

Drive Analytic  
Innovation  
Through SAS®  
and Open Source  
Integration

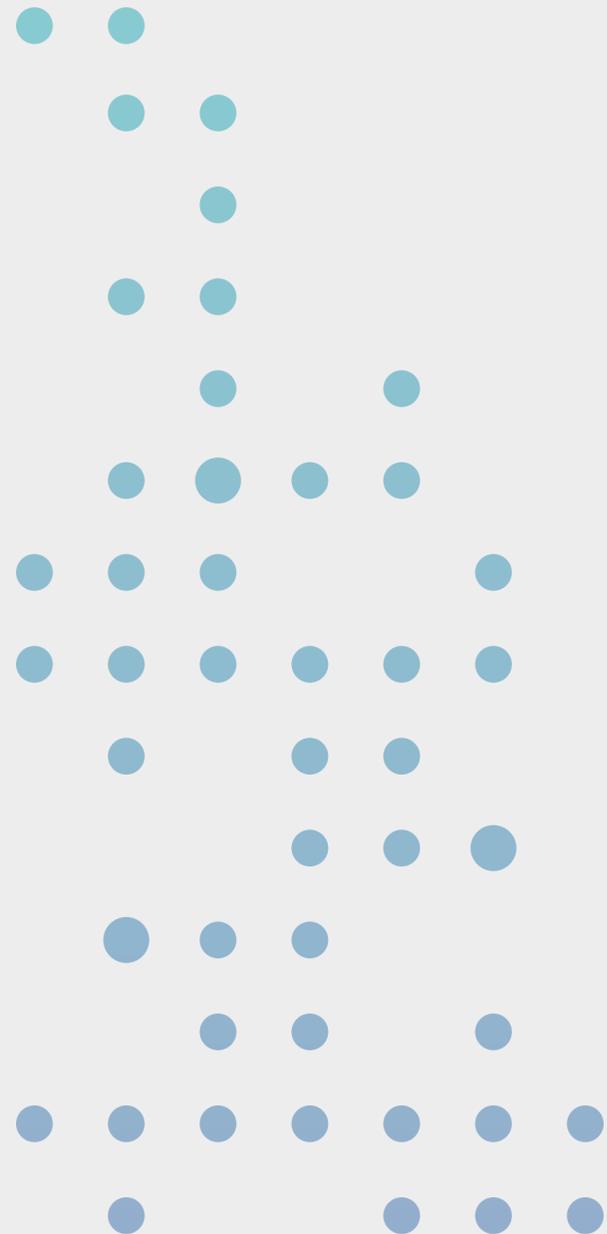




## PURPOSE OF THIS E-BOOK

This e-book is intended for those who want to learn about how to use SAS with open source to drive analytic value and achieve trusted decisions. Whether you are a SAS user interested in dabbling in open source or an open source user who wants to work with SAS, this e-book will help you get started.

# CONTENT



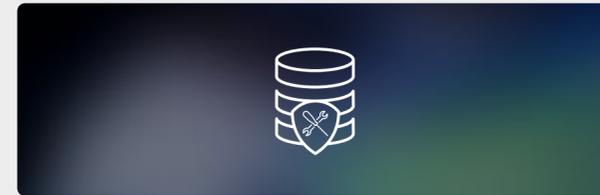
WHAT DRIVES ANALYTIC INNOVATION TODAY?



INTEGRATION ACROSS THE ANALYTICAL LIFE CYCLE



BUILDING MODELS



MANAGING MODELS



DEPLOYING MODELS



NEXT STEPS



# WHAT DRIVES ANALYTIC INNOVATION TODAY?

# AI & ANALYTICS CHALLENGES TODAY

*Four main challenges that influence business outcomes, increase the cost of analytics and slow the pace of innovation.*

**Data:** In many organizations, we need more collaboration between businesses, analytic teams, application developers and IT operations. These teams often work with data in silos and end up duplicating efforts, failing to integrate, or missing opportunities to deliver value from data.

In addition to siloed efforts, data scientists are also faced with ever-increasing volumes and speeds of data. And the reality is they're expected to answer questions just as fast as - or faster than - before.

It's important that we're using the right data and the right techniques to ensure optimal outcomes.

**Technology:** What we learned in college is that data doesn't magically appear in these analytically ready formats. There's a ton of transformation work required to take relational data and put it into formats that support model development.

It comes down to how quickly you can explore and identify the right algorithms, and ultimately train one or more models to achieve your analytic goals.

