



Out in the open with analytics

Combining the benefits of SAS®
and open source analytics



Out in the open with analytics

How can you open your analytics program to all types of programming languages and all levels of users? And how can you ensure consistency across your models and your resulting actions no matter where they initiate in the company?

With today's analytics technologies, the conversation about open analytics and commercial analytics is no longer an either/or discussion. You can now combine the benefits of SAS and open source analytics technology systems within your organization.

As we think about the entire analytics life cycle, it's important to consider data preparation, deployment, performance, scalability and governance, in addition to algorithms. Within that cycle, there's a role for open source and commercial analytics.

For example, machine learning algorithms can be developed in SAS or Python, then deployed in real-time data streams within SAS Event Stream Processing, while also integrating with open systems through Java and C APIs, RESTful web services, Apache Kafka, HDFS and more.

In this e-book you'll learn how organizations in many industries are using open source analytics and SAS, getting the most from both, and what role SAS plays throughout the analytics life cycle.



3

Keeping an open mind about open analytics



6

Industrialized modeling helps free-to-play video games earn big profits



9

5 questions about open analytics



12

Open source momentum in the new world of analytics



15

Analytics for all, all in one place



Keeping an open
mind about open
analytics

What if we stopped arguing over which analytics software is best, and decided instead to use them all?

Today's data scientists come from many different backgrounds, and they bring a wide range of skills to the job. If they have access to a variety of analytics tools, along with a system to govern and deploy models consistently, they have more options for solving complex problems.

Cleveland Clinic and Cox Automotive are two organizations that have benefited from this realization. As a result, their data science programs are thriving – and so are their larger organizations.

"We have employees who are trained in multiple languages and technologies. We want to enable people to access and use languages they're comfortable with but using a common approach," says Chris Donovan, Executive Director of Analytics for Cleveland Clinic.

Cleveland Clinic hopes to grow analytics maturity across the whole health care system, says Donovan. Instead of centralizing analytics skills and capabilities in one team, they're building a broad program across the enterprise. "Having a platform that enables that is critical for analytics to be successful."

This focus on analytics has helped Cleveland Clinic transform, along with the industry, from a focus on palliative care to preventive care. Donovan explains: "How can we move away from just taking care of you when you show up as an individual patient in the ER or the doctor's office, to looking at a population of patients and thinking about how to prevent people from getting sick in the first place?"

Becoming 'code agnostic'

At Cox Automotive, SAS® Viya® is the glue holding the analytics organization together, says Shawn Hushman, former Vice President of Decision Sciences for Cox Automotive. "It removes the different political debates

between the source systems, so we can focus on the modeling itself, the versioning of the models and the delivery of the models. The open platform allows everyone to use their code, leverage open source opportunities, and it opens everyone up to new code bases."

In particular, Hushman praises the ability to integrate disparate code, processes, and information into one hub that provides consistent delivery of information.

Hushman's team includes people around the world who code in Python, R and SAS. "We have people programming in multiple interfaces, and they're using different ways to collaborate on model development. They have different ways they like to publish and show the output and different ways to deploy the models. SAS allows us to stitch it all together."

The solution allows Cox Automotive to be "code agnostic," says Hushman. Instead of debating over a preferred code base, everyone can discuss the best solution for a problem together. "We don't care about your code



Shawn Hushman
Former VP of Decision Sciences
Cox Automotive

"SAS has the ability to bring together the data scientist community like no other solution can."

preferences,” he says. “Let’s look at frameworks that can bring real change to the organization, instead of battling over which package I’m going to use for modeling.”

With that mindset, Cox Automotive has been able to make the most of its data as it transitions offline businesses like Autotrader and Kelley Blue Book to the online world. “SAS has the ability to bring together the data scientist community like no other solution can,” says Hushman. “We think of model management as the center of our hub, because that’s where we can be agnostic and make sure everything connects.

“Our responsibility is to deliver results efficiently and make it seamless,” he says. “We want all the goodness that comes with the diversity of different algorithms, and then we bring alignment around our delivery.”

