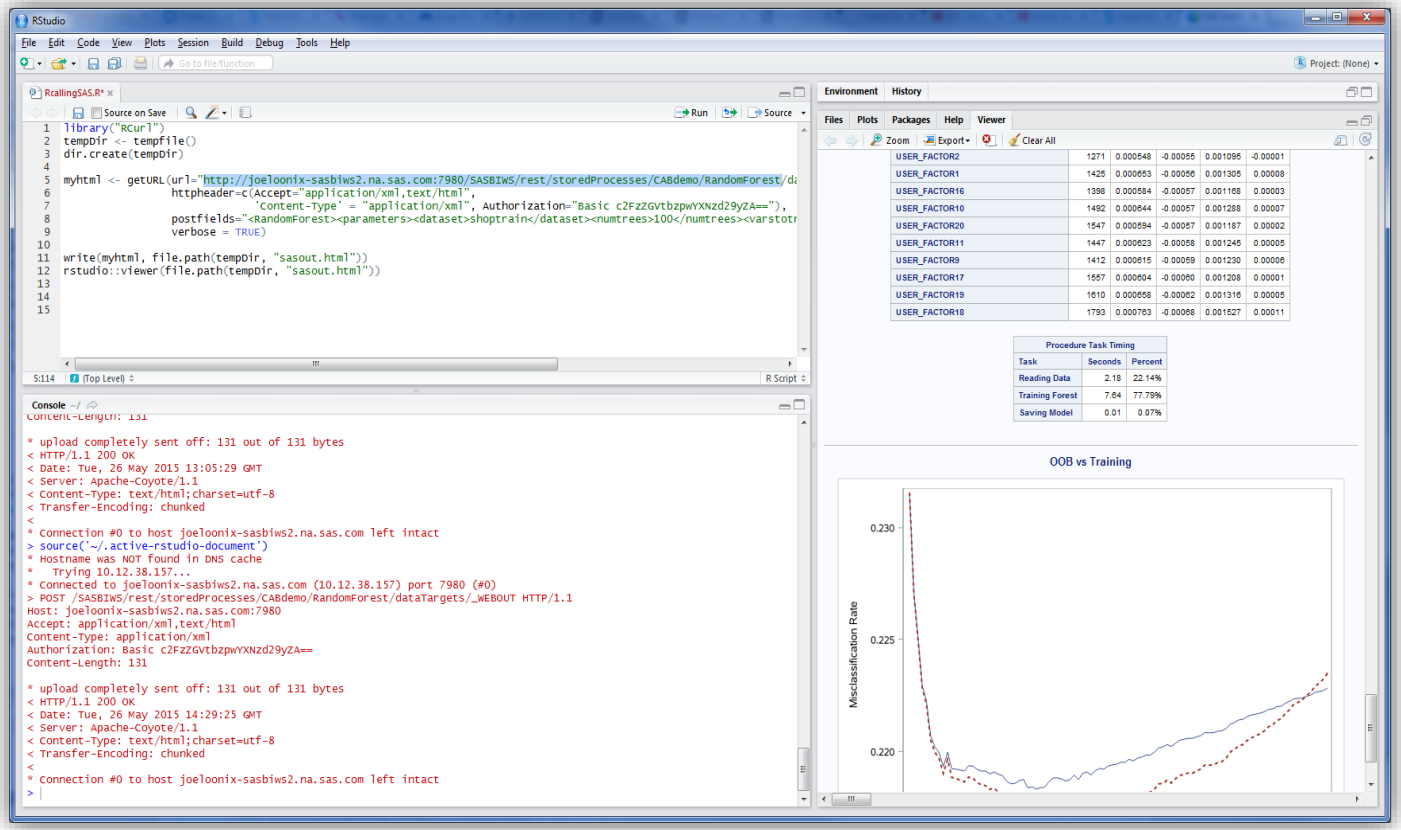
- Best possible analytics
- Flexibility of tools
- Productivity
- Greater insights = models
- Trusted models

