

# SAS IS OPEN (FOR BUSINESS)

MATT MALCZEWSKI, SAS CANADA



# ACKNOWLEDGEMENTS

TAMARA DULL, SAS BEST PRACTICES  
STEVE HOLDER, NATIONAL ANALYTICS LEAD, SAS CANADA  
TINA SCHWEIHOFFER, SENIOR SOLUTION SPECIALIST, SAS  
CANADA  
CHRIS HEMEDINGER, MANAGER, COMMUNITIES.SAS.COM



© 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%<sup>1</sup>

It's hard to support.

Strength of community!

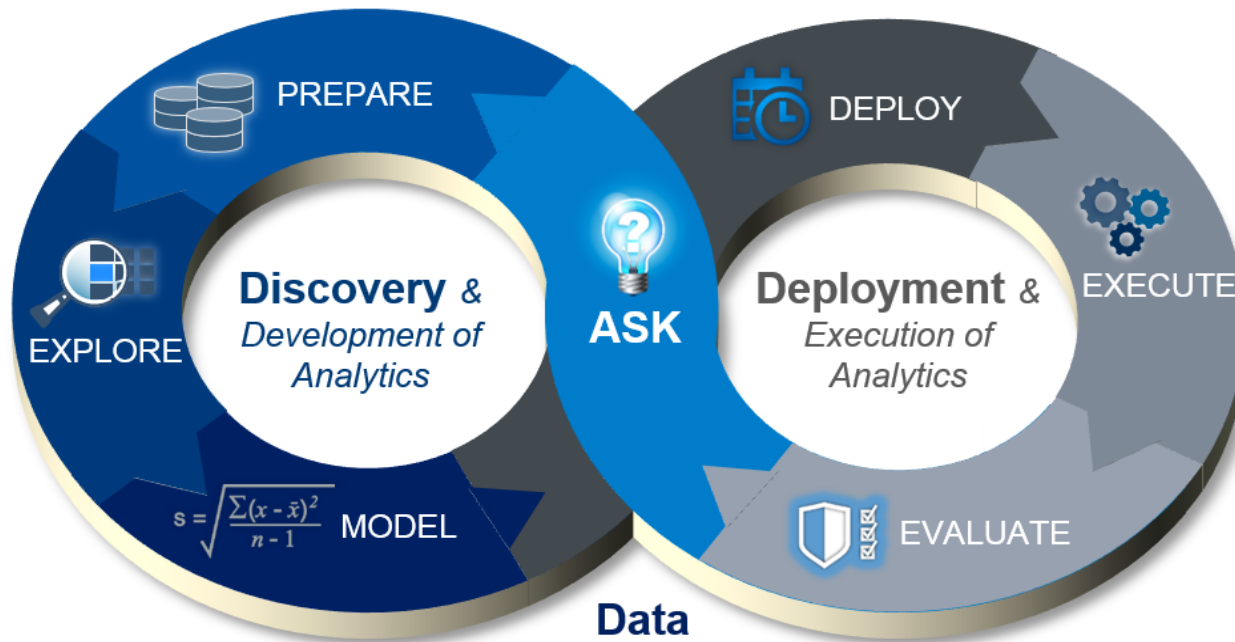
It's not secure.

55% believe it's *more* secure.<sup>1</sup>

<sup>1</sup> 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"> <li>+ Native access to <b>all data</b> including Hadoop.</li> <li>+ Ability to <b>run key analytic functions</b> in-database to <b>reduce data movement</b>.</li> </ul>	<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"> <li>+ Allow <b>non-technical users</b> to get started with the data in a <b>visual interface</b>.</li> <li>+ <b>Embedded</b> data preparation.</li> <li>+ <b>Data quality</b> and <b>governance</b>.</li> </ul>	<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"> <li>+ <b>Code-based</b> and <b>visual</b> user experiences provide <b>flexibility</b> and <b>productivity</b>.</li> <li>+ <b>Approachable analytics</b> designed for non data scientists.</li> <li>+ <b>Robust algorithms that scale</b> to all data.</li> </ul>	<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"> <li>+ <b>Model management</b> platform to inventory <b>all models</b> - SAS and Open Source.</li> <li>+ <b>Collaborative</b> modeling environment.</li> <li>+ Documentation, versioning and <b>model lineage</b>.</li> </ul>	<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"> <li>+ <b>Complete</b> model execution platform.</li> <li>+ Models <b>deployed in-database</b>.</li> <li>+ <b>Automated</b> execution processes.</li> </ul>	
MONITOR	●	<ul style="list-style-type: none"> <li>+ Robust analytics to enable <b>visibility into model performance</b> including retraining.</li> <li>+ <b>Champion/challenger</b> modeling.</li> </ul>	



