

SAS IS OPEN (FOR BUSINESS)

MATT MALCZEWSKI, SAS CANADA

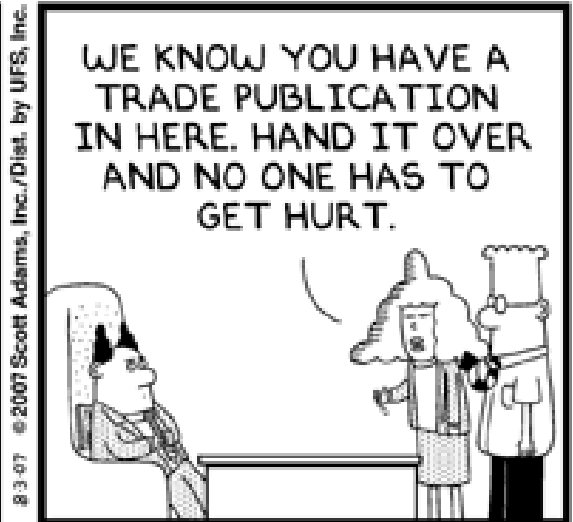
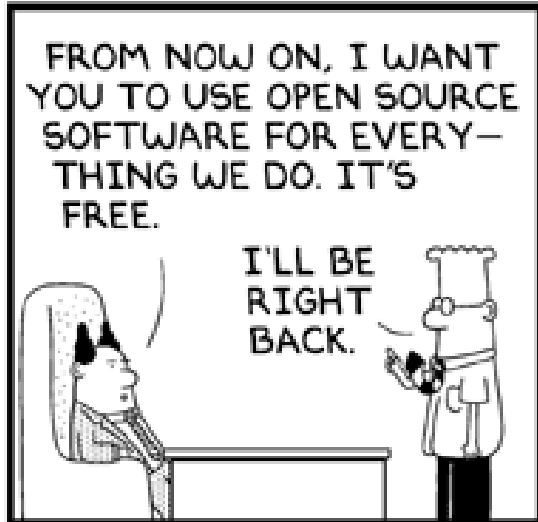


TAMARA DULL, SAS BEST PRACTICES

STEVE HOLDER, NATIONAL ANALYTICS LEAD, SAS CANADA

TINA SCHWEIHOFFER, SENIOR SOLUTION SPECIALIST, SAS CANADA

ACKNOWLEDGEMENTS



© Scott Adams, Inc./Dist. by UFS, Inc.

5 OPEN SOURCE MYTHS

the open source myth...

...and the reality

It's free.

Licensing is free: that's it.

It's 'Geekware'.

At first... but not over time.

It's 'not ready' for the Enterprise.

2010: 42%. 2015: 78%¹

It's hard to support.

Strength of community!

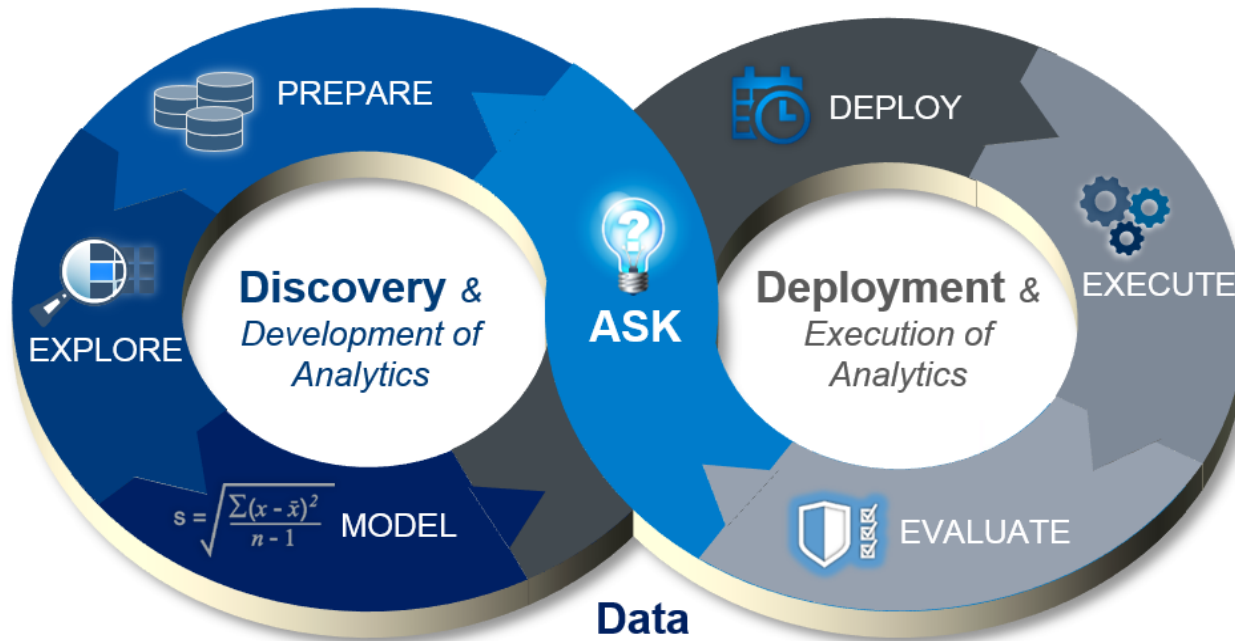
It's not secure.

