

SAS, OPEN SOURCE & VIYA

MATT MALCZEWSKI, SAS CANADA



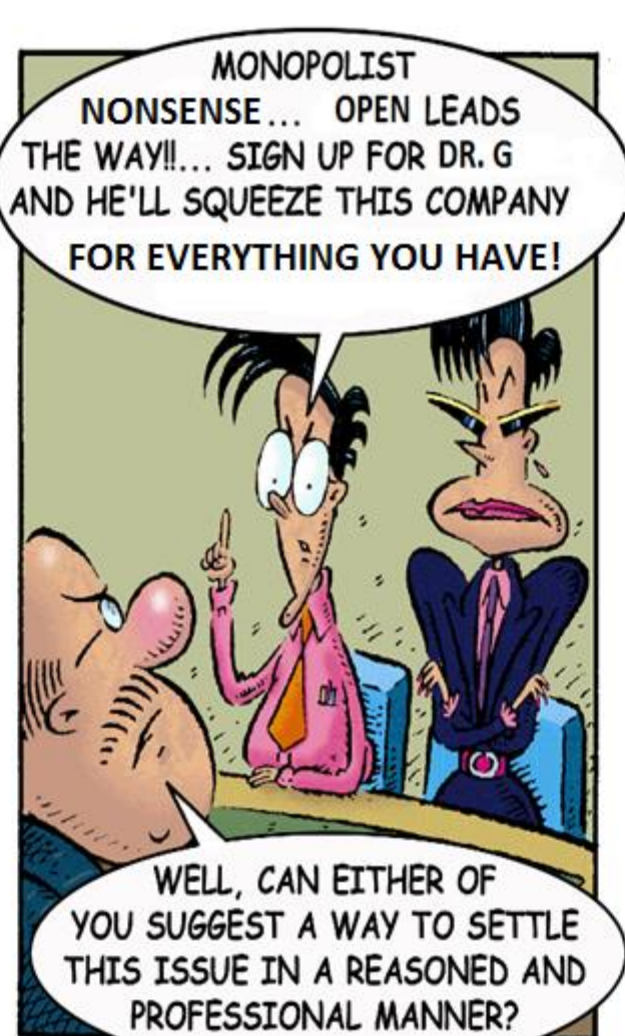
ACKNOWLEDGEMENTS

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



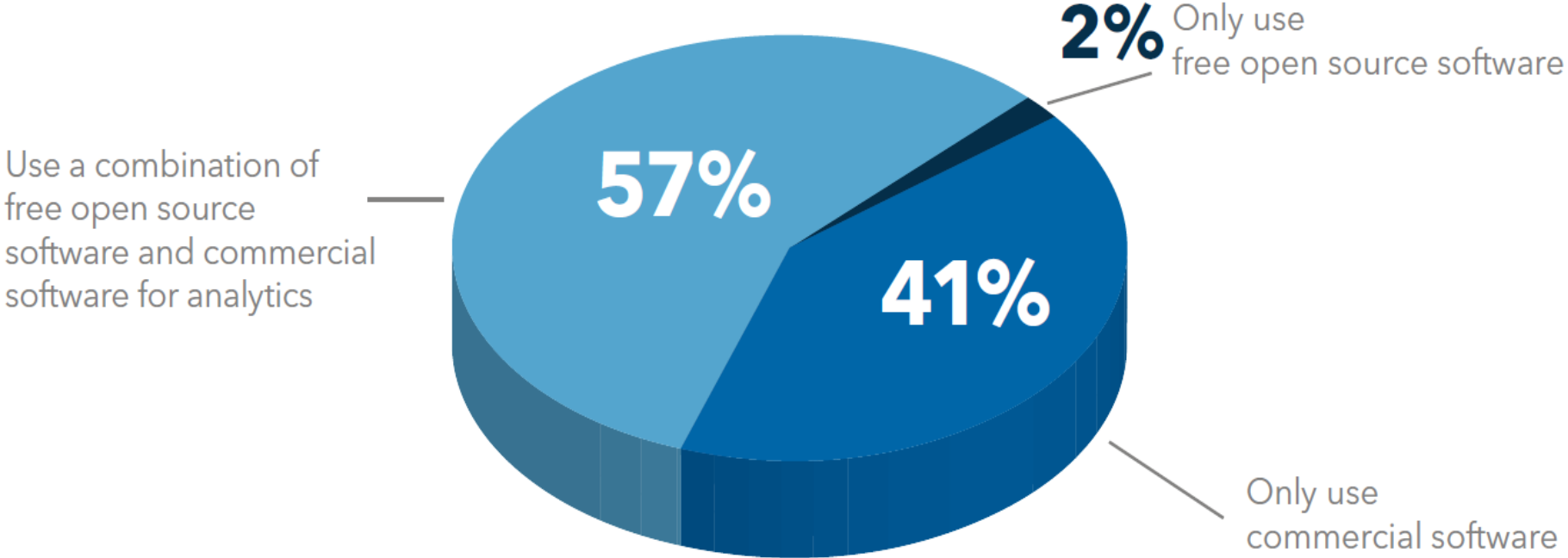
Commercial Software

“Ability to produce complex and reliable results, more secure and better support for troubleshooting.”

Free Open Source Software

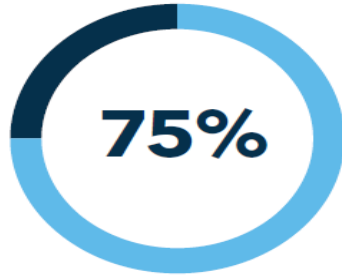
“Quicker releases and ability to add functionality with ease.”

For data, discovery and analytical insight...

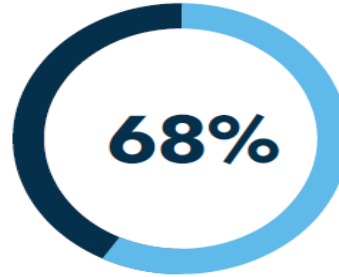


[SAS' Open Source Survey, September 2016](#)

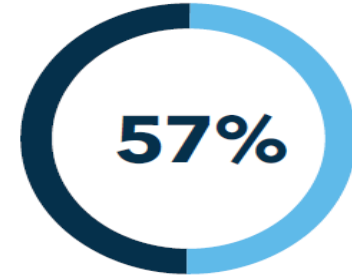
Organizations use the combined powers of commercial software and FOSS for:



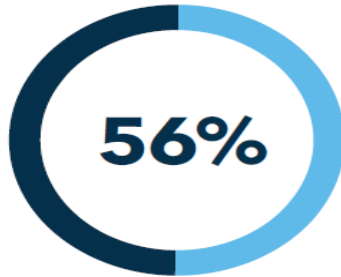
Analytical exploration



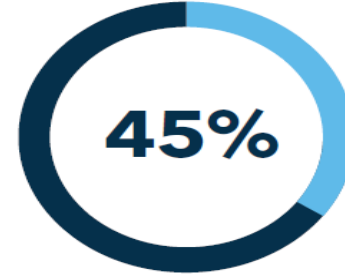
Building models



Analytics reporting



Developing applications



Data management for analytics

A leather belt with a solid steel buckle and numerous metal compartments placed along its length, the utility belt is one of the foremost items in the Batman's arsenal against crime.

It must be noted that the Caped Crusader is continuously updating and improving the contents of the utility belt, so the accompanying illustration shows merely an average sampling of its contents.

TWO-WAY RADIO
(CONCEALED IN BELT BUCKLE)

BATARANG FITS INTO COMPARTMENT BEHIND BELT

BATLINE IS DRAWN OUT OF BELT LINING FROM SPRING-LOADED REEL

BATMAN'S UTILITY BELT



OPEN SOURCE REVOLUTION....

