



WHITE PAPER

Migration from SAS to Open-Source in the Data-Driven Era



Introduction

Commercial decision-making is increasingly the driver for change in modern analytic practice and one such decision is in the choice of open-source or proprietary software. Open-source software such as R or Python, is “freely available and may be redistributed and modified”, with the code contributed to and supported by, the community. Proprietary software, such as SAS, is developed and maintained by a company, who charge a licence fee to end users.

SAS has been long established as the leader in the commercial analytics space, however recent years have seen an evolution in, and growing commercial adoption of, more flexible and cost-effective open-source alternatives.

SAS migration projects are often a consequence of a broader company-wide transformation around the use of data to drive strategic decision-making.

The reason for trends in migration, associated challenges and general tips for driving this change are shared here.

About SAS technology

SAS is a long standing, comprehensive platform widely used by companies across a range of industries. Easy to use, it offers the user a range of statistical functionalities and a straightforward GUI; SAS is easy to learn and backed by technical support. However, unlike the free open-source technologies, the usage of SAS means that organisations are committed to costly licencing contracts without necessarily having access to much of the functionality required by modern data science practices.



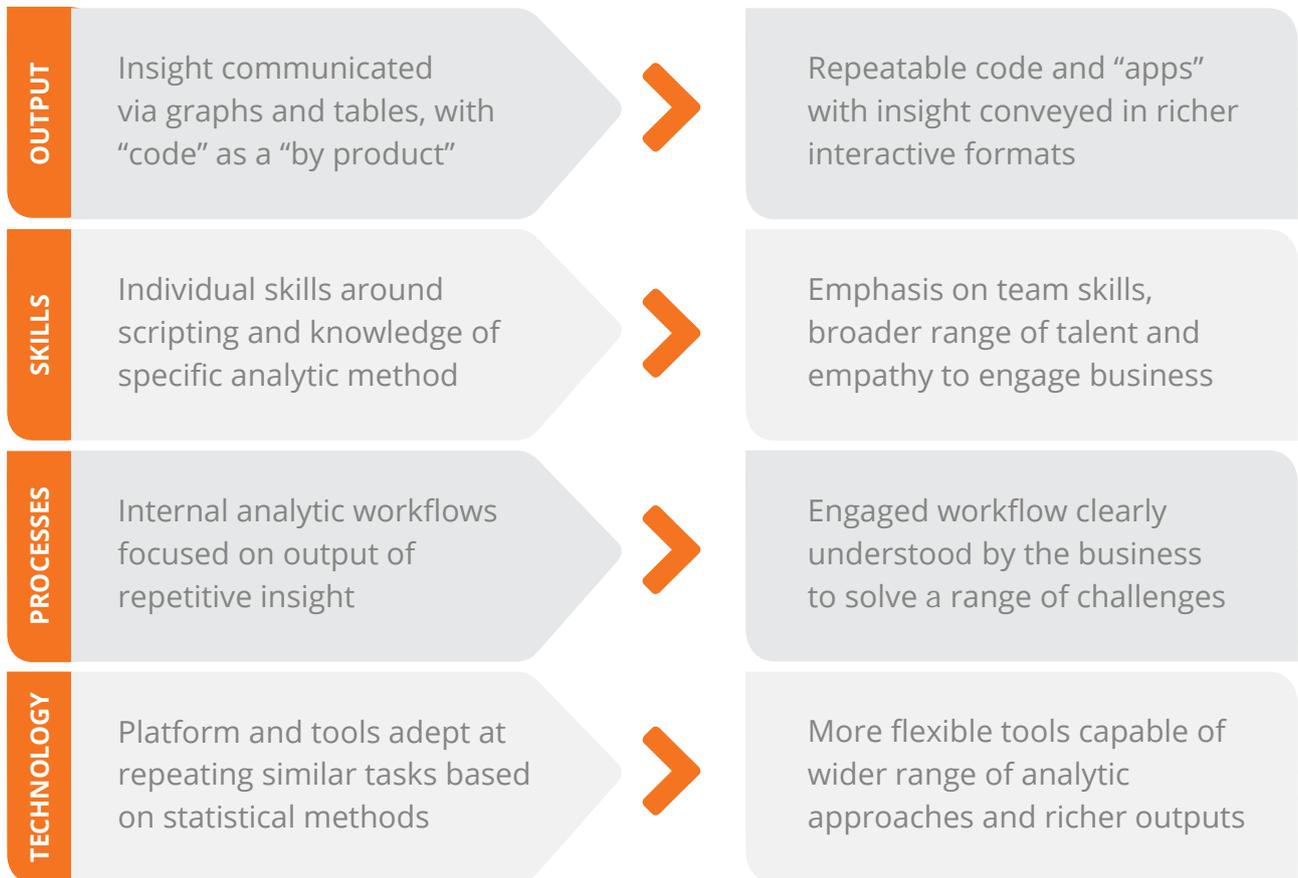
Modern Analytic Practices

The visibility and remit of analytics has changed over the years as companies increasingly understand that future success depends largely on the use of data to make optimal business decisions and drive efficiencies. This may require a change from reactive statistical analyses around a narrow range of challenges to more proactive open-source data science methods to drive business decisions in a repeatable and scalable manner. Such a change often impacts working practices and workflows across the organisation including the technology platform to render it fit for purpose in the modern analytic world.