Opening analytics to executives

At Cleveland Clinic, giving more people access to analytics is also a priority. Beyond data scientists and programmers who are adept at writing code and doing advanced analytics, Donovan and other leaders at Cleveland Clinic want to make data easily available to executives and managers with drag-and-drop capabilities and simple interfaces.

“Our leaders may not know how to build a predictive model, but they need to be able to use data to make better decisions,” says Donovan. “Not everyone is a data scientist. But everyone needs to be able to interact with data at their level.”

Donovan says Cleveland Clinic is redefining what an analyst is and working to create a common entry point for all levels of users. “Before, we had data everywhere and multiple tools, but we’re trying to invert that. Instead of taking the data to the people, it’s bringing the people to the data. We believe that if we create a world-class platform, that will draw the people to the platform – which will drive consistency, build communities of practice, and link people across the organization to find standard approaches.”

Opening your business strategy with analytics

Hushman also emphasizes the importance of analytics across the organization at Cox Automotive. “I prefer to view analytics as the heart and soul of our organization and the foundational element of everything we do. Analytics is improving all our products, driving new products, and growing revenue across all our product suites. Analytics doesn’t drive our business; it is our business.”

Broadening the use of analytics to support all users is more than a technology tactic. It’s a business strategy. “Analytics is not just a capability that supports your core strategy,” says Donovan. “It has to be a core strategy of its own. You need to become an analytically mature organization, and you need to be world-class in that space or people will leapfrog you.”



Chris Donovan
*Executive Director
of Analytics*
Cleveland Clinic

“Analytics is not just a capability that supports your core strategy. It has to be a core strategy of its own.”



Industrialized
modeling helps
free-to-play
video games
earn big profits

Wargaming scales its analytics to
understand terabytes of daily
customer data

Every day, millions of people play World of Tanks to clash with other tanks in virtual battlefields – and many of them have never paid a penny to play the game. Yet Wargaming, World of Tanks' developer, has made billions in the gaming industry.

How did this gaming juggernaut monetize its free-to-play gaming structure and come to see the priceless value of every customer – even the ones who don't pay a dime? In part, through its use of analytics.

The goal of Wargaming is to offer a great experience for players of all levels. Anyone can play for free on Mac, Windows, console or mobile versions, and players who want more can make in-game purchases.

Wargaming's most played game, World of Tanks, has 110 million registered users online, and Wargaming collects data on every shot fired and each move made in every online game. "At any given point daily, we have about 4 million players playing our game," says Alexander Ryabov, head of Wargaming Business Intelligence Data Services. "They play multiple battles, and those battles have multiple events in them that all generate close to several terabytes of data daily."

Wargaming captures data from the second players log in to a game to when they log out. The company also collects and analyzes in-game chat logs, along with mentions of its games on social media sites and in many gaming discussion communities. Using this data, they can run models to retain customers, cross-sell other games, convert players into paid users, monitor the player journey and reduce friction points in the games.

In all, Wargaming processes more than 30 terabytes of data per month. It stores 98 percent of its data in Hadoop on an Oracle Big Data Appliance, with Cloudera

managing the Hadoop implementation. Once the data is in Hadoop, ETL developers create data marts that integrate with SAS® to generate models and put them into production.

Improving gameplay and customer offers with analytics

A team of data scientists at Wargaming develops models whose scores can be sent to an event-processing component in the game, to the company's CRM systems and back to the team for additional modeling.

Recently, for example, the team recognized in the data that players kept dying in one particular place. "So they put up a hill in that place to balance the map," explains Ryabov. "Our data scientists have created a heat map where you can see, on a game map, every shot fired during a certain period of time."

The team also uses analytics to see if players are missing out on certain elements of the game so the game can send notifications for a better experience the next time. The message might tell a player where to access certain weapons or identify overlooked locations from a previous game.

"It will help players have a better experience in a game the next time," says Ryabov. "This is just one example, but a lot of things like that can be accomplished using modeling and putting those models into production."

To improve customer experiences even further, Wargaming applies text analytics to feedback collected on social media and in direct conversations with customers. "We can put certain filters in social media to get a sentiment analysis of overall play. We can also use sentiment analysis for customer support and to identify our all-star players on multiple channels," says Ryabov.