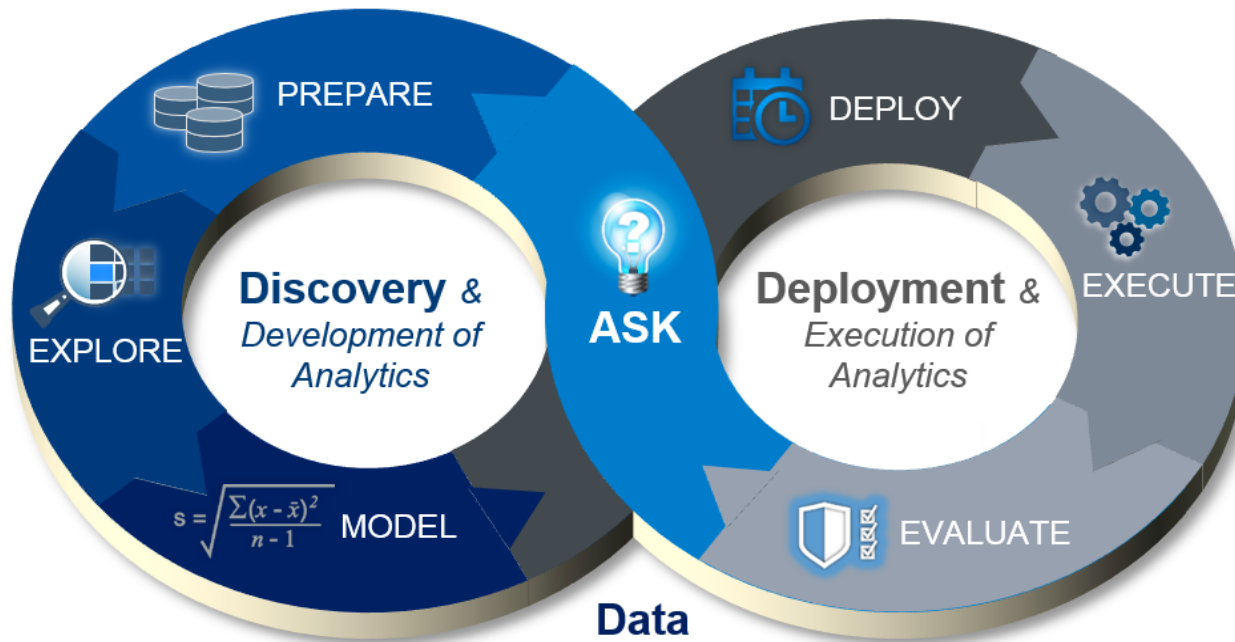
... means the evolution of SAS to ***embrace*** and ***extend*** the capabilities of open source as part of an analytics ecosystem.

WHAT IS 'OPEN' TO SAS?

- Not just about open source software
- Refers to the openness of SAS' platform and the flexibility, creativity and stability and inclusivity it offers

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



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**.