REACTIVE

PROACTIVE



The driving force for change

There are a number of contributing factors that may induce a shift of analytic platform away from SAS to open-source technology.

These include:

- ✓ The desire to reduce high annual licencing costs
- ✓ The desire to adopt modern data science techniques and methodologies
- ✓ The ability to formulate efficient data science practices with delivery-focused methodology
- ✓ To entice the modern data scientist to join the organisation

The desire to reduce high annual licencing costs

It is quite well documented that SAS is expensive and used routinely by large organisations. Part of the expense is because of the way SAS has been procured historically, piecemeal by both business and IT, the “fractured” nature of SAS adding to the costs.

It might be assumed that cost is nearly always the primary migration driver, though in our experience of SAS to open-source projects, this is often not the case. However, IT budgets have come under increasing pressure over the last few years and software licencing is an obvious area where cost reductions can be achieved.

It's clear then that the adoption of open-source technology is hugely cost-effective in comparison to SAS, offering an immediate cost-saving advantage.

The desire to adopt modern data science techniques & methodologies

A primary reason for migrating away from SAS to R or Python is the quest for the latest functionality and access to richer capabilities. Historically there was a feeling that a specific statistical routine was not widely accepted until it was implemented in SAS. Such days are long gone and modern analysts or data scientists need access to cutting edge models and approaches that are just not incorporated in SAS quickly enough.

SAS is a large piece of software designed to perform standard ETL and analytic processes in a repetitive manner. SAS is also used interactively, but challenges can occur when you try experiment and innovate. Generally, SAS is great when you want to do standard things but can be painful when trying to do something it was never designed to do.

Today there are a vast range of support options and products available that allow analytics teams to deploy R in production environments for example, with a sense of safety and comfort that allow a more flexible environment and an all-round scalable solution.

Businesses looking to get one step ahead of their competition by utilising the latest and greatest algorithms simply don't have the time to wait for the next SAS release.



The ability to formulate efficient data science practices with delivery-focused methodology

A SAS to open-source migration may first appear as a simple technical project. However, it can impact a range of technical and, importantly, cultural factors.

Imagine you take a Software Developer with a formal programming education and ask them to transition from (say) Java to C# - maybe they'll be frustrated but they understand they have a "generic" skill called "programming" that they are happy to apply in Java today - so they can go from this base to C# and find familiar structures and concepts,

A SAS user typically doesn't have a programming background, so they don't recognise they have a "generic" ability to create code - all they know is SAS, and that's their only mechanism for doing their job. A migration from SAS to another system is a very worrying prospect for many long-term SAS users. Strong objections to a migration are often driven from this perspective, so critical to success is an understanding of the context of the migration, and an empathy with the user community.

This must include a large amount of appropriate communication, training and support for any technology adoption to be successful.

Another cultural challenge is the fact that legacy SAS usage may be based around set coding standards and procedures which can cause barriers when adopting new technologies. Similar standards and procedures are required for R or Python, so care must be taken to understand how R /Python code is to be developed, tested, and deployed to end users.

To entice the modern data scientist to join the organisation

It's clear that the need to attract, resource and upskill data science teams has become far more pronounced. With the skills gap growing ever wider, employers need to take positive measure towards retaining leading talent.

R and Python are recognised leading technologies for data science today and have been widely adopted within many major organisations. There are an estimated 2m R users in the world today and 8.2m Python users worldwide and these numbers are set to grow.

Universities have adopted R as their primary analytical programming language for mathematics, science and statistics courses meaning graduates are joining the workplace fully skilled and suitably ready to deploy their newfound skills. For GCSE age school children, Python coding is now on the National Curriculum, as part of the Computer Science course.

By contrast, SAS skills are declining as universities shift away from teaching it and unlike other languages, SAS has limited free online resources which means costly training. It is simply not cost effective or sensible for any organisation to take people already skilled in R or Python and train them to use SAS.

Such grassroots formation in open source languages is fuelling the trend away from SAS. This is starting to take shape with the emergence of centralised data science and analytics capabilities in open-source technologies across a much wider range of sectors.