55% believe it's *more* secure.¹

¹ Source: 2015 Future of Open Source Survey, North Bridge and Black Duck Software, April 2015

The Analytics Lifecycle

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



Regulated
 Automated
 Governed
 Embedded
 Reliable
 Decisions
 Consistent
 Documented
 Actions
 IT



COMPARISONS

- **Open Source Offers:**

- + A robust online community.
- + An extensive array of algorithms.
- + Low cost barriers to entry.
- + Fast adoption of new innovation.

- **SAS Offers:**

- + Productivity for users regardless of skillset.
- + Scalability to address any problem or dataset.
- + Governed analytics and data.
- + The support organizations require for production and operational analytics.

OPEN SOURCE REVOLUTION....

... means the evolution of SAS to ***embrace*** and ***extend*** the capabilities of open source as part of an analytics ecosystem.

	OPEN SOURCE	SAS EMBRACES	RESULTS
PREPARE DATA	●	<ul style="list-style-type: none"> + Native access to all data including Hadoop. + Ability to run key analytic functions in-database to reduce data movement. 	<p><i>Work with more data, identify new patterns and anomalies and uncover new insights.</i></p> <p><i>Minimize movement of data increasing performance.</i></p>
EXPLORE	●	<ul style="list-style-type: none"> + Allow non-technical users to get started with the data in a visual interface. + Embedded data preparation. + Data quality and governance. 	<p><i>Give more people access to data stored in Hadoop.</i></p> <p><i>Provision trusted, high-quality data for all.</i></p> <p><i>Improve governance by working with data inside Hadoop.</i></p>
MODEL	●	<ul style="list-style-type: none"> + Code-based and visual user experiences provide flexibility and productivity. + Approachable analytics designed for non data scientists. + Robust algorithms that scale to all data. 	<p><i>Democratize analytics.</i></p> <p><i>Free up data science resources and solve more complex business problems by shortening model development time.</i></p> <p><i>Increase model accuracy by using all your data – not just a sample – and running more iterations, more frequently.</i></p>

	OPEN SOURCE	SAS EXTENDS	RESULTS
INVENTORY	●	<ul style="list-style-type: none"> + Model management platform to inventory all models - SAS and Open Source. + Collaborative modeling environment. + Documentation, versioning and model lineage. 	<p><i>Manage analytics as an enterprise asset.</i></p> <p><i>Run your business on fact-based decisions.</i></p> <p><i>Create trusted models with visibility.</i></p> <p><i>Manage risk and compliance.</i></p> <p><i>Embed analytics into production systems.</i></p>
EXECUTE	●	<ul style="list-style-type: none"> + Complete model execution platform. + Models deployed in-database. + Automated execution processes. 	
MONITOR	●	<ul style="list-style-type: none"> + Robust analytics to enable visibility into model performance including retraining. + Champion/challenger modeling. 	