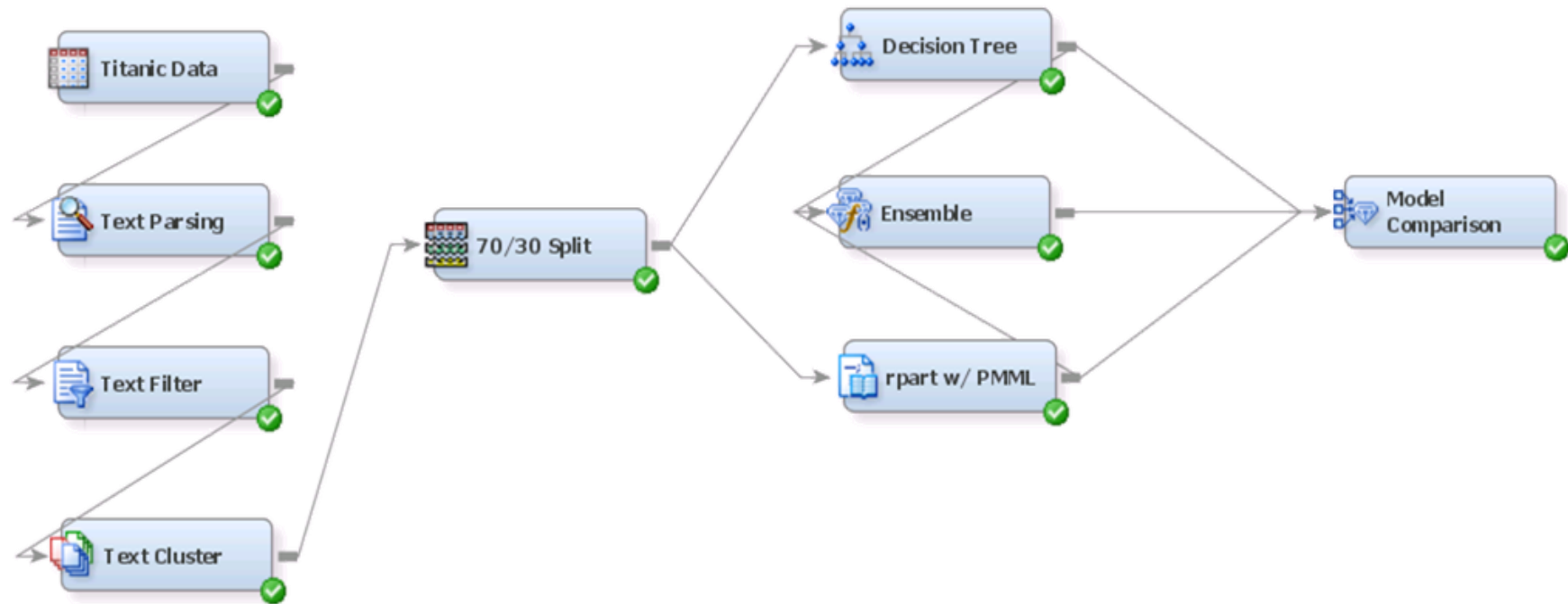
WHY IS THIS IMPORTANT?