Okay, so you've created a model. But this won't affect business outcomes until it's deployed and integrated into an operational environment.

The longer this entire process takes, the greater the risk that your models in production are working on assumptions in the data that are no longer valid.

If the market conditions have shifted, how quickly can you determine if your models are deteriorating? How quickly can you retrain? How quickly can you redeploy?

Data scientists must answer all of these questions and more.

**Process:** When it comes to development, data scientists are usually not interested in spending time with mundane activities such as documentation, lineage, traceability, versioning, or explainability. These provide visibility into what models are doing and how they are performing.

These factors increase the likelihood of a failed model experiment. One that never makes it into production.

But putting in place an effective process to ensure governance and transparency will increase your overall efficiency, productivity, and ultimately... repeatability.

**People:** An effective process only matters if you can put people in a position to be successful and give them access to the right tools to support innovation. Data scientists must be able to rapidly adapt and refresh these tools over time in a consistent way allowing them to spend more time applying the tools to achieve value rather than maintaining the tools.



# SEAMLESS INTEGRATION WITH OPEN SOURCE AS SUCCESS FACTOR

*There's a vibrant ecosystem of choices available for data scientists.*

This can span languages like SAS, Python or R, integrated development environments, deployment technologies, virtual machines, Kubernetes and more.

However, this can create option fatigue, resulting in an inconsistent landscape that makes it difficult to scale analytics. There is increasing recognition among companies that it may be helpful to draw in other technology to integrate open source and other software - such as analytics platforms like SAS® Viya® - to create interoperability and utility from open source.

## EXPECTATIONS AND BUSINESS NEEDS



# WHY USE AN AI, ANALYTICAL AND DATA MANAGEMENT PLATFORM?

In her article [“Six trends that will influence Data Science practitioners’ priorities in 2021”](#) Marinela Profi, Product Marketing Manager for Data Science at SAS, mentions “an ongoing reliance on open source, and a focus on its integration” as one of the trends in 2021.

A critical success factor for these platforms is not limiting the languages that data scientists or IT developers can use, including open source. They also need to integrate via open APIs and ensure endless scalability.

Organizations are depending on open source (like Python and R). And it's proven to be a valuable tool for specific analytical tasks. However, when it comes to building a long-term analytic strategy - a self-sustaining

and seamlessly integrated life cycle, open source is creating massive challenges around coordination, integration and as a result delivering business value.

There is increasing recognition among companies that it may be helpful to draw in other technology to integrate open source and other software - such as analytics platforms like Viya - to create interoperability and utility from open source.

No matter which technologies you choose, you must work on simplifying and automating this process. You need to figure this out for your own ecosystem. And when you do, the payoff is huge.

# HOW DATA SCIENTISTS WILL BENEFIT

Data scientists, according to interviews and expert estimates, spend 50 percent to 80 percent of their time mired in the mundane labor of collecting and preparing unruly digital data before it can be explored for useful nuggets. And, what's more, once they have built the model, deployment can be another nightmare.

Citizen data scientists will bring their work to the business faster, being appreciated by their organization and recognized as innovators.

## BENEFITS OF USING OPEN-SOURCE WITH SAS:

- Harness the cloud-native high-performance architecture of SAS Viya
- Massively distributed parallel processing for endless scalability
- API first development strategy coupled with containerized microservices architecture
- Automated feature engineering with ML powered data preparation
- Use Python or R directly with SAS or integrate SAS into applications using REST APIs
- Run open-source models without recoding
- Container Deployment of SAS and open-source models
- Deploy SAS, R, Python models in batch, streaming, cloud or edge device
- Use automation in continuous integration and continuous deployment pipelines to manage code artifacts
- Composite AI to combine different AI and analytics techniques in the same environment (i.e. computer vision and optimization)
- Data Lineage and Auditability so you can see how data moves through the entire system
- Monitor open-source and SAS models in same repository and build customized performance reports
- Business rule and formal decisioning integration
- Go from code to point-and-click, and then back to code if desired
- Ensure explainability of your data science projects through natural language-powered built-in report



