Hybrid analytics: why and how to make them work for you

Key insights to optimise operational value, reduce business risk and propel your career

By John Spooner, Chief Data Scientist, SAS UKI

As organisations become increasingly reliant on data-driven insights for their efficiency, responsiveness, productivity and competitive advantage, data scientists become increasingly valuable.

Data science is rarely an activity assigned to a single department. The profession and activity permeates organisations, supporting all manner of activities and departments. Slice through any organisation and you'll see layers of hybridisation. You'll have hybrid personas – data scientists with a range of skills, some with sharp business acumen, others with PhDs. Next, you'll find hybrid tasks – from data management to building machine learning models and deploying insights back into the business. Then of course there's a hybrid of data sources. All of which combines to a hybrid of data science activities and processes, which of course lends itself ultimately to a hybrid of software – proprietary and open source.

It's a complex environment with nothing to link the disparate threads of activities, skills, software and data. What connects the work being achieved in open source with that being undertaken in proprietary (otherwise known as Independent Software Vendor, or ISV) software? How is collaboration enabled? What about the continuity of projects and governance? Surely the beauty of this very environment is that data scientists could, in principle, dip in and out of ISV

analytics and open source as they see fit?

Let's take a closer look at the issues and opportunities, and I'll talk you through why an integrated strategy that links open source and proprietary analytics will help you drive up the value of this most important business activity.

Change is coming…

According to newly published research by Vanson Bourne, entitled 'Open Source versus Proprietary: What organisations need to know', "most organisations are looking to expand their use of open source while retaining a solid foundation of proprietary technologies". Largely, this is because proprietary is optimised for operational and production analytics and includes integrated capabilities for data management, audit, governance and lineage plus much more. Open source can quickly bring new analytic algorithms to market and test new models very quickly.

If hybrid is the future, what does it mean for the data science community and users in general?
Should you embrace this approach? How could it impact your daily operations? What will it mean for your career and the value of the work you do?

Your peers think…

According to Vanson Bourne research, using open source is predicted to bring new opportunities to both the organisation (95%) and to customers (89%).

While some 75% would consider using an open source platform for data analytics, 95% believe this poses several challenges for their organisation.

Insight 1: More creativity and freedom

I believe that one of the most liberating things about deploying a hybrid approach is the scientific freedom you will gain to experiment with new types of models, to test, customise and then, most importantly, to productionise them. A hybrid analytics strategy will ensure that you have the right tools for any job – a significant improvement in the working life of a data scientist.

If you have already developed analytics in open source, you will benefit from the governance and efficacy of powering them with proprietary software. Most importantly, in terms of 'making a difference to your business' you will be able to benefit from significant repeatability via automation.

The scale and repeatability of a proprietary solution will cut time-to-insight, potentially giving you and your analytical organisation an important

'edge'. This should also allow you to spend longer thinking about the kind of analytical insights that would support your organisation's business objectives and less time tweaking code for versions of the same use case.

You will gain or maintain the freedom to choose the language to work in, and flip between them for different use cases. These are really important points that could materially improve your daily working life.

Insight 2: More pressure to demonstrate value

If highly successful organisations such as Uber can disrupt entire markets using sophisticated algorithms, CEOs from every enterprise will be asking their data science teams why they can't do the same.

Perhaps this is a question of where analytics are being deployed? Until now, the most common uses have included operational management, strategy and planning, and financial planning and forecasting. In future, it seems that analytics will drive risk assessment, people management, demand forecasting, and customer experience. Does this highlight a shift towards using analytics in value-generating, front-end processes? I think so, and consequently, you will need the tools with which to generate ROI more rapidly. And you will also need to expand your repertoire, including techniques such as machine learning.

However, gaining value quickly can be problematical. The Harvard Business Review tells a concerning story about ROI. They say that amongst a group of

150 specialist data scientists focusing on machine learning, only 50 had developed a functioning model and none had managed to generate value from it.

It's headline grabbing figures like these that will be making CEOs and Line of Business managers twitchy. However, in a hybrid world, this time-to-value problem will improve. Why? You will be supported by the proven maths of proprietary systems that will have been finely honed over many years. The huge processing power will help to accelerate ROI – ideal for analytical activities such as machine learning. Perhaps most importantly, you'll be able to tap into vendor support. And if you're a staunch 'open-sourcer' you may rely on 'the community' to come to your aid. However, as Tony Wasserman, professor of software management at Carnegie Mellon University notes, there will be times, working on mission critical projects, that it will be incredibly useful to know that immediate expert support is only a phone call away.

In Wasserman's words:

"People often respond very quickly to queries posted on the forum pages of widely-used open source projects, but that's not the same thing as a guaranteed vendor response in response to a toll-free telephone call."

Furthermore, if you can influence the choice of a world-class vendor for your organisation, you will be able to tap into expert onsite support during the early phases of integration and model building to help you drive faster ROI.

I believe this pressure to demonstrate 'business value delivered' will only intensify as analytics becomes an increasingly important tool with which to generate competitive advantage. So why not lean on the support a hybrid approach can provide?

This increasing reliance brings me to my next point – data quality and management in the hybrid landscape.