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

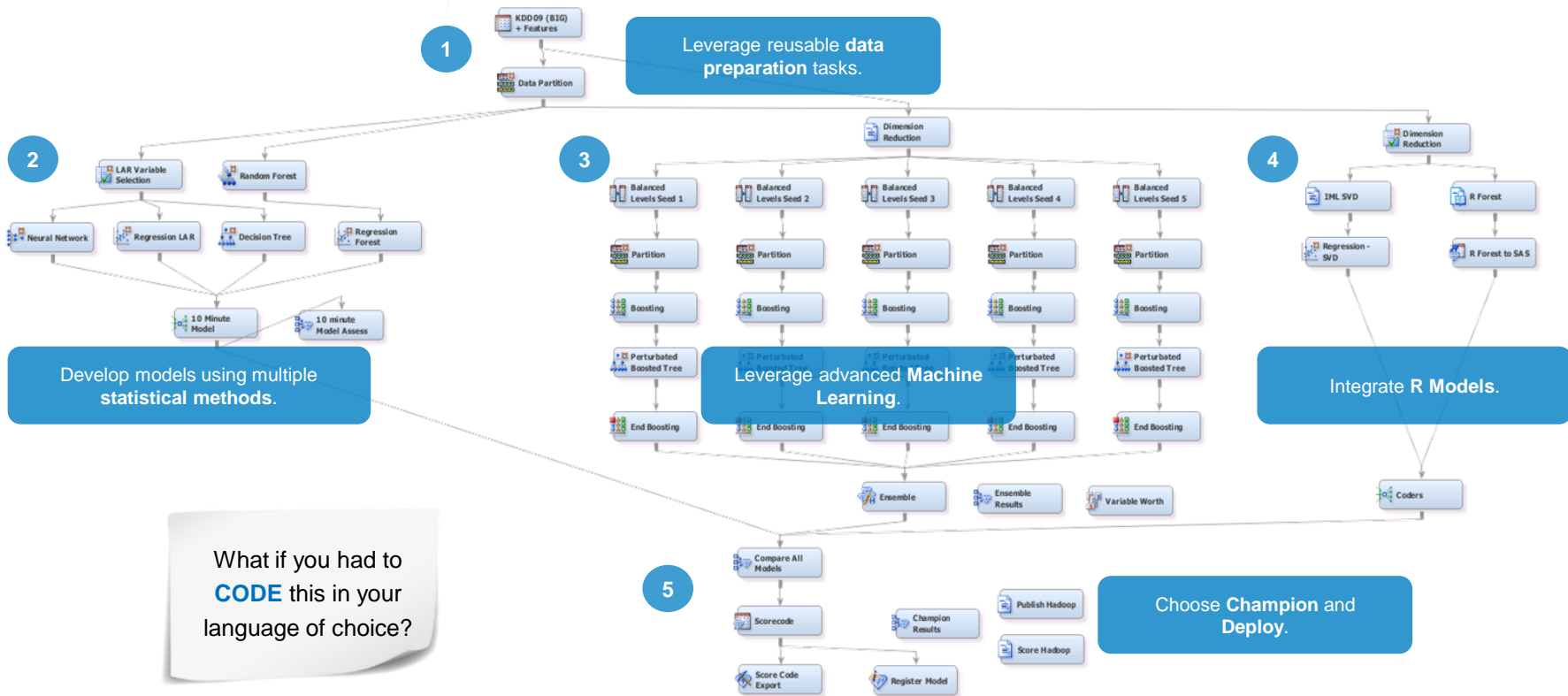
SAS automatically generates documentation capturing best practices, **promoting collaboration** and helping **reduce turnover risk**.

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.

SASPY

```
In[3]: cars.describe()
```

```
Out[3]:
```

	Variable	Label	N	NMiss	Median	Mean	StdDev	\
0	MSRP	.	428	0	27635.0	32774.855140	19431.716674	
1	Invoice	.	428	0	25294.5	30014.700935	17642.117750	
2	EngineSize	.	428	0	3.0	3.196729	1.108595	
3	Cylinders	.	426	2	6.0	5.807512	1.558443	
4	Horsepower	.	428	0	210.0	215.885514	71.836032	
5	MPG_City	.	428	0	19.0	20.060748	5.238218	
6	MPG_Highway	.	428	0	26.0	26.843458	5.741201	
7	Weight	.	428	0	3474.5	3577.953271	758.983215	
8	Wheelbase	.	428	0	107.0	108.154206	8.311813	
9	Length	.	428	0	187.0	186.362150	14.357991	

	Min	P25	P50	P75	Max
0	10280.0	20329.50	27635.0	39215.0	192465.0
1	9875.0	18851.00	25294.5	35732.5	173560.0
2	1.3	2.35	3.0	3.9	8.3
3	3.0	4.00	6.0	6.0	12.0
4	73.0	165.00	210.0	255.0	500.0
5	10.0	17.00	19.0	21.5	60.0

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	1482 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%