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

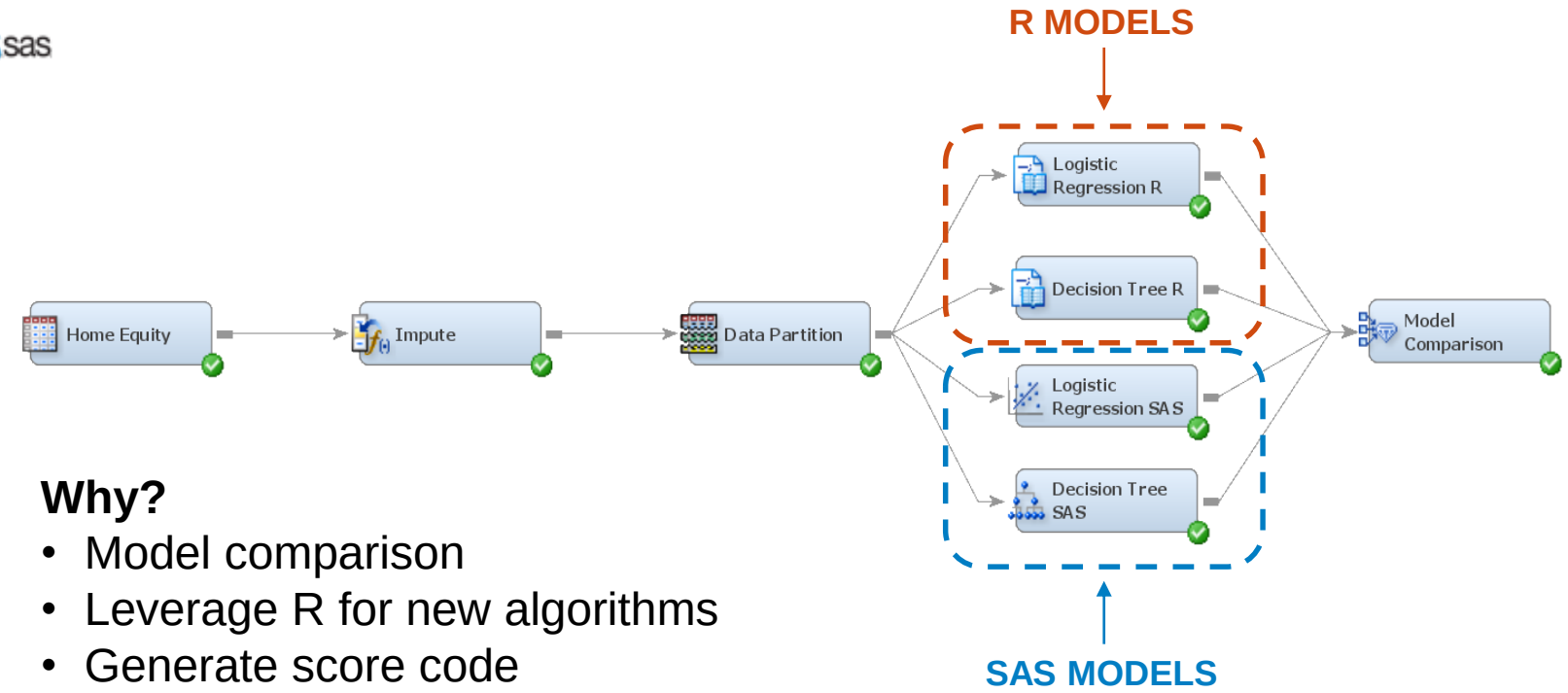
SAS FROM R

EXTEND



USE SAS TO INTEGRATE R