# 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**.

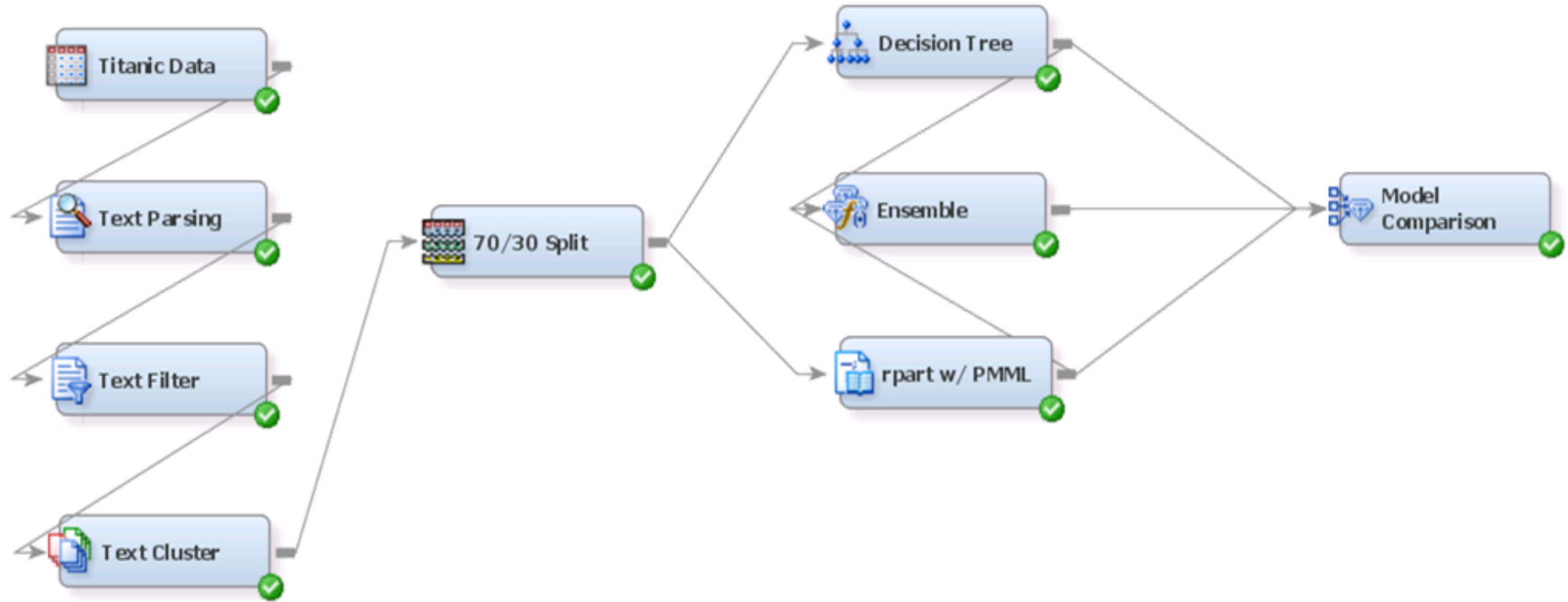
## WHY IS THIS IMPORTANT?