Scaling analytics to the massively multiplayer online experience

When Wargaming created its business intelligence program three years ago, it gravitated toward open source technologies. "Once we understood the need for in-depth data analysis and data mining, we started doing some initial, advanced analytics modeling in R, Spark, Python and all the other open source solutions," says Ryabov.

The benefits of industrialized modeling

Automated and industrialized modeling has created many benefits for Wargaming:

- Transitioned most coding to a point-and-click based workflow for model building efficiencies.
- Reduced the amount of time needed to develop and deploy models by 60 percent.
- Reduced the need for data warehouse administration in the deployment and automation of models by 80 percent.

But the team realized that scaling those initial efforts to thousands of models and more and more data every day posed big challenges. Describing Wargaming's early use of open source analytics, Ryabov says, "The biggest issue for us was scalability. Our data scientists come up with a model concept, do some data wrangling, some data extraction and then we need to automate the results. It was all manual. It was a lot of work for our developers."

According to Ryabov, the first models his team created took three to six months to implement. "Once we realized that we're going to be running hundreds or even thousands of models for all of our games, all of our regions and in all of our time frames, we started looking for the solution that could make it scalable for us."

After some in-depth research, Ryabov and his team found what they needed. "SAS Factory Miner and SAS Model Manager were perfect for our use cases," he says, "because we can

take the same model and multiply it by time frames, regions and by different products. So a model is virtually the same, but we can put it into the production environment, where we run, maintain and promote it over and over in an industrial sort of way. In our research, SAS was the only viable option."

After the data is prepared and the modeling methodology is established, Ryabov says multiplying the model to thousands of similar models has become a one-person job. "To manually create and maintain that many models would take something like 10 to 20 people, and naturally they'll make mistakes. An automated production environment like SAS does not make mistakes."

Overall, Wargaming data scientists are able to create and deploy more models in less time, which will result in higher revenues, better use of resources and lower opportunity costs. As the market grows and Wargaming continues to

diversify into other platforms, it will be able to run even more models, retain more customers, acquire more customers and apply more complex analytics, all within the same analytics platform.

Most importantly, the players benefit too. "Our data scientists are a group of talented people that have very innovative ideas on how to offer players exactly what they want at the right time," says Ryabov. "And SAS helps increase overall satisfaction and make the player experience even better."

Plus, improving the gaming experience encourages more players to become long-term customers with a desire to invest in the game. "As our founder, Victor Kislyi, says, 'Our goal ultimately is happy players.' If players are happy, you know everything else will come."



Alex Ryabov
Head of Business Intelligence Data Services
Wargaming

"And SAS helps increase overall satisfaction and make the player experience even better."



5 questions about open analytics

Wyndham Destination Network
embraces openness in its
analytics programs

When you need to price vacation rentals for 3.8 million vacation members who have access to 2 million distinct SKUs, you need accurate, reliable, nonstop analytics.

At Wyndham Destination Network, the analytics team uses a combination of open source and enterprise analytics to find innovative solutions to complex problems from pricing elasticity to resort operations and customer retention.

Recently, we talked to Jeremy TerBush, Senior Vice President of Analytics at Wyndham Destination Network, about his team's use of analytics, ranging from open source technologies to enterprise solutions.

How do you define open source vs. open software?

Jeremy TerBush: Open source to me means I can download the system right now and create a program directly on my machine without any licensing. It's free, easy to use, and there are a lot of communities out there talking about it and working on improving the software.

Open software, on the other hand, is software that makes it easy to connect with other pieces of software. We have systems that produce forecasts, and they need to send those forecasts over to another routine. Open software would allow that to happen seamlessly and easily from a development and execution perspective.

How have you embraced openness at Wyndham?

TerBush: We don't mandate that everyone use the same tools. When people are coming across a new problem, and it's not part of an implemented system we've already built, we've seen our analysts turn to more of the open tools. We've had a lot of success developing and solving smaller scale projects with open source software. Now, if we need to operationalize that work, typically we find a way to move it into our larger environment and use SAS.

What we have is somewhat of an open ecosystem because we allow the group to use whatever tools it wants given the problem at hand. Open source is used in the prototyping stage for one-off problems. That's where we've found a lot of success using open source technology.