Project Execution

As with any major projects, there are potential issues and pitfalls involved in converting from a single tool such as SAS. Typically, organisations have built up large amounts of legacy code over the years; much of this may be undocumented and untested meaning a conversion project can be a huge task. SAS teams have typically invested years of effort building large libraries of SAS macros and code. The move to open-source is not something, therefore, which can happen overnight.

In choosing a strategic move away from SAS, companies must carefully consider a variety of factors including re-training staff and crucially, the migration of what is typically a very large code base developed over many years by a variety of analysts.

How do you migrate 20 years' worth of SAS code to R/Python?

How do you begin tackling these challenges and delivering your migration to open-source?

Here are our 6 key components to a successful SAS migration project

STEP 1

STRATEGY: Create an effective strategy for migration, covering each of the following 5 elements, with a clearly defined roadmap and success criteria



STEP 2

SKILLS: Upskill SAS users in the chosen open-source language, and any other aspects of more modern data science methodologies and approaches



STEP 3

DATA: Migrate any data stored in specific SAS formats onto more accessible and appropriate modern data infrastructure



STEP 4

PLATFORM: Move from a SAS system to a trusted, R-centric or Python data platform, with all technical elements required for modern data science



STEP 5

PROCESS: Change workflows to represent new ways of working with the business, within the analytic team, and with R/Python itself



STEP 6

ACTION: Convert legacy SAS code to R/Python and provide evidence and assurance to stakeholders around the newly migrated code base



Converting to Open Source: A Case Study

Whilst the need for a specific, documented strategy can vary, SAS to open-source projects often require learning from people who have been on this journey before. The following is a specific case study that saw a company take a decisive step into data-led open-source migration.

Mango undertook a review of the SAS Estate of a directorate of a UK Government department.

This project initially outlined the steps needed to migrate away from SAS with Mango subsequently undertaking this migration work.

Aims, Objectives & Scope

To outline a clear pathway for migration away from SAS in order to save c£1m pa on licence fees, Mango analysed 20 SAS applications consisting, in total, of approximately 1.5m lines of code.

Methodology

- Use of SASMap to analyse the code <https://github.com/MangoTheCat/sasMap>
- In depth discussions with the application users to understand their requirements and what additional functionality they needed
- Review of the client's current Technology & Skills landscape
- Analysis of SAS licenses to understand how costs could be incrementally saved
- Recommendations for moving forward in a pragmatic and cost-effective way

Outcome

The recommendations were positively received. Mango were awarded a contract to undertake the migration work over an 18 month period, which resulted in both cost savings in licence fees and an increased analytics capability.

Next Steps

The creation of a pragmatic, practical strategy aligned to internal business objectives, helps to identify the key roadmap for a migration. Some essential elements might include:



DEFINING THE BEST APPROACH TO TAKE

Discuss with Mango the best approach to take utilising our experience of successful migration projects. Our data science team will work with you to independently guide you on the right approach for a successful migration.



UNDERSTAND YOUR BUSINESS CASE AND SUCCESS FACTORS

Why are you looking to do this?
What are the key drivers?
Do you have buy-in from key stakeholders in the organisation?



AUDIT YOUR ESTATE

Consider allowing us to audit your estate. This approach will help you get a better understanding of your existing code base. We can look at such points as: What is the current status of the analytics portfolio and its currency (e.g. existing SAS IP)? Of the existing analytic IP, what needs to be migrated, re-written or discarded? What technologies are best suited to the modernization program (e.g. R, Spark, Python)? What constraints exist (e.g. financial, technical) that would hamper the integration and adoption of any analytic technology into the overall roadmap?



CREATE POC'S TO VALIDATE BENEFITS

Look to create POC's (Proof of Concept) in order to test out the approach and add value.

Conclusion

There are many defensive and proactive reasons for a migration away from SAS to open-source, whether this is cost reduction, access to richer capabilities, more flexibility for experimentation or simply the availability of technical skills. Frequently, however, a migration is a consequence of a broader data-driven transformation and a business wide adoption of modern analytical practices. Whatever the drivers for the migration may be, the considerations and steps outlined above provide an essential framework for a successful transition and are based on experience and expertise honed from many such projects.

About Mango Solutions

Mango Solutions is a data science consultancy that specialises in enabling a data-driven culture within businesses, helping them derive value from data science and data analytics initiatives. It does this via a collection of products and services, delivered by Mango's team of data scientists and data engineers, which help companies use advanced analytics to create operational acumen that improves business performance.

Services include strategic advice that addresses fundamental business challenges to transform data into a business asset; data science solutions aligned to driving use case value; training and upskilling through educational programmes and capability assessments and data engineering and data platforms to create an analytic environment.

Mango also offers products for regulated open-source development, including ValidR, ModSpace, Navigator Workbench and Data Science Radar to identify and build world-class data science talent. Visit www.mango-solutions.com or follow [@MangoTheCat](https://twitter.com/MangoTheCat).