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

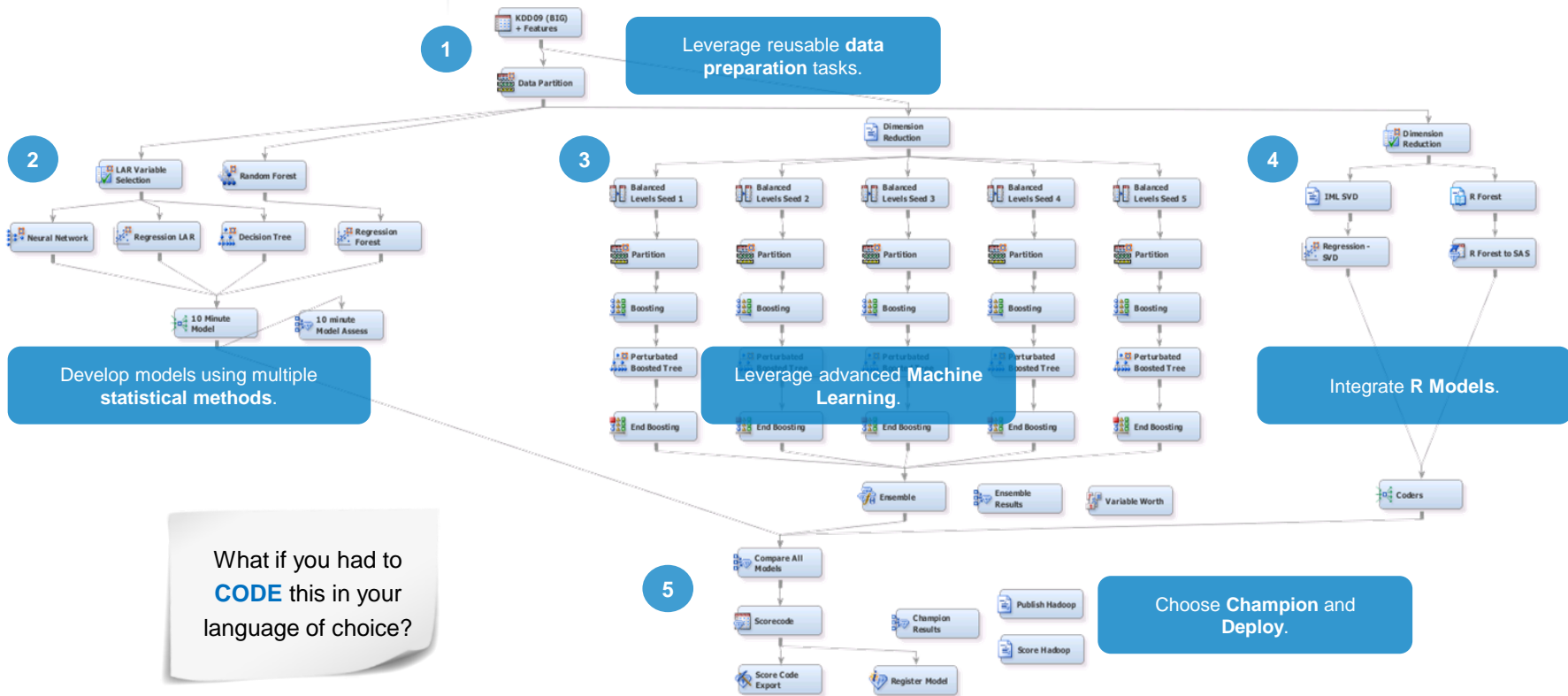
SAS automatically generates documentation capturing best practices, **promoting collaboration** and helping **reduce turnover risk**.

# 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://joeloonix-sasbius2.na.sas.com:7980/SASBIUS/rest/storedProcesses/CABdemo/RandomForest/dataTargets/_WEBOUT"
headers = {"Accept": "application/xml,text/html", "Content-Type": "application/xml", "Authorization": "Basic c2F2ZGV0b2pwbWVXNzZ299"}
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 =
```

249	1120	0.158	0.163
250	139950	0.158	0.163

```
ca = stat.reg(model="horsepower = length weight mwp mpg city", data=ca)
ca.DIAGNOSTICSPANEL
res[4]
```

**The SAS System**

The REG Procedure  
Model REG1  
Dependent Variable: horsepower

**Fit Diagnostics for Horsepower**

Statistics	Value	DF
Sum of Squares	102000.0000	1
Mean Square	102000.0000	1
F Value	102000.0000	1
Root Mean Square	319.5294	1

```
proc reg data=ca;
  model horsepower = length weight mwp mpg city;
run;
```



THE POWER TO KNOW.



# Bringing SAS to R



The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains R code for calling SAS via REST API to retrieve a Random Forest model.
- Console:** Shows the output of the REST API call, including headers and status information.
- Environment:** Lists the loaded packages and the environment variables for the Random Forest model.
- Plots:** Displays a plot titled "OOB vs Training" showing the misclassification rate.

```
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:

Package	Version	Loaded	Source
USER_FACTOR2	1271	0.000548	-0.00055
USER_FACTOR1	1425	0.000853	-0.00056
USER_FACTOR16	1398	0.000584	-0.00057
USER_FACTOR10	1492	0.000844	-0.00057
USER_FACTOR20	1547	0.000594	-0.00057
USER_FACTOR11	1447	0.000623	-0.00058
USER_FACTOR9	1412	0.000615	-0.00059
USER_FACTOR17	1557	0.000604	-0.00060
USER_FACTOR19	1610	0.000858	-0.00062
USER_FACTOR18	1793	0.000763	-0.00068