INTEGRATION POSSIBILITIES

Integrating SAS with Open Source

Open Source into
SAS environment



SAS into Open
Source environment



OPEN SOURCE IN SAS

SAS® Enterprise Miner offers score code support for **6 different R packages** and allows users to **import any type of R code**. The **open source node** can also be used to import any open source model.

Allows users to create **ensemble models** using open source and SAS.

Models can be converted to score code for operational deployment from within a drag-and-drop interface. This results in improved **productivity**, **deployment** and **scalability**.

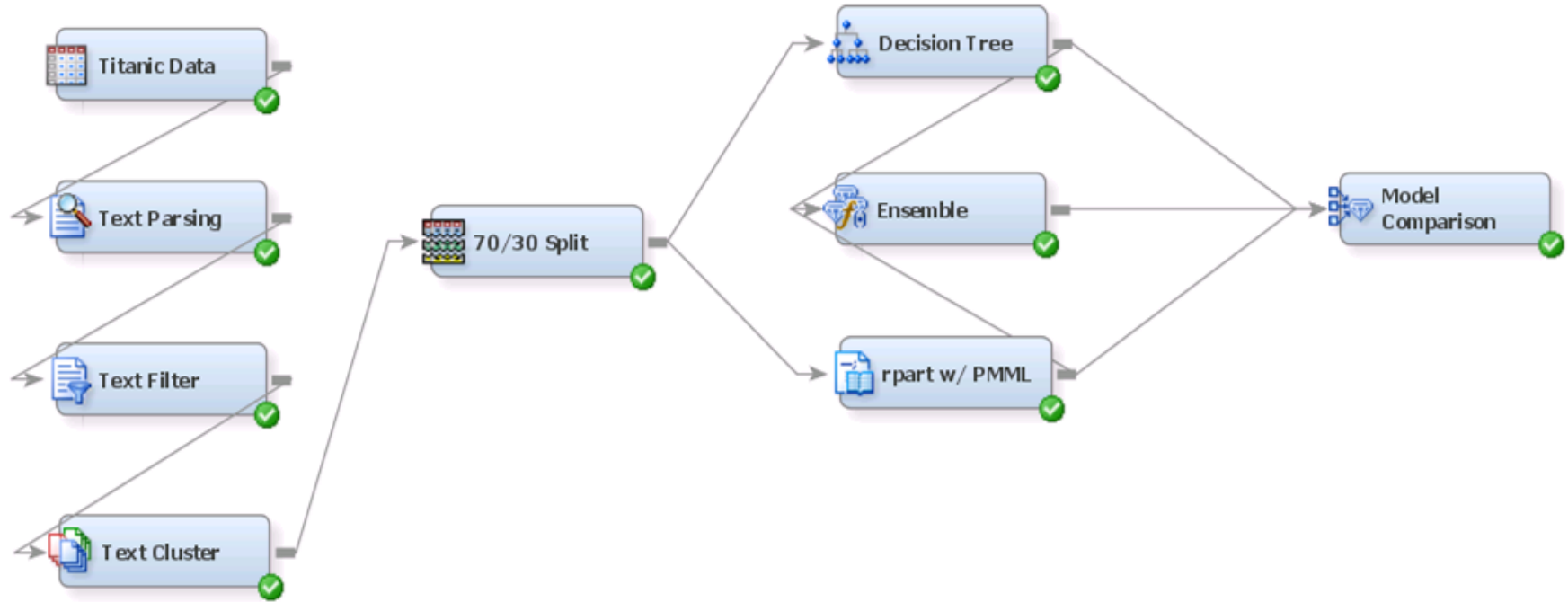
WHY IS THIS IMPORTANT?

Improve model lift and **performance** by creating blended models that **combine the best** of SAS and Open Source.

