Paper 4708-2020
Open Source Python & R Lang on our SAS® Shared Grid
Thomas Ball, Christopher Hughes, NatWest Group

ABSTRACT

In November 2018 the requirement was received to enable Open Source Python and R Lang on our SAS Shared Grid Environment to provide enhanced functionality for our Data Science community.

Our platform is one of the Biggest SAS Environment's in the world.

In May 2019 the functionality was enabled on our SAS platform following successful **User Acceptance Testing. This functionality along with SAS' interoperability with** other data source types has provided new opportunities for our business users.

INTRODUCTION

This paper gives an insight into the process we experienced when introducing Open Source Python & R Lang onto our SAS Platform.

Engaging with Business Users and Infrastructure Colleagues we faced several challenges in this deployment with the key parts being the desire for Agile package deployment whilst also minimizing the risk of deploying packages to our platform.

WHY PROVIDE OPEN SOURCE CONNECTIVITY?

- Python is one of the Most Popular coding languages on the Planet! And R Lang usage is increasing.
- The Software comes at No License Cost.
- This added functionality provides new opportunities for our user base.
- Open Source functionality enables us to build on the insights we have already developed using our SAS toolset.
- The 2 languages can be harnessed in SAS Enterprise Miner and deployed using SAS Model Manager enabling a true end to end Modelling Life Cycle with Open Source Integration.

OUR PLATFORM

NatWest Group's SAS Shared Grid is one of the world's largest SAS Platforms. Our environment includes SAS tools such as Enterprise Guide, Enterprise Miner, Model Manager and Visual Analytics.

Our user base includes Data Scientist's, Data **Analyst**'s and **Viewer's.**

THIS ENCOMPASSES…

- 12,000+ REGISTERED SAS USERS
- 10 SAS PRODUCTS
- 3 METADATA/MID-TIER SERVERS
- 7 GRID SERVERS
- 59 LASR SERVERS

DATA SOURCES

We have vast amounts of data in many different technologies that include:

- Teradata
- SQL Server
- Oracle
- Hadoop

OUR JOURNEY TO OPEN SOURCE ON SAS

Our journey to Open Source enablement began in Q3 2018, with our SAS data science community expressing a desire to expand on their capabilities – the start point of this project.

Through project management using focus meetings, we sought formal requirements and identified timelines for items such as delivery into test. Using this coordinated approach, we were able to organise test enablement with our infrastructure colleagues and user testing with our focus group.

This approach resulted in us identifying key technical challenges, which required further technical discussion, assessment of risk and the requirement for formal sign off.

This successful approach led us to the delivery of Open Source to our SAS Grid Platform in Q2 2019 with which many future opportunities are now available.

THE PROJECT TIMELINE

2018

- In Q3 2018 the requirement is formally received from our data science community in the bank which is a fairly new team at this stage.
- In Q4 2018 we start working with our data science community to enable this functionality on our SAS grid to look the actions involved.

2019

- In Q1 2019 we undertake non-functional testing and user acceptance testing on our test SAS environment.
- In Q2 2019 we enabled R and Python on our SAS shared grid after test signoff.

THE TECHNICAL CHALLENGES WE FACED:

'HOW CAN WE ENABLE OPEN SOURCE SECURELY?'

XCMD Enablement
It became apparent after analysis alongside our SAS infrastructure team colleagues that the SAS XCMD system option would need to be enabled.

This presents challenges as the XCMD option when enabled provides users with the ability to run OS level terminal commands from SAS.

After extensive discussions considering the risk of enablement it was felt that the OS level permissions in place provide enough control to minimise the risk of this enablement

More info on XCMD can be found here:
https://documentation.sas.com/?docsetId=hostunx&docsetTarget=n02699r064oqccn19vblte8xzr8m.htm&docsetVersion=9.4&locale=en

Open Source Package Safety Checks
With the deployment of Open Source we also identified that we needed to consider how we are actually able to support users in their appetite to use of the latest Open Source package versions whilst ensuring that any packages are reviewed and signed off as safe to minimise the risk to the platform.

This required plenty of discussion and consideration about Service Level Agreements on time from request deployment (to manage user expectations), the process of checking packages (who / when / how) and the actual deployment into production. This area is still evolving.

THE FUTURE

We aim to integrate Open Source into the model life cycle using SAS Model Manager. Combining Open Source with the machine learning and model management capabilities that Model Manager has to offer.

We anticipate an increase in usage of Open Source on our SAS platform not only as an enhancement to our data science teams capabilities yet also as an opportunity for our wider userbase to increase their knowledge in different software technologies.

Our process of deploying Open Source packages is a space which is evolving. Collaborating with other technology team colleagues we are working on creating a safe but efficient method of deploying packages to the SAS servers. With our aim to try and meet our users desire for agile package deployment whilst retaining the integral safety of our platform.

This is all part of our aim of enabling our users in their use of our platform and our Open Source and SAS environment contributing towards **NatWest Group's** future enterprise, learning, climate and sustainability goals!

CONCLUSION

Open Source integration Enables our Data Science community to improve their insights from our vast amounts of Data.

The Interoperability of SAS provides colleagues with the quite unique opportunity to work with Python & R Lang Alongside SAS code using our vast amounts of different Data Source Technologies.

This combined gives our community opportunity to develop new skills, enables huge potential to develop new insights whilst also moving us into the area of model automation and machine learning. Meaning the sky really is the limit for our data!

RECOMMENDED READING

Open Your Mind: Use Cases for SAS® and Open-Source Analytics

https://support.sas.com/resources/papers/proceedings17/SAS0747-2017.pdf

SAS® and Open Source: Two Integrated Worlds

https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2019/3415-2019.pdf

How to Win Friends and Influence Platforms: Solving Non-Technical Challenges of SAS® Estates

https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2019/3716-2019.pdf

How many Genies can you fit in one bottle? – Multi Tenancy and all the wishes you want

https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2019/3717-2019.pdf

CONTACT INFORMATION

Christopher Hughes
The NatWest Group
christopher.hughes@rbs.co.uk

Tom Ball
The NatWest Group
thomas.b.ball@rbs.co.uk