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

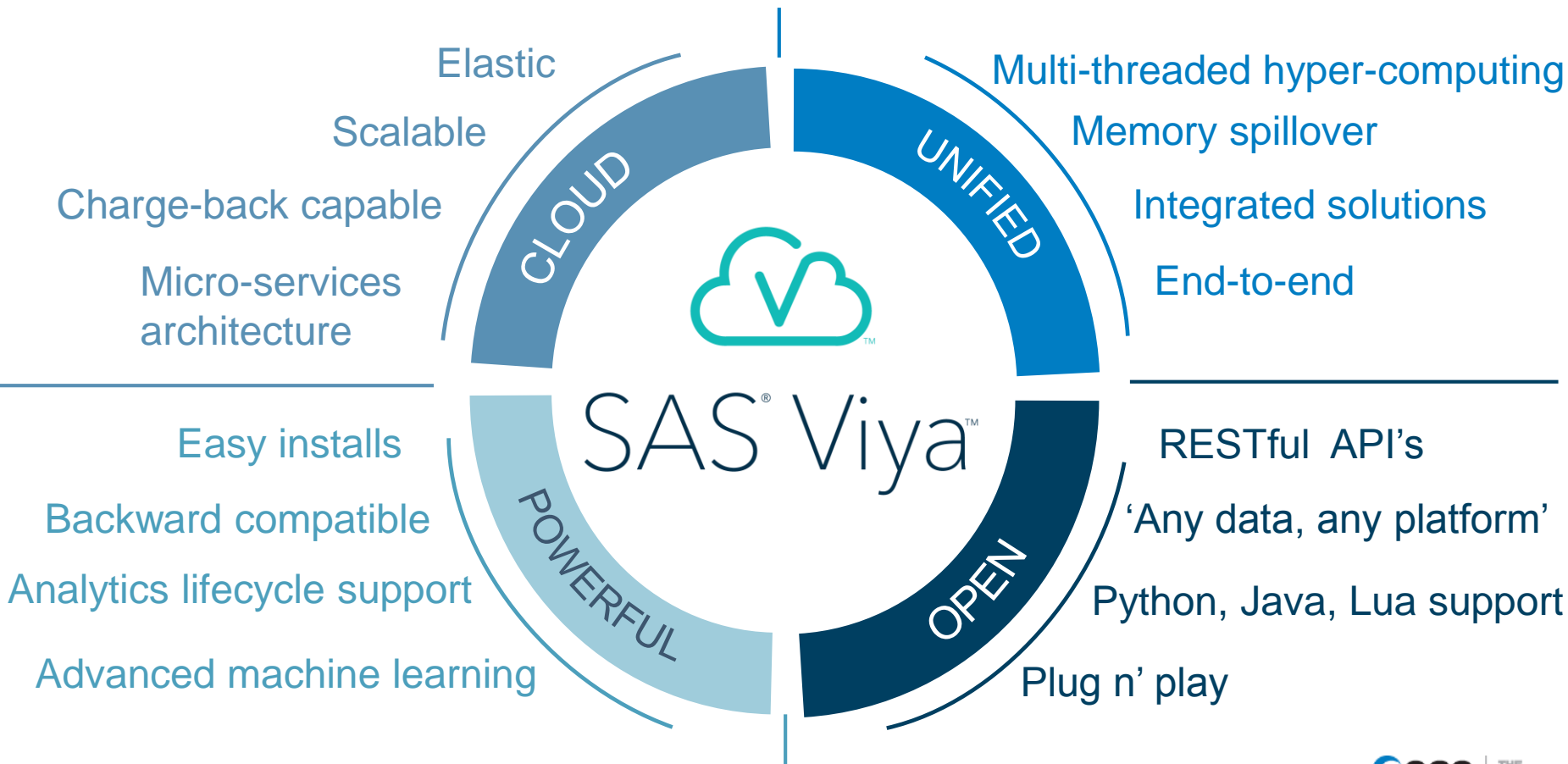
# THE POWER OF MODELS

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

			 Supported PMML	 Non-PMML		
<b>Inventory</b>		●	●	●	●	●
<b>Publish &amp; Score</b>	Batch	●	●	●	●	●
	In-Database	●	●	●	●	●
	Web Service	●	●	●	●	●
	Streaming	●	●	●	●	●
<b>Monitor</b>	●	●	●	●	●	
<b>Retrain</b>	●	●	●	●	●	

**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

## RESOURCES

- [How SAS Adds Value to Open Source](#)
- VIDEO: [SAS Enterprise Miner and R](#)
- Whitepaper: [The Use of Open Source is Growing. So Why Do Organizations Still Turn to SAS?](#)
- Whitepaper: [SAS Analytics and Open Source](#)
- [Using R in the Open Source Integration Node in SAS® Enterprise Miner](#)
- [What is the value of R in SAS Enterprise Miner?](#)
- [SAS AND OPEN SOURCE –A MATCH MADE IN ENTERPRISE MINER](#)
- [Using R in SAS Enterprise Miner – YouTube ▶ 8:24](#)