# **INTEGRATION ACROSS THE ANALYTICAL LIFE CYCLE**

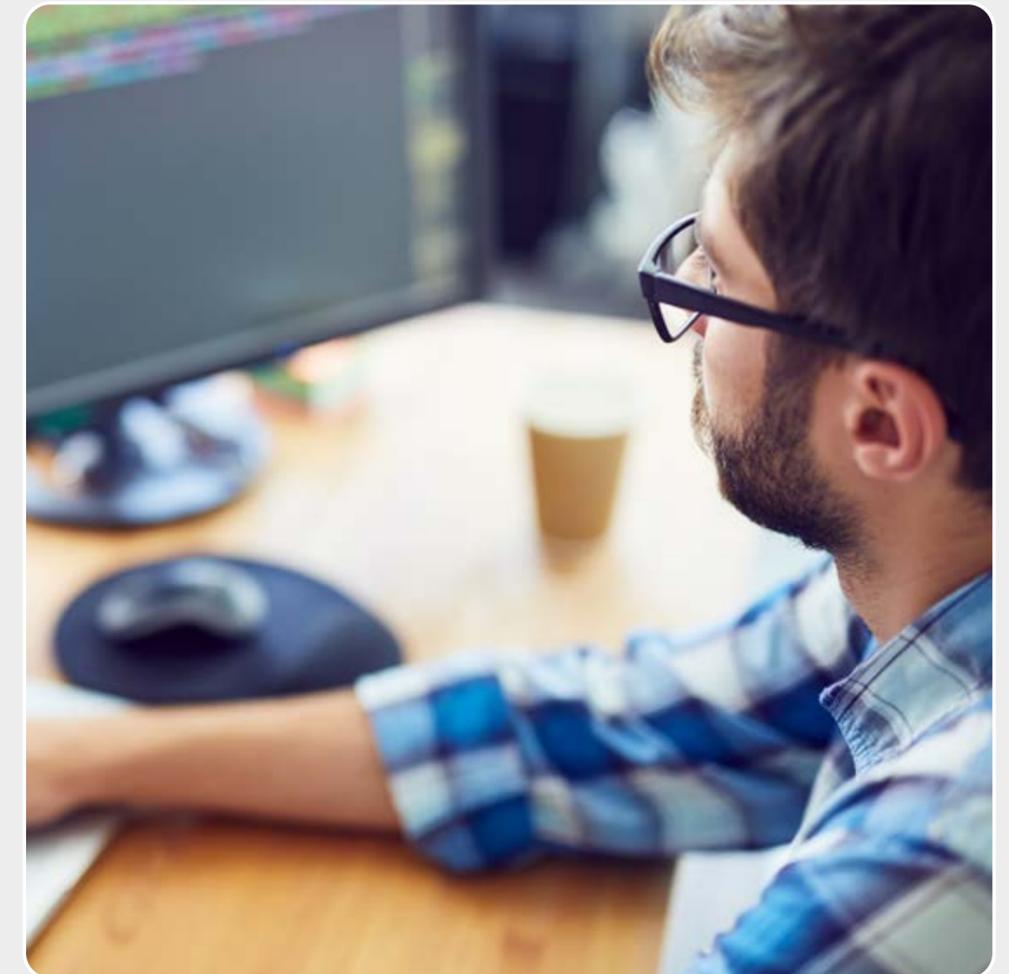


# INTEGRATION ACROSS THE ANALYTICAL LIFE CYCLE

As the analytical life cycle is a broad topic, for the rest of this booklet, we will focus on a specific perspective: The Model Life Cycle.