When would you switch from open source technology to enterprise software?

TerBush: The processes we're setting up that need to be run daily are typically pricing processes, like forecasting demand and price elasticity for our rental units. To do that, we have to take all the recent historical and transaction behavior and feed it into a forecasting routine to ultimately recommend a daily price for that unit. That model includes optimization and forecasting that were prototyped with open source analytics software. When we were happy with the results, we moved the algorithm to SAS for daily scheduling.



Jeremy TerBush
Senior Vice President of Analytics
Wyndham Destination Network

"So we'll do the forecasting with open analytics as an initial prototype. When we're happy with that first stage, we run it in SAS. For us, it's easier to create an industrial-strength product in SAS."

If it's a job that has to run on a daily basis for pricing forecasts, I don't have the confidence that any open implementation would give the team what it needs on a daily basis to optimize prices. So we'll do the forecasting with open analytics as an initial prototype. When we're happy with that first stage, we run it in SAS. For us, it's easier to create an industrial-strength product in SAS.

What are the benefits of using both?

TerBush: One of the main benefits of using both SAS and open source software is that you can let people use the platform they're most comfortable with. It's a learning curve for them to completely shift their thinking. To allow everyone to use whatever platform they're comfortable with is a big productive boost for us. I think we get better answers too. We can collect the best practices and draw from communities of all platforms and have a bigger pool of ideas. It gives us more flexibility within the team to solve the problem in a number of different ways.

What other technologies are important for remaining open?

TerBush: Within our organization and across other organizations, it's becoming much easier for companies to share data by developing common-sense APIs. For example, we did some work recently where we needed to use the Google Maps API, and we went to Python immediately.

We're seeing more and more that there's external data that we can tap into. When building solutions internally, we want to communicate our models to multiple platforms and send results to APIs to whatever platform needs to consume it. That's definitely the way things are going: being able to produce results so any platform can consume them, and APIs are the way to do it.





Open source momentum in the new world of analytics

Excerpted from a Quantum
Insights research report

With business units playing a stronger role in corporate strategic decisions, and with core development projects moving so much faster than before, the move toward open source for analytics starts to make sense. Consider these words from an innovation CTO at a large global manufacturer:

“Why open source? Cost counts, of course, especially in data analytics, where people might be unsure of the business case, so they feel they ought to try out ideas before committing a lot of time and money. If they are not absolutely clear as to whether investment is justified, the entry barrier with open source is much, much lower. This means lower risk combined with greater speed.”

Likewise, we have seen a senior IT management team tasked with delivering a robust platform for integrated service management and delivery, for internal and external use, with an 18-month combined development, production and delivery schedule.