EXTEND

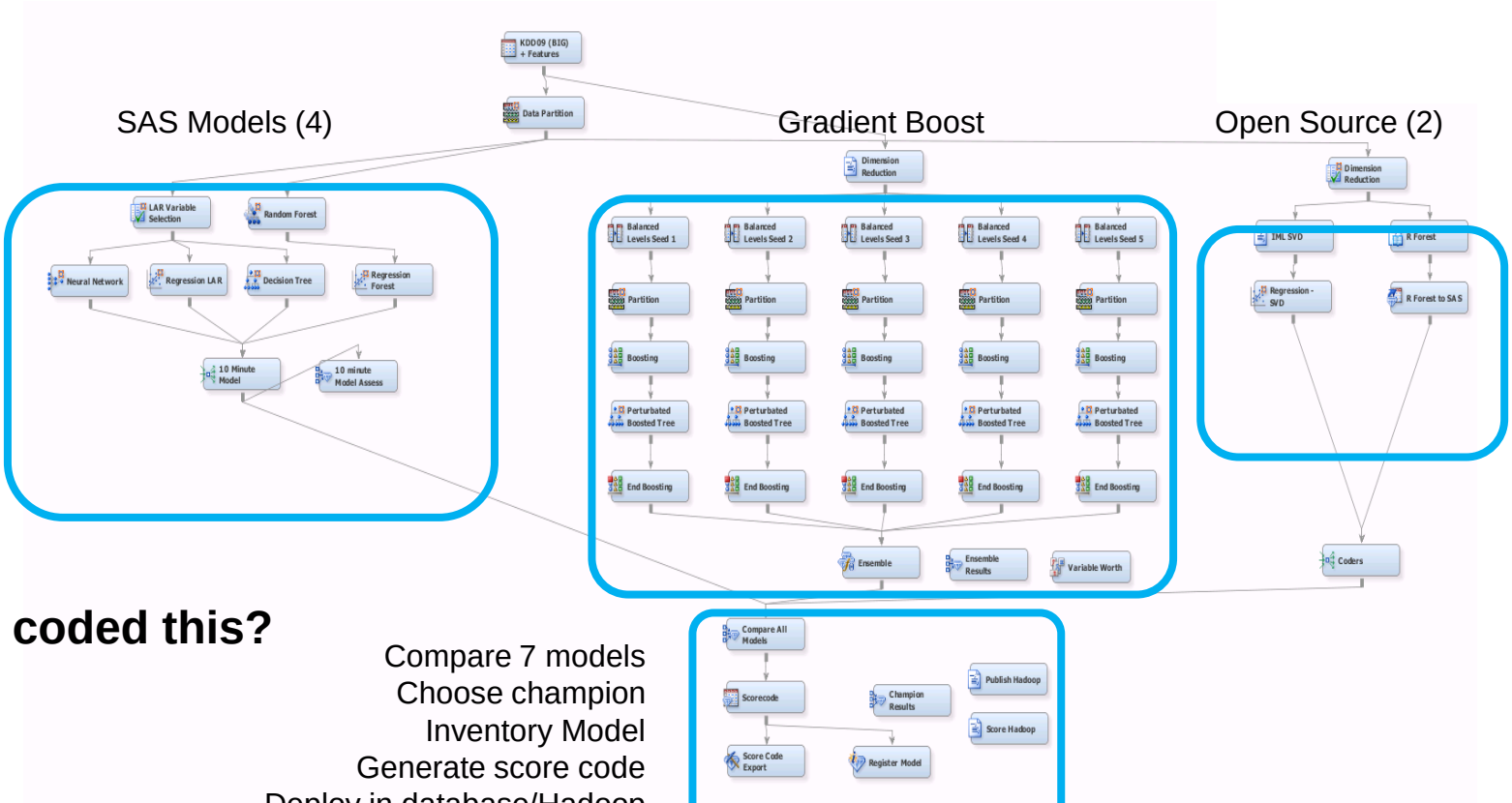


Why?

- Model comparison
- Leverage R for new algorithms
- Generate score code
- Deploy R models