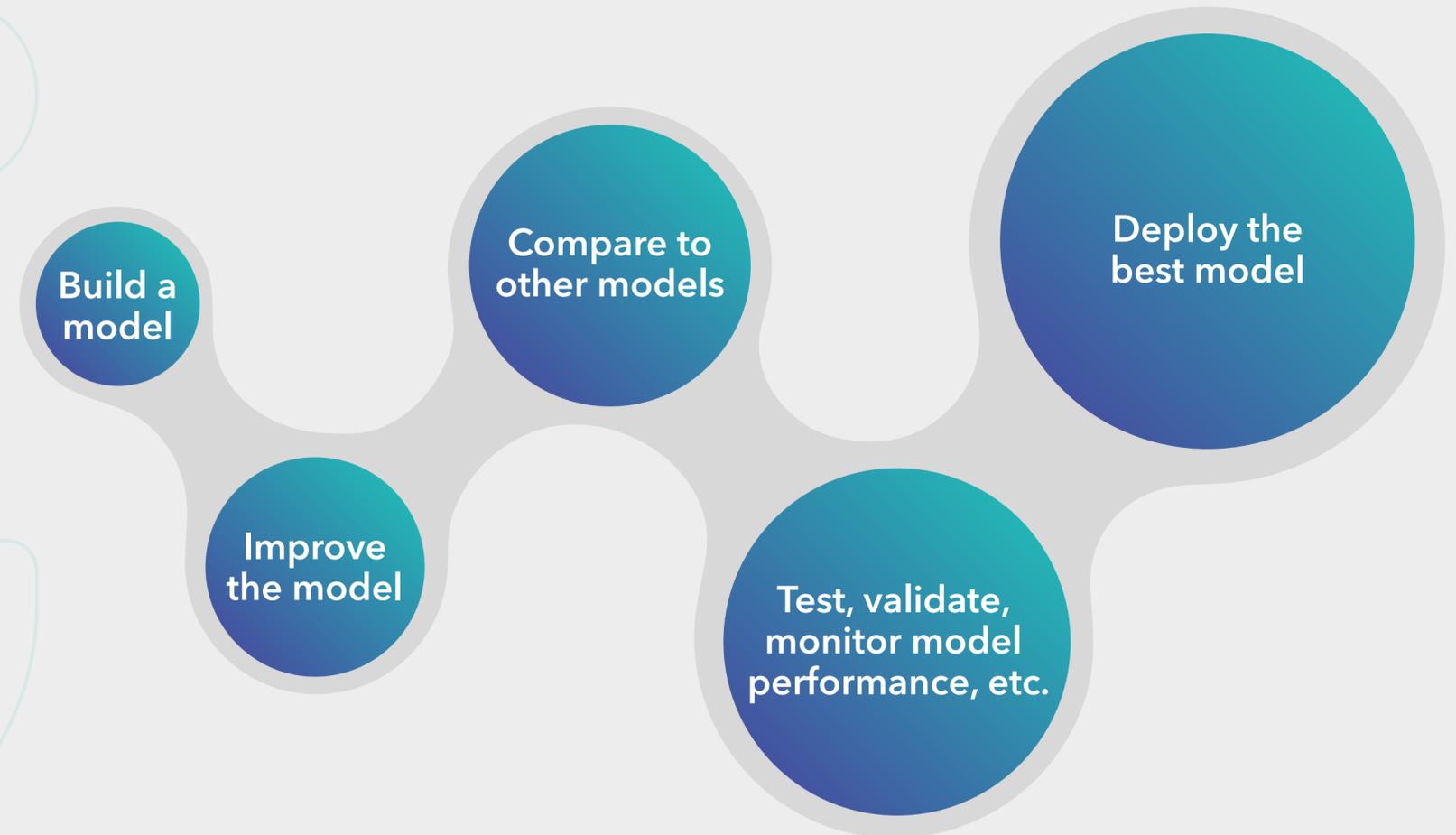
The Model Life cycle is a methodological process that data scientists and practitioners apply to build, manage and deploy analytical models to generate business value.

The Model Life Cycle is where we see the most interactions and integration possibilities between open source technology and SAS technology, and hence will be our focal point of discussion.



# MODEL LIFE CYCLE PROCESS FLOW

Note: Modelling is an iterative process. Continuous model building, testing and monitoring is needed for any healthy model life cycle.

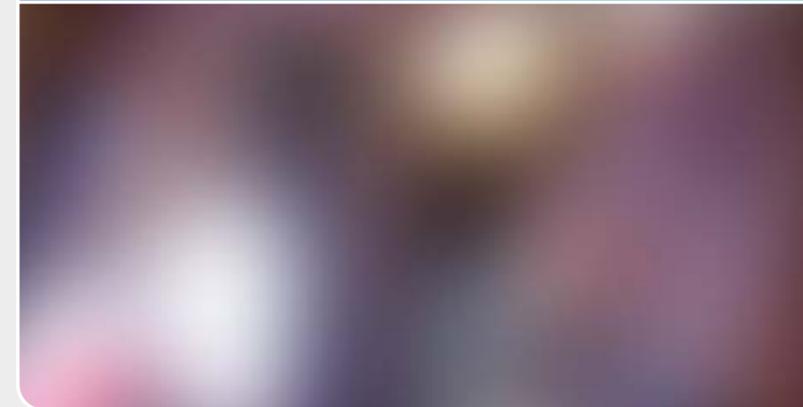


# INTEGRATION IN MODELING

Integration in modelling provides more capabilities and flexibility to users. The next question is, how can integration be achieved?

To answer this, we will look at two important areas before getting into the more technical aspects of integration.

## Empowering Open Integration



**Empowering Open Integration With Viya:**  
Introduction and Overview

- 1 How does integration occur?
- 2 How do I integrate in the model life cycle?