Insight 3: Data management must be addressed

Imagine that your organisation has a big data strategy. Perhaps it employs machine learning or artificial intelligence, embedding advanced analytics into key business processes. Your organisation will be taking in potentially net new data sources for example from IoT sensors, or voice logs from contact centres, perhaps even social media feeds. And let's not forget that the velocity of data analysis is also on the increase. In these circumstances the way you manage data is almost as important as the analytics you use.

It's easy to see why data management is such a fraught issue – and one that puts the brakes on time to value. However, as your organisation becomes an increasingly sophisticated user of analytics you will need to be able to integrate the processes of data management, discovery and insight deployment. The question to ask is how you will stitch all these activities together into a coherent process.

Unsurprisingly, data quality is also being attributed to organisations' inability to generate business value from data science. In fact, Harvard Business Review is quoted as saying that data scientists themselves believe lack of value is because "data is a mess".

Without good quality, well organised data, the credibility of your insights will be undermined. However, in a hybrid environment, a proprietary system will make provision for best practice data management. It will help to convert different types of data into a format that makes data science activities easy. You and your colleagues will have a common understanding of the data you're working and you'll be able to more quickly understand what data is relevant to the work you're undertaking.

Insight 4: Integration issues – as tricky as you think?

On first appearance, it might seem that a hybrid approach adds another level of complexity to your data science and analytics activities. Actually, what you acquire is the ability to work in the language of your choice, while powering your models with the commercially proven algorithms of your proprietary system. The integration, when it needs to happen, occurs seamlessly through open REST APIs.

For those organisations adding open source software to the mix, there will be a need to hire in more people with integration skills. Could this be an opportunity for you to add skills to your personal mix, and drive your personal value up, and make a hybrid approach really work for you?

Insight 5: Your career could get a boost

As analytics techniques develop and the capability spreads throughout organisations, some data scientists will be required to take less of a siloed approach. This trend is termed the rise of the 'hybrid data scientist'. In recent US research by Analytics Week and Business over Broadway, nearly 50% self-identified as a hybrid data scientist of one form or another. Of them, 32% believed they held two functions, 13% held three and 4% held four roles.

Clearly, the more skills, the greater the salary. Even in 2014, TechRepublic was reporting that merely including the term data scientist in your title, rather than data analyst, could add around 40% to a pay packet. One can only imagine how actually skilling up is beginning to impact remuneration packages!

I believe it is the adoption of a hybrid approach with open-compatible technology that will provide the lever to accelerate the trend towards the hybrid data scientist – and its potential to transform your personal value. Why? Because of its flexibility, scalability and agility, allowing all lines of business to readily and rapidly access transformative insights. All while having the mechanisms of governance and best practice data management in place to secure profound peace of mind about the credibility of analytic-based answers to complex questions.

Therefore, whether you are approaching hybrid from the perspective of taking proprietary software and 'making it better' with open source, or the other way around, integration skills certainly add value to your business and also to your career.

How can you make the shift to hybrid as effortless as possible?

It probably won't surprise you that my opinion is this: partner the experts.

However, to make hybrid as easy to use, quick to deliver ROI and as beneficial to you personally, I would recommend partnering experts that can provide a holistic approach. Those that are responsible for developing software with the power to create leading algorithms; those that understand the evolving working practices you apply; those who are at the leading edge of innovation; and those who cover all the essential bases, such as data management, compliance and security as part of their open platform.

At SAS, we are supporting the evolving needs of data scientists in all sorts of commercial and public sector organisations with our new, open analytics platform, called SAS® Viya™.

Our cloud-based platform helps you drive business value by encouraging innovation through collaboration. You and your colleagues can work together to solve complex challenges using a common and consistent interpretation of data, as well as multiple types of data from structured to unstructured, historical to real-time. Importantly for those used to open source, SAS® Viya™ allows you to create relevant, portable analytic assets using consistent processes for model assessment, benchmarking and monitoring. This means your business can trust the insights you deliver, getting them back into processes rapidly. Why faster? Because you'll have computational processing that's optimised for analytic workloads – however complex and heavy. Finally, SAS® Viya™ gives you the all-important repeatability and automation that open source alone cannot. Compliance is effortless because all assets and actions are fully traceable.

In conclusion

A large percent of those respondents to the Vanson Bourne research believe that a combination of open source and proprietary software improves the ROI of analytics.

Embracing a hybrid approach, when backed by a holistic, single open platform will give you the very best of both open source and proprietary worlds. Most importantly, as data and analytics become the new battleground of competitive advantage, I believe your future success really requires the scale, flexibility, speed and performance hybrid delivers. You will be free of mundane data manipulation tasks and free to spending more time thinking and exploring. And isn't that the reason we all love working with data in the information economy?

"We observe an increasing number of our customers are embracing the idea of a third way in analytics. A hybrid approach that gives organisations greater analytic flexibility and speed, while underpinned by a market-leading engine, sophisticated data management and GDPR-compliant governance capabilities. They are empowered by its cloud model, meaning that working together to find answers to today's complex business decisions is made easier with SAS® Viya™." **David Shannon, Technical Director, Amadeus Software**

Get the facts about the hybrid migration

Download our research Open Source versus Proprietary: What organisations need to know