PRODUCTIVITY

EXTEND

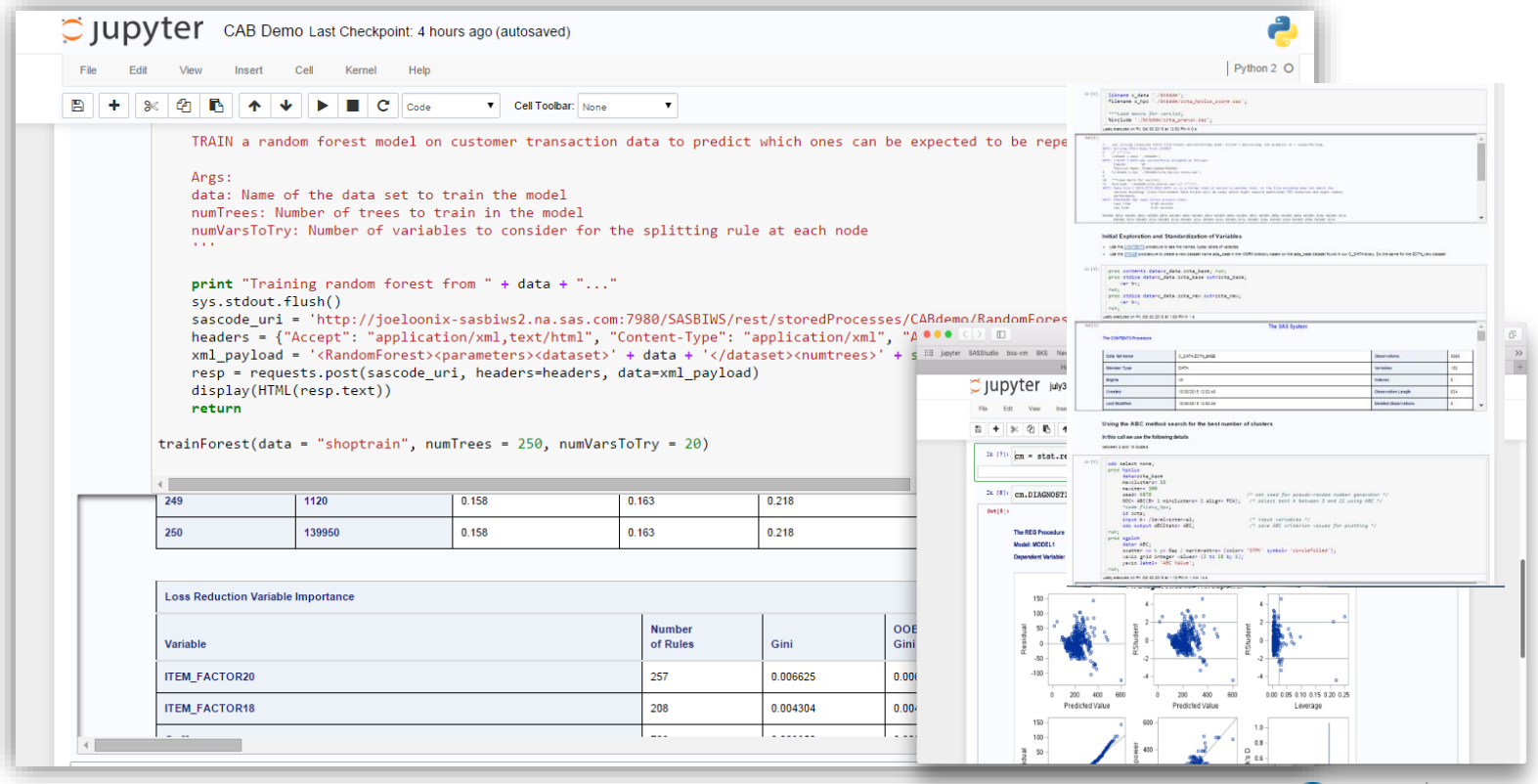


What if you coded this?