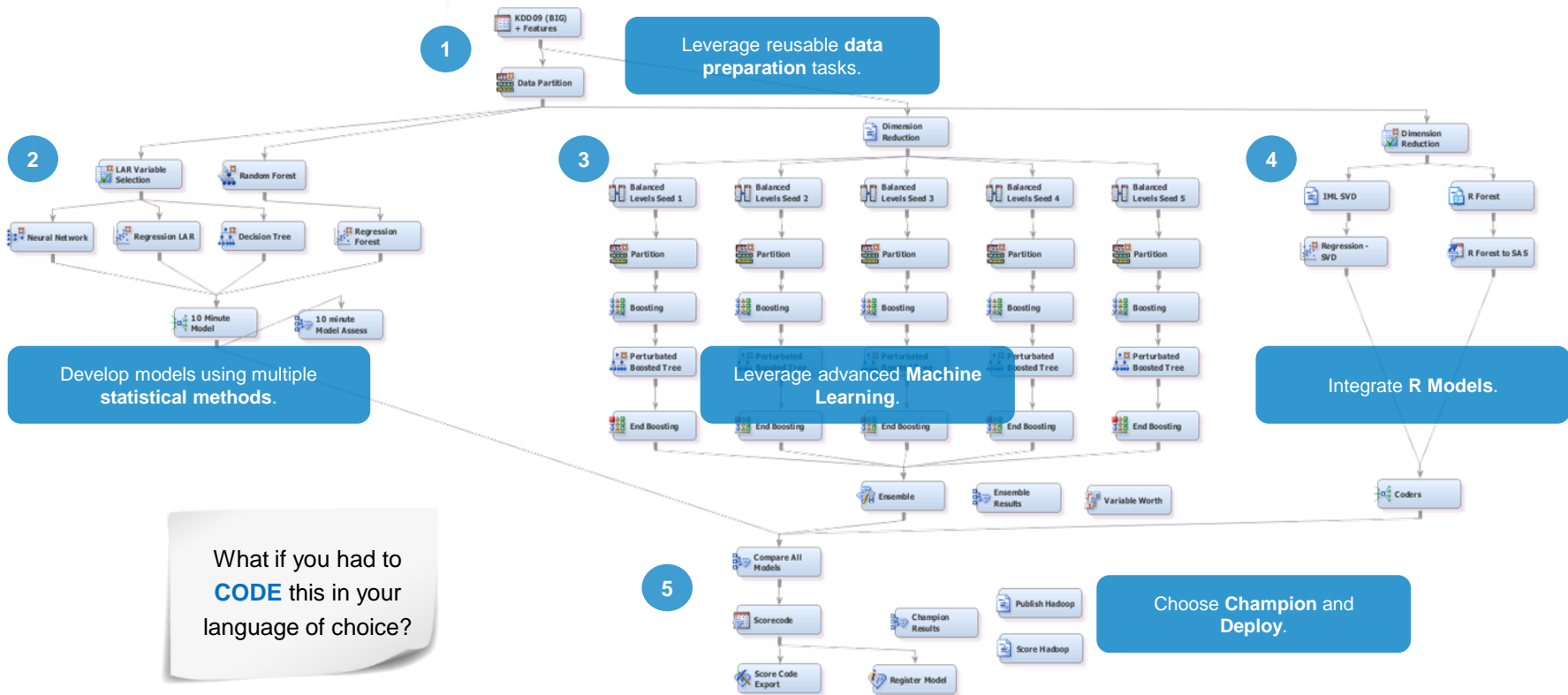
SAS automatically generates documentation capturing best practices, **promoting collaboration** and helping **reduce turnover risk**.

Facilitating Interoperability

A simple example



IN REAL LIFE



SAS IN OPEN SOURCE

- + **Base SAS** offers a Java object to incorporate a variety of external languages, including Python.
- + This allows **SAS Procedures** to be **called from** open source tools.
- + The **Jupyter kernel for SAS** brings the power of SAS data manipulation and analytics capabilities to the Jupyter notebook.

WHY IS THIS IMPORTANT?

This allows data scientists to code in their language and interface of choice, while allowing SAS to **extend** open source applications with **productivity** and the ability to **scale to any data volume**.

Using SAS in open source can ease the transition of non-SAS users: calling SAS via stored processes/APIs from other programming interfaces is a simple way for open source programmers to access SAS.