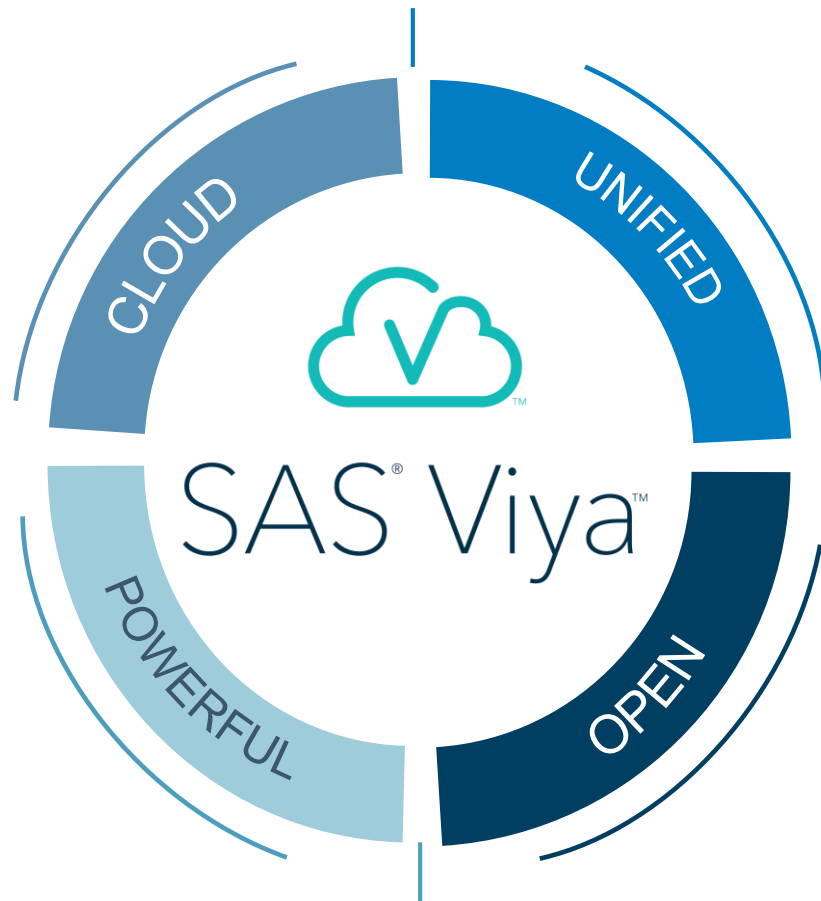
OOB vs Training

THE FUTURE IS NOW...



SOME CHALLENGES

- Computational issues with big data
- Pressure to provide scalable solutions within constrained budgets
- Ever-changing everything
- Lack of analytical governance
- Difficulty managing portfolios of diverse technologies
- Suboptimal technologies impede decision making



VIYA IS...

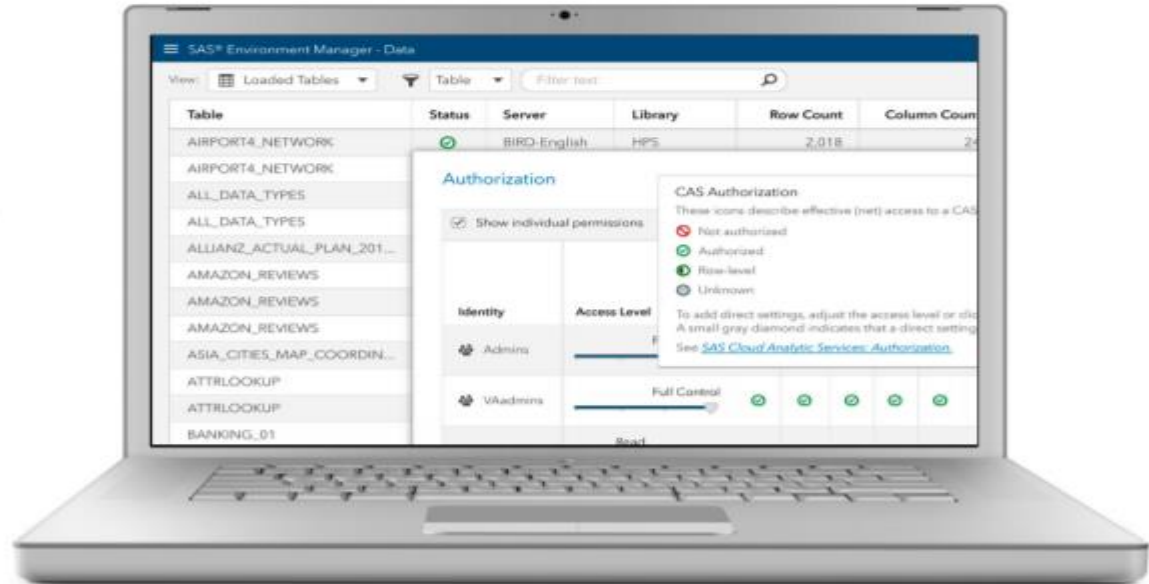
- **Unified.** A single, cohesive and governed analytics environment
- **Open.** An environment custom built for analytics that supports programming in SAS and access from non-SAS languages, such as Python, Java, R and Lua
- **Powerful.** An architecture designed with resilient distributed processing for faster answers
- **Adaptive.** A modern, cloud-ready analytics platform designed for elastic processing and easy scalability