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

SAS FROM JUPYTER

EXTEND



THE ANALYTIC LIFECYCLE

EXTEND

Discovery *& Development of Analytics*

Deployment *& Execution of Analytics*

Enterprise Wish List

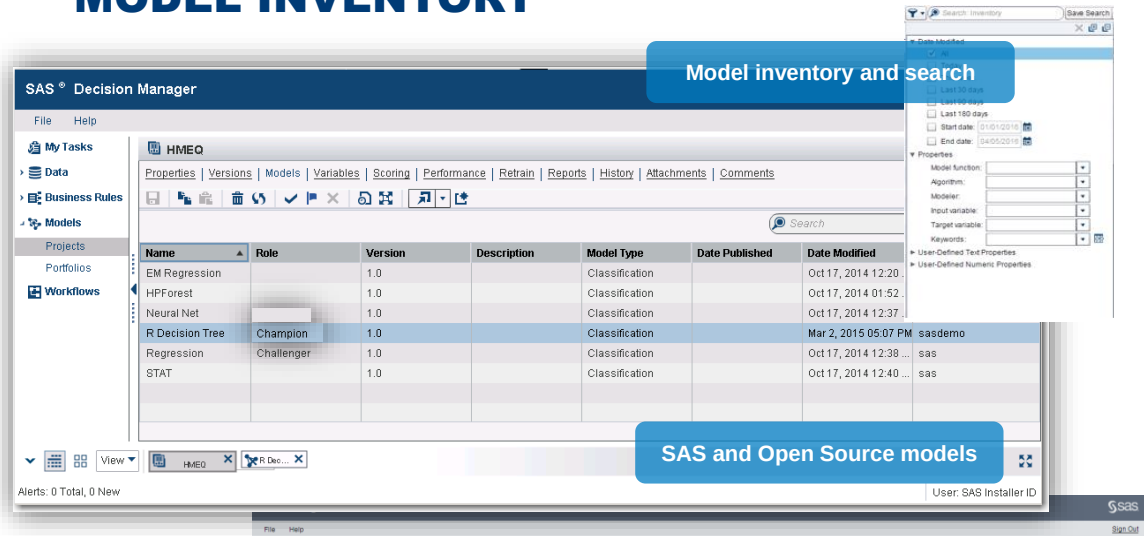
- Model management platform
- Inventory ALL models
- Know who's working on what
- Ability to deploy models
- Auditable models

How SAS Extends…

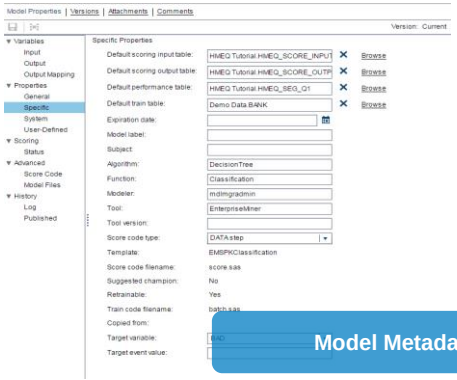
- Central model management platform
- Repository for SAS models and open source (R, Python, PMML)
- Model history
- Version control
- Model and data lineage
- Model governance