SAS and Python in Jupyter Notebook



TRAIN a random forest model on customer transaction data to predict which ones can be expected to be repeat customers

```
Args:
data: Name of the data set to train the model
numTrees: Number of trees to train in the model
numVarsToTry: Number of variables to consider for the splitting rule at each node
...

print "Training random forest from " + data + "..."
sys.stdout.flush()
sascode_uri = "http://joelooxix-sasbiws2.na.sas.com:7980/SASBIWS/rest/storedProcesses/CABdemo/RandomForest/dataTargets/_WEBOUT"
headers = {"Accept": "application/xml,text/html", "Content-Type": "application/xml", "Authorization": "Basic c2F2ZGV0b2pvc2pwYXNzd29y"}
xml_payload = "<RandomForest><parameters><dataset>" + data + "</dataset><numtrees>" + str(numTrees) + "</numtrees><varstotry>" +
resp = requests.post(sascode_uri, headers=headers, data=xml_payload)
display(HTML(resp.text))
return

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

249	1120	0.158	0.163
250	139950	0.158	0.163

ca.DIAGNOSTICSPANEL

The SAS System

data setname	numtrees	varstotry	maxdepth
shoptrain	250	250	10
shoptrain	100	250	10
shoptrain	500	250	10
shoptrain	1000	250	10

Using the ABC method search for the best mixture of elements

numtrees	varstotry	maxdepth	score
100	250	10	0.158
250	250	10	0.163
500	250	10	0.163
1000	250	10	0.163

Fit Diagnostics for Horsepower

Bringing SAS to R



RStudio interface showing R code, Environment pane, and a plot.