WHY SHOULD I CARE?

- Reduce the cost of maintaining and managing multiple code bases
- Easily govern and maintain all of your analytical assets
- Scale up and out as needed
- Streamline day-to-day workloads and increase productivity
- Increase collaboration among teams

CURRENTLY AVAILABLE...

- SAS Visual Analytics
- SAS Visual Statistics
- SAS Visual Data Mining and Machine Learning
- SAS Visual Forecasting
- SAS Optimization
- SAS Econometrics
- SAS Visual Investigator



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

A leather belt with a solid steel buckle and numerous metal compartments placed along its length, the utility belt is one of the foremost items in the Batman's arsenal against crime.

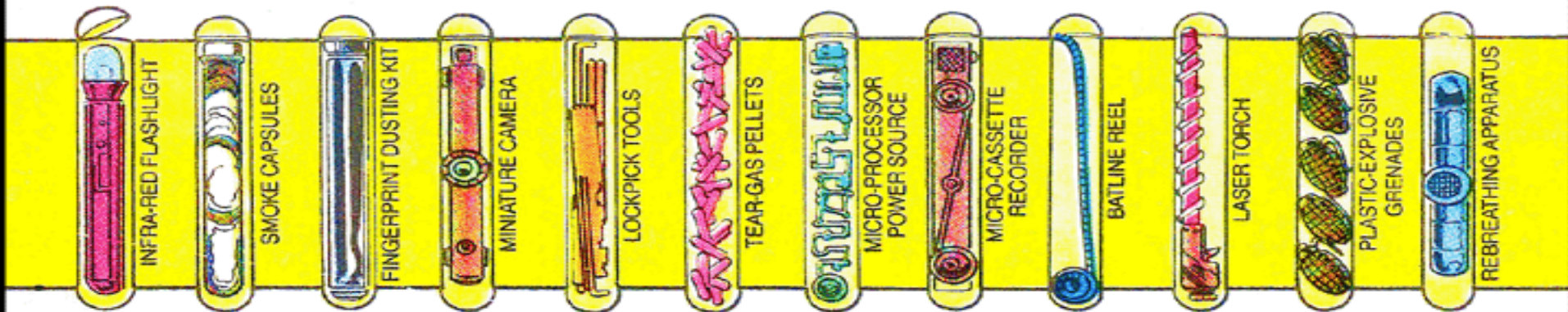
It must be noted that the Caped Crusader is continuously updating and improving the contents of the utility belt, so the accompanying illustration shows merely an average sampling of its contents.

TWO-WAY RADIO
(CONCEALED IN BELT BUCKLE)

BATARANG FITS INTO COMPARTMENT BEHIND BELT

BATLINE IS DRAWN OUT OF BELT LINING FROM SPRING-LOADED REEL

BATMAN'S UTILITY BELT



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](#)
- Viya – [Try it Now](#)

THANK YOU

MATT.MALCZEWSKI@SAS.COM

TWITTER: MALCHEW

LINKEDIN: MATT MALCZEWSKI