## CLOSER TO HOME | EXCITING NEWS FOR SAS CUSTOMERS

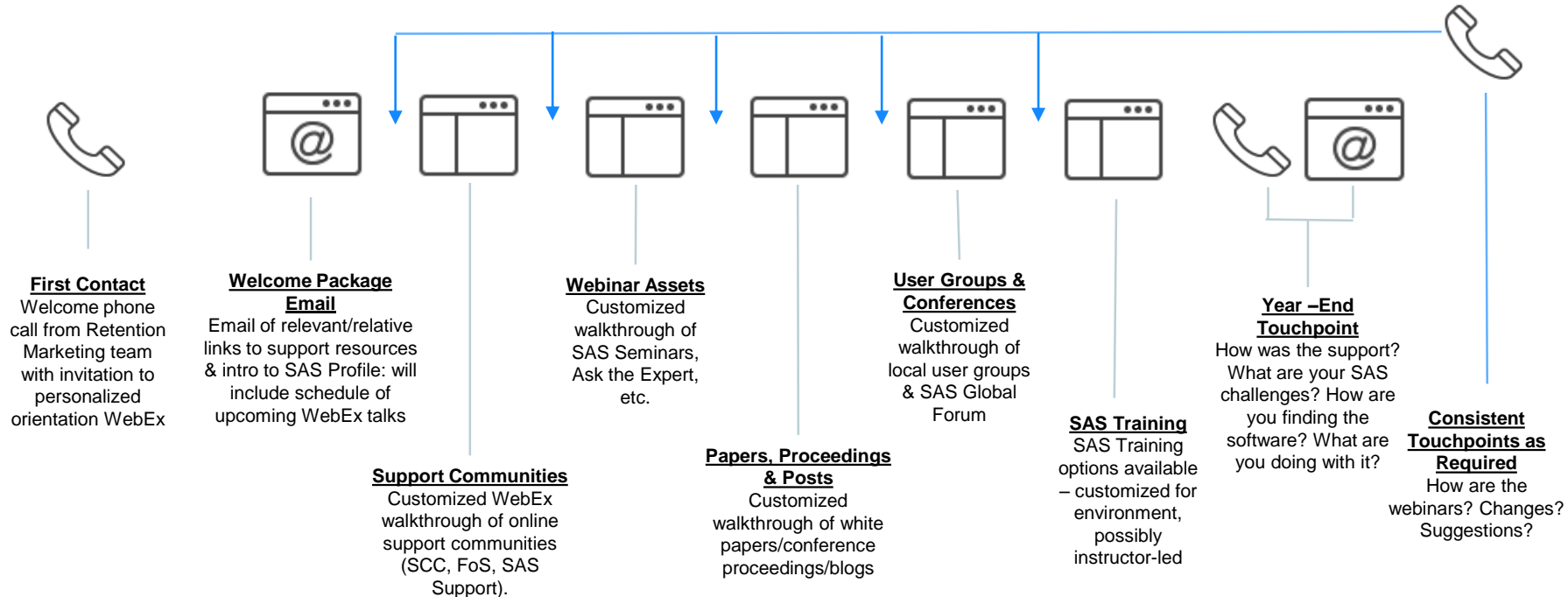
- Ever feel overwhelmed as a newer (or long-time) SAS user when facing all the support options available?
- Lots of options – not so great with wayfinding
- Which is best for *me*?
- How do I search?
- What is the value/benefit?
- Why should I bother?

## INTRODUCING... SAS SUPPORT ONBOARDING PROGRAM

- Developed here in Canada
- Designed for newer SAS customers... but also, for seasoned experts that want some guidance around navigating SAS Support
- SAS has a *responsibility* to help our customers find value in their investment
- We can achieve this by facilitating...
  - Comfort with the tools/solutions
  - Guidance through support and knowledge landscape
  - Dialogue around challenges/opportunities
  - Connection points with other SAS users



# PERSONA: USER/PRACTITIONER (6 MONTH PROCESS)



# SOUND INTERESTING?

- Email Matt Malczewski at [matt.malczewski@sas.com](mailto:matt.malczewski@sas.com)
- If personalized onboarding can't be arranged, assets can certainly be shared!

**THANK YOU**

**MATT.MALCZEWSKI@SAS.COM**

**TWITTER: MALCHEW**

**LINKEDIN: MATT MALCZEWSKI**



THE POWER TO KNOW.