# HOW DOES INTEGRATION OCCUR?

There are two approaches to integration: going from open source to SAS and from SAS to open source. The benefit of having two options is that it opens and extends analytics in your organization to multiple types of users, regardless of their technical background.



## OPEN SOURCE TO SAS

For users who have an open source background looking to explore SAS capabilities right from the open source interface.



## SAS TO OPEN SOURCE

For users who want to utilize both SAS and open source assets in SAS' visual interface.



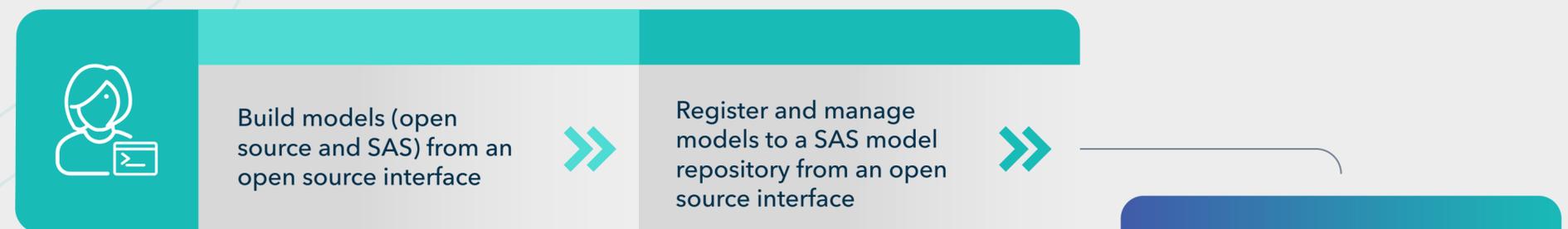
## VALUE

Users can share and collaborate irrespective of which approach they've taken.

# HOW DO I INTEGRATE IN THE MODEL LIFE CYCLE?

SAS can integrate with open source technologies at any point of the model life cycle via APIs. Your integration path is dependent on how you prefer to integrate, and where in the model life cycle you would like to integrate.

## OPEN SOURCE TO SAS / START HERE!



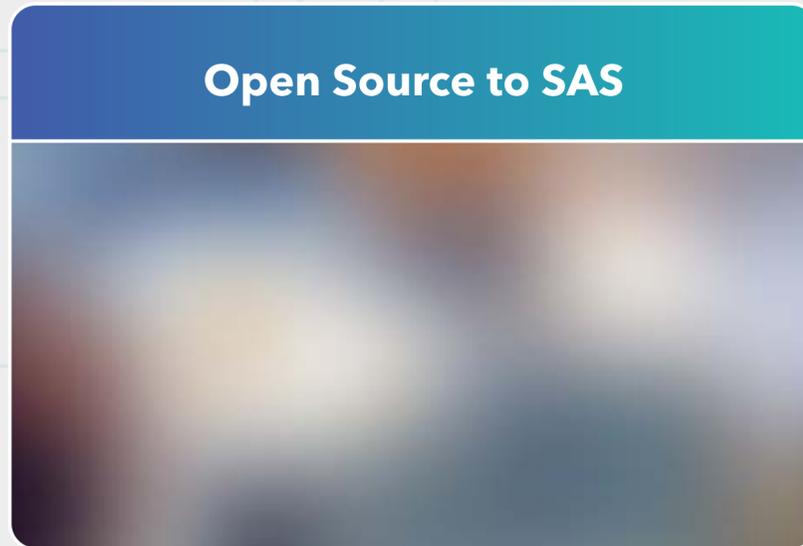
## SAS TO OPEN SOURCE / START HERE!



Publish models into a range of SAS and open source execution engines for batch, single call or real-time processing