```
1 library("RCurl")
2 tempDir <- tempfile()
3 dir.create(tempDir)
4
5 myhtml <- getURL(url="http://joeloonix-sasbiws2.na.sas.com:7980/SASBIWS/rest/storedProcesses/CABdemo/RandomForest/d
6             httpheader=c(Accept="application/xml,text/html",
7             "Content-Type" = "application/xml", Authorization="Basic c2FzZGVtbzpwYXNzd29yZA=="),
8             postfields="<RandomForest><parameters><dataset>shoptrain</dataset><numtrees>100</numtrees><varstot
9             verbose = TRUE)
10
11 write(myhtml, file.path(tempDir, "sasout.html"))
12 rstudio::viewer(file.path(tempDir, "sasout.html"))
13
14
15
```

Environment pane showing variables:

Variable	Value
USER_FACTOR2	1271 0.000548 -0.00055 0.001095 -0.00001
USER_FACTOR1	1425 0.000853 -0.00056 0.001305 0.00008
USER_FACTOR16	1398 0.000584 -0.00057 0.001168 0.00003
USER_FACTOR10	1492 0.000844 -0.00057 0.001288 0.00007
USER_FACTOR20	1547 0.000594 -0.00057 0.001187 0.00002
USER_FACTOR11	1447 0.000623 -0.00058 0.001245 0.00005
USER_FACTOR9	1412 0.000615 -0.00059 0.001230 0.00006
USER_FACTOR17	1557 0.000604 -0.00060 0.001208 0.00001
USER_FACTOR19	1610 0.000858 -0.00062 0.001316 0.00005
USER_FACTOR18	1793 0.000763 -0.00068 0.001527 0.00011






Procedure Task Timing

Task	Seconds	Percent
Reading Data	2.18	22.14%
Training Forest	7.64	77.79%
Saving Model	0.01	0.07%

OOB vs Training

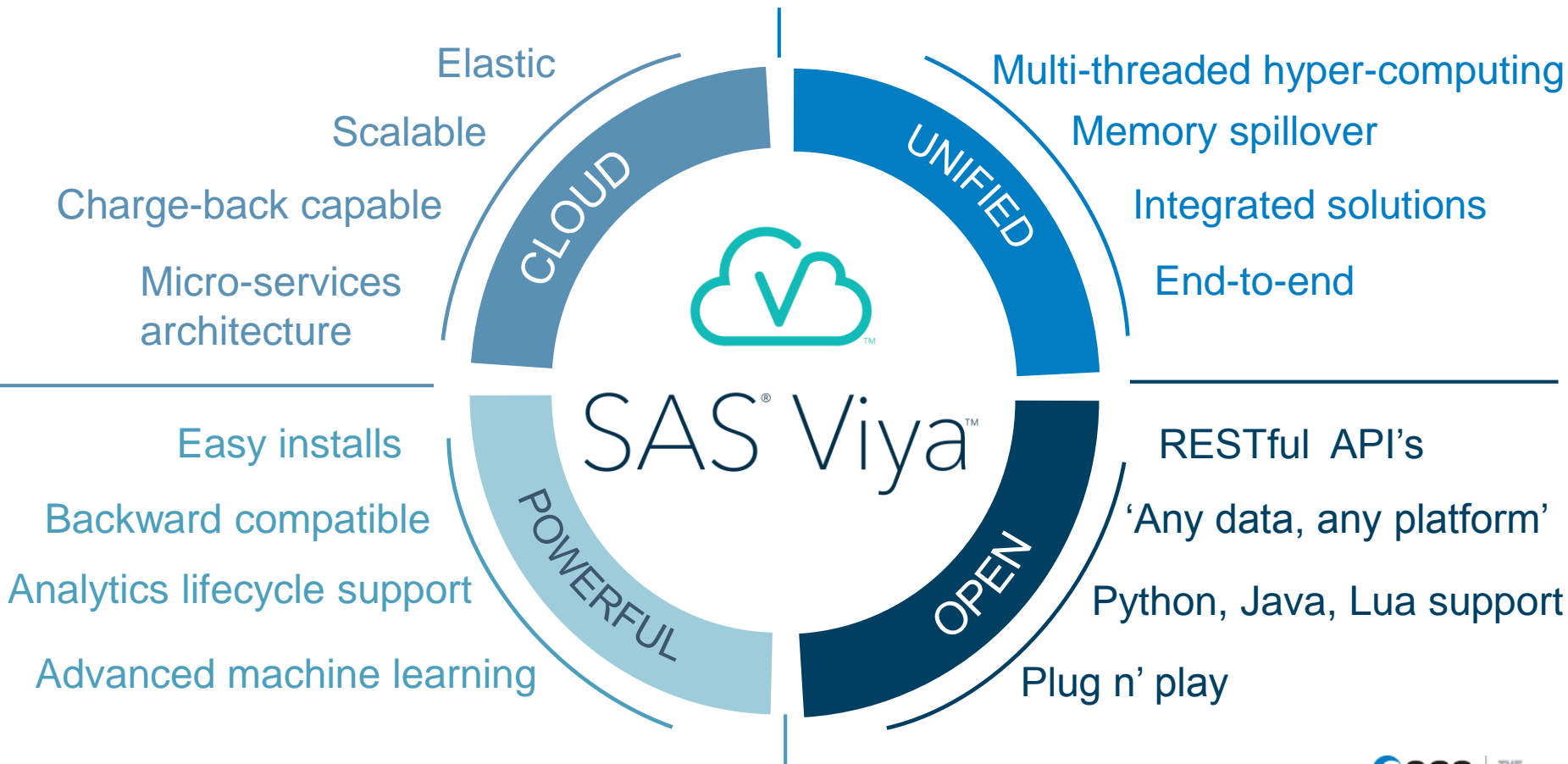
THE POWER OF MODELS

- SAS supports analytic model deployment with inventory, scoring, monitoring and retraining capabilities for **SAS and Open Source** models.

			 Supported PMML	 Non-PMML		
Inventory		●	●	●	●	●
Publish & Score	Batch	●	●	●	●	●
	In-Database	●	●	●	●	●
	Web Service	●	●	●	●	●
	Streaming	●	●	●	●	●
Monitor	●	●	●	●	●	
Retrain	●	●	●	●	●	

THE FUTURE IS NOW...





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@SAS.COM

TWITTER: MALCHEW

LINKEDIN: MATT MALCZEWSKI



THE POWER TO KNOW.