MODEL INVENTORY

EXTEND

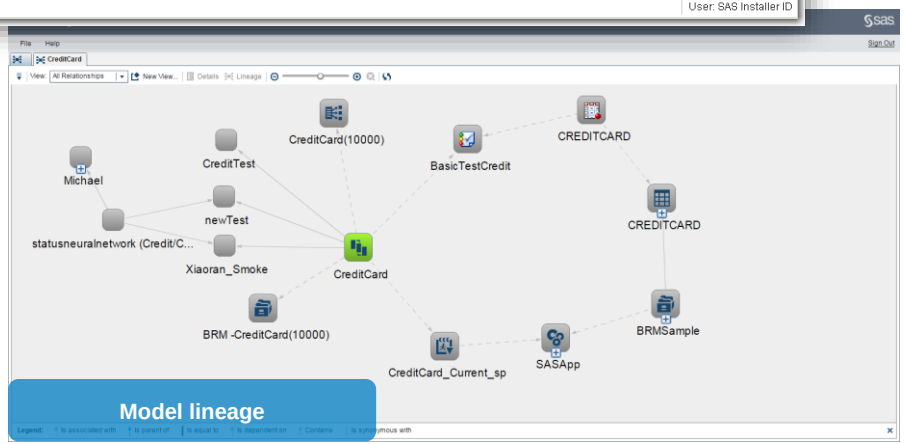


Model inventory and search



Model Metadata

SAS and Open Source models



Model lineage

THE ANALYTIC LIFECYCLE

EXTEND

Discovery *& Development of Analytics*

Deployment *& Execution of Analytics*

Enterprise Wish List

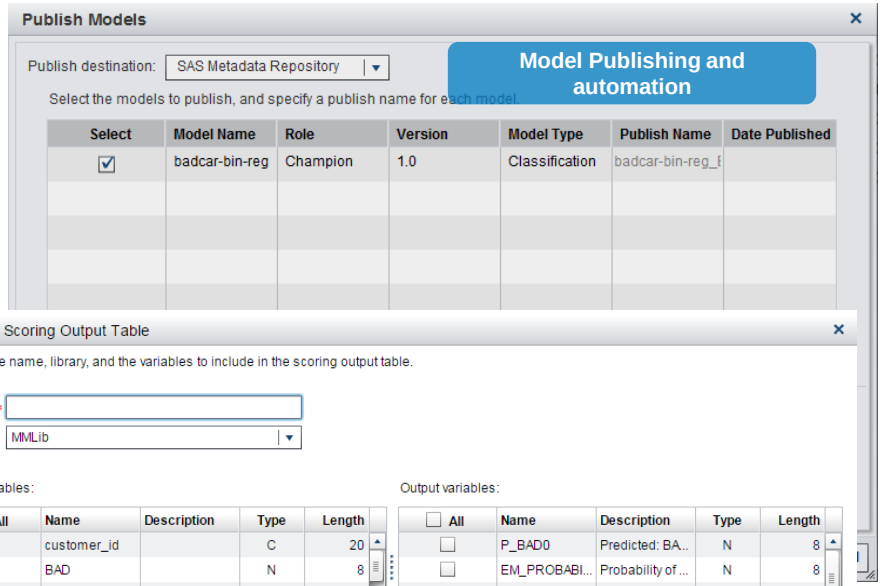
- Deployable analytics
- Automation
- Faster time to model execution
- In Hadoop/database model execution