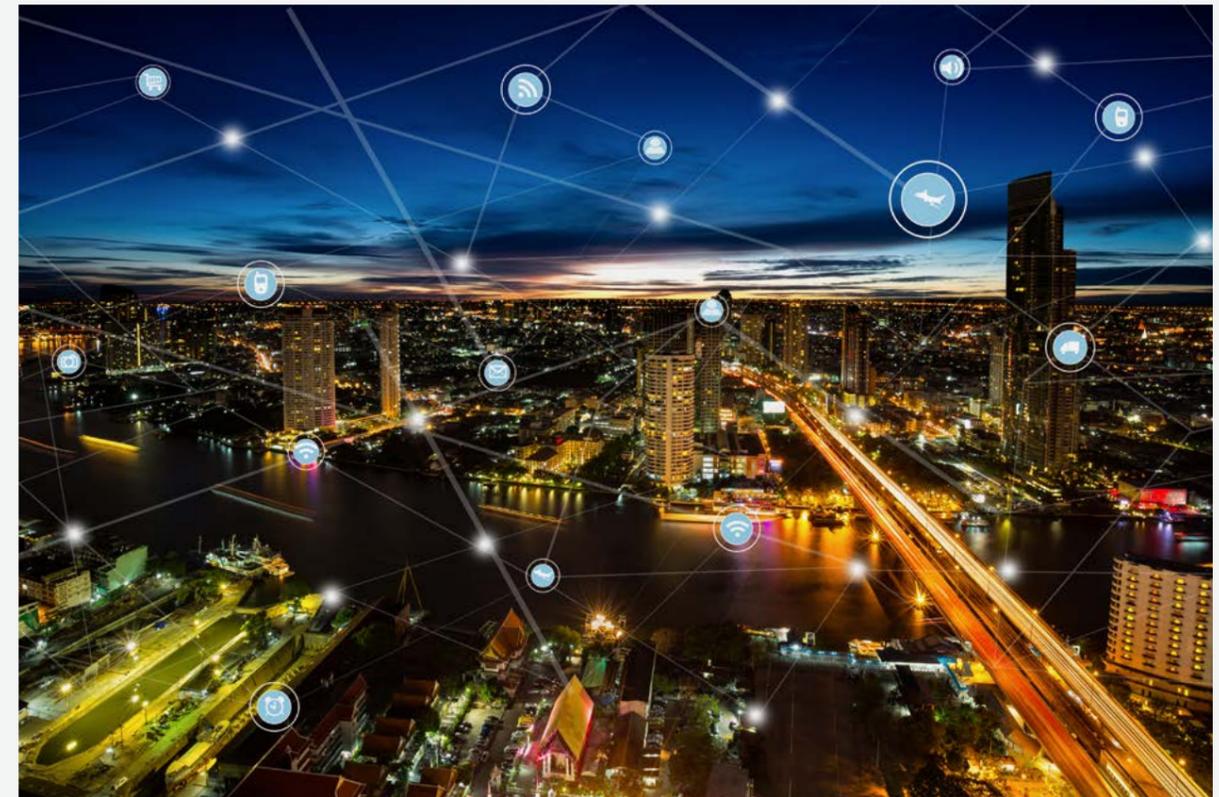
This directly drove a large-scale move toward open source technologies. The need to deliver very fast results means that it becomes easier to fail fast at low cost and move on, rather than prepare meticulously and hope not to fail at all.

A new risk equation is developing, together with new working methods. Devolved working is now the norm, with individual units and teams delivering specific projects to suit their own goals, and corporate IT accountable for ensuring that they all fit into an ecosystem.

This is causing a “cultural revolution” in many companies, involving much higher levels of collaborative working, not just across internal boundaries but often also with small specialist partner companies or even communities of external developers.

It is also notable that the move to open source for analytics software has led to a consequent reduction in the use of traditional vendor software, as explained by a CTO of analytics and innovation at a large engineering company:

“One of the worries with traditional analytics players is that you lose a lot of time in negotiating with salespeople. There is a lot of money involved, and that slows you down. It is very hard to make a call and then change it later. With open source, even choosing the wrong technology entirely is not necessarily a disaster. The more agile your own processes are, the more logical it is to choose open source suppliers.”



What do we conclude from this?

The move to greater use of open source makes the enterprise IT environment more complex, with the need for very strong processes to handle a software stack that sometimes threatens to become uncontrollably large and diverse.

CIOs accept that managing large IT environments is “what they do” and they simply have to make it happen. Yet much of the apparent savings made through the lower costs of open source can be lost due to extra consulting costs and higher internal IT spending on building a robust, integrated and secure stack capable of handling this growing software diversity.

The cost calculations, therefore, may seem simple to the business users who commission projects using open source software, but it looks very different from where the CIO sits!

Recommendation for a new world of analytics

Based on our findings, we think the logical answer is to put in place an industry standard, software-agnostic analytics platform that is constantly updated and modular, and always represents industry best practice.

This represents low maintenance risk, low investment cost and high operational flexibility in a standard, largely off-the-shelf package.

It enables business users within large enterprises to use the open source analytics software of their choice without risking the integrity of the wider environment.

What would such a specialist analytics management system look like?

- Not simply a general-purpose integration layer of the kind common in all enterprise IT landscapes today.

Every aspect of business management today depends on analytics, so the decisions made in the near future about sourcing, procurement, implementation, management and continuous evolution of analytics tools will be critical.

- Instead, this should be a specialized analytics workload operating environment.
- Designed and optimized for management of analytics assets and development tools.
- This makes it possible to apply secure, professional service management