# BUILDING MODELS



Open Source to SAS

**Building Models 1:**  
Open Source to SAS via SWAT and DLPy

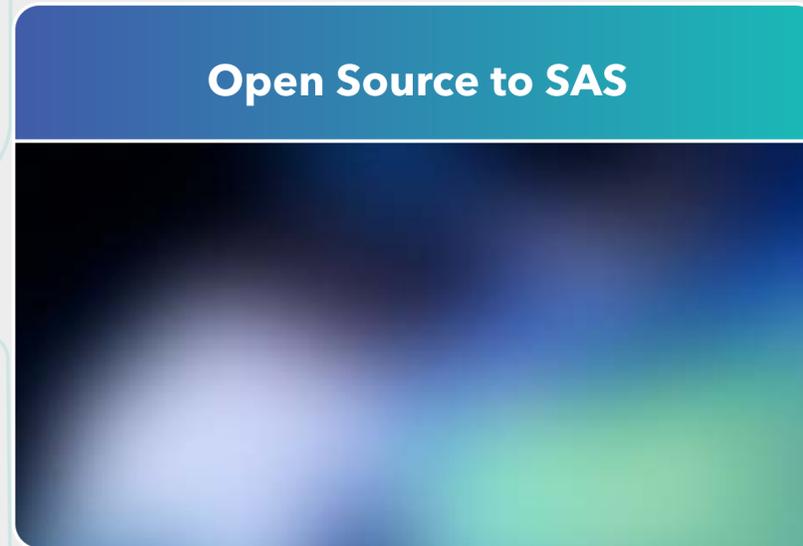


SAS to Open Source

**Building Models 2:**  
SAS to Open Source via Model Studio

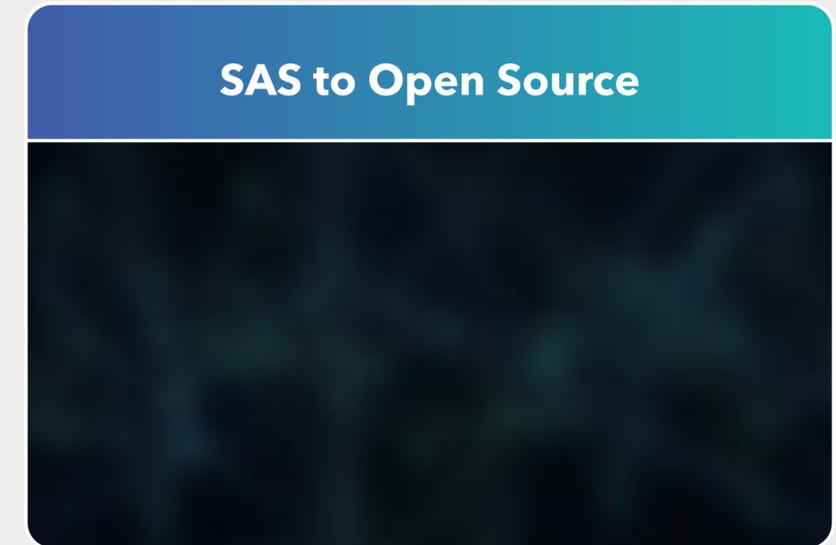


# MANAGING MODELS



Open Source to SAS

**Managing Models 1:**  
Open Source to SAS via SAS Ctl/Pzmm



SAS to Open Source

**Managing Models 2:**  
SAS to Open Source via Model Manager

# I HAVE MY MODEL RUNNING, WHAT'S NEXT?

Once you have a Python or R model created, you can then register those models to SAS Model Manager to compare, evaluate and monitor the performance of the models before publishing them to a test or production environment.

Model management is an important step after building models. The benefits of using SAS Model Manager include:

It ensures you are consistently running your best model at any given time **to minimize the impact of model decay on business.**

Allows for complete traceability and analytics governance through a centralized model repository, and version control **for higher governance of a model workflow.**

Can deploy models with just a few clicks, both in batch and in real time **for quicker value realization.**

*Note: To manage models in SAS, you would need to license SAS Model Manager on your environment.*



# DEPLOYING MODELS

# MY MODEL IS READY! WHAT'S NEXT?

## *Understanding the open source ecosystem*

Now that you have your model running, it's time to deploy it. Deploying your model simply means to have your models be used in a production environment.

This is usually done by the DevOps or IT team in your organization.

In these videos, we will demonstrate deployment methods using SAS tools, which will make deployment for both SAS and open source models more robust for production environment using a combination of SAS and open source technologies.

**Deploying Models:**  
SAS Cloud Analytic Services

**Deploying Models:**  
SAS Micro Analytic Service

**Deploying Models:**  
Docker/Kubernetes

**Deploying Models:**  
SAS Event Stream Processing



# NEXT STEPS

By this point, you should have a good grasp of why SAS integrates with open source, the benefits of doing so, and the various ways in which you can do it.

However, to have a successful integration, you need to give careful consideration of whether an integration is needed in your context, and if so, to what extent.

Here are some sample questions to ask:

**What is the complexity and/or volume growth of your data?**

**What is the mix of skillsets in your team/organization?**

**What are the analytics problems that you are trying to solve? Are they big? Are they complex? Are they urgent?**

**How does your IT environment operate? Will you generate more technical debts?**

For further information please visit the [sas.com/viya](https://sas.com/viya) page.