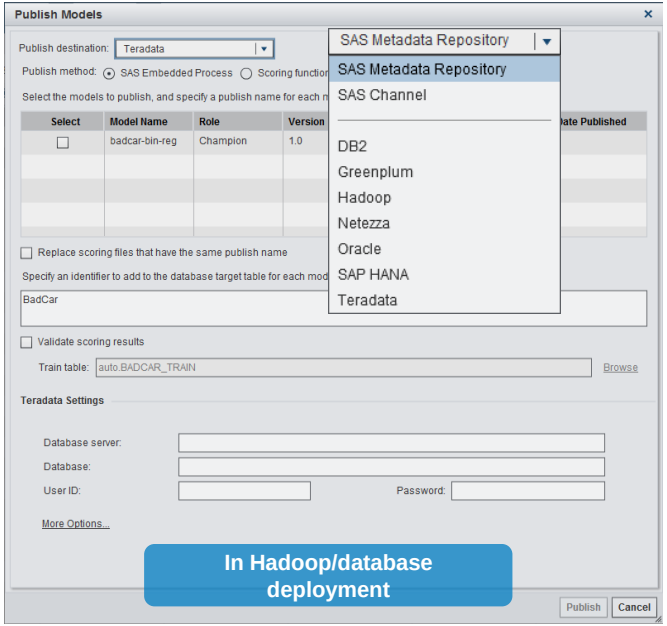
How SAS Extends…

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

MODEL EXECUTION

EXTEND





THE ANALYTIC LIFECYCLE

EXTEND

Discovery *& Development of Analytics*

Deployment *& Execution of Analytics*

Enterprise Wish List

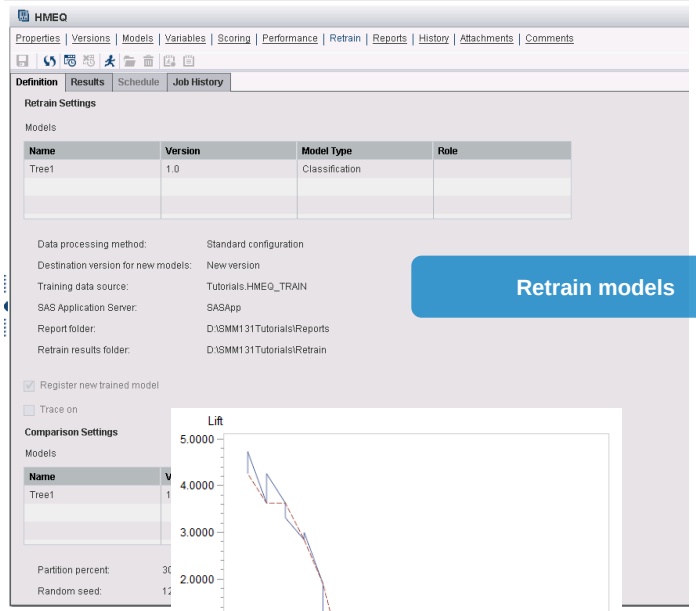
- Best possible models
- Model tournaments
- Visibility into performance
- Easy retraining
- Champion/challenger modelling

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

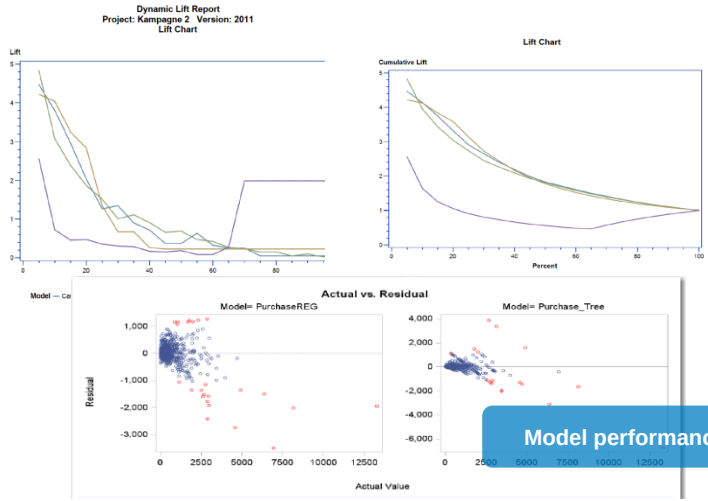
MODEL PERFORMANCE

EXTEND

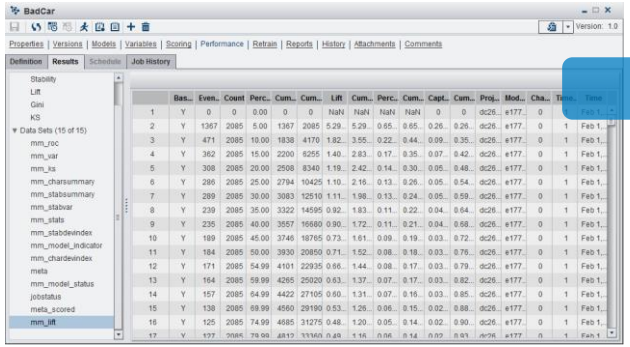


Retrain models

Model comparisons



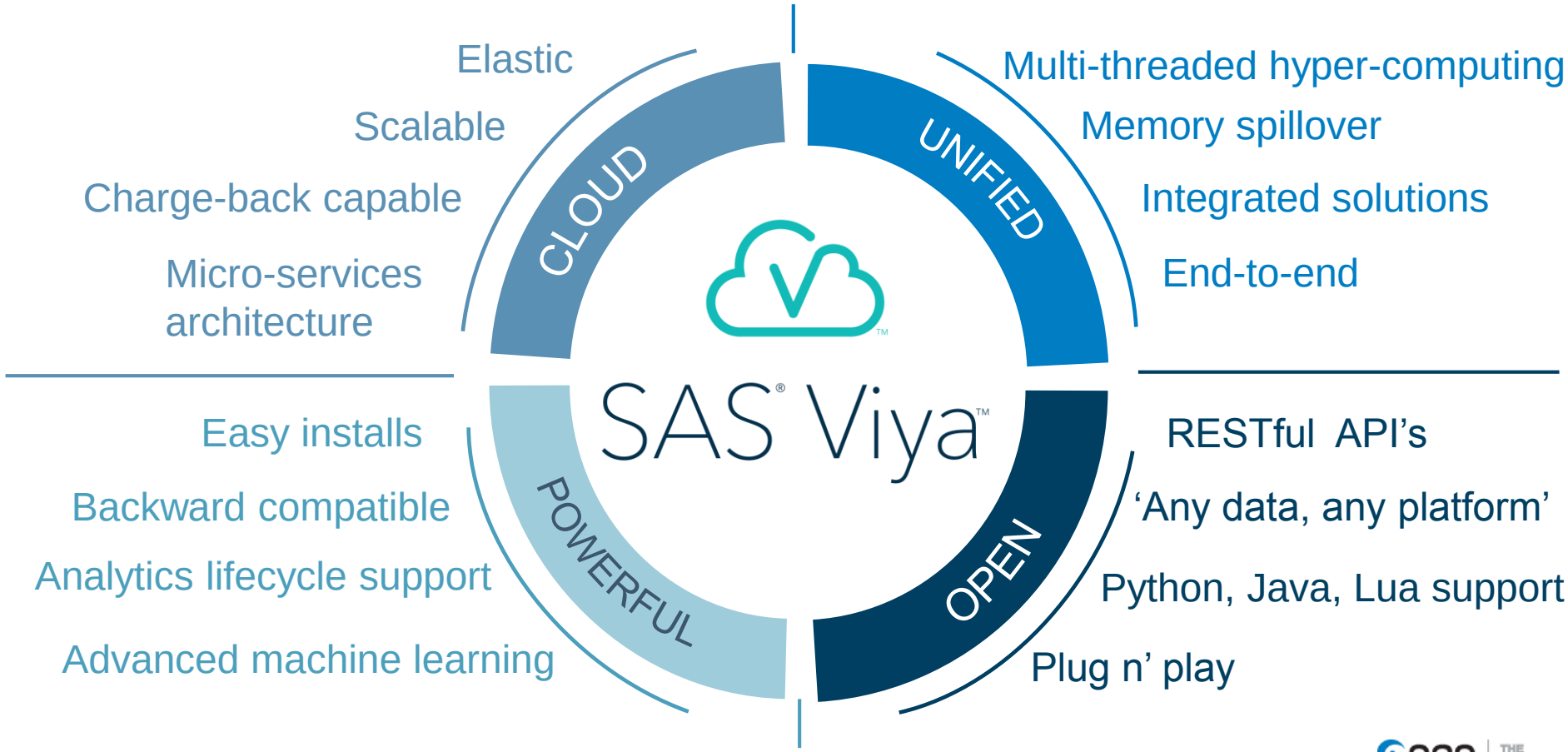
Model performance reports



Monitor data drift

THE FUTURE IS NOW...

SUPPORTING CURRENT INDUSTRY TRENDS



SAS AND OPEN SOURCE

SAS 9.4

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*
http://support.sas.com/documentation/cdl/en/imlug/66845/HTML/default/viewer.htm#imlug_r_toc.htm

Video: *Calling R Procedures from SAS/IML® Software*
https://www.youtube.com/watch?v=rUaTTre24kI

Video: *SAS/IML and R: Using Them Together*
https://www.youtube.com/watch?v=nmRQ3MtkG6A

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*
http://support.sas.com/resources/papers/proceedings10/329-2010.pdf

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

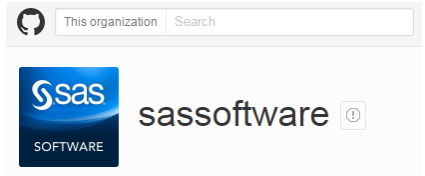
Blogs: *How to execute a Python script in SAS® Enterprise Miner™*
https://communities.sas.com/docs/DOC-10832

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:
http://support.sas.com/dsearch?Find=Search&ct=&qt=open+source&col=suppprd&nh=25&qp=&qc=suppsas&ws=1&qm=1&st=1&lk=1&rf=0&oq=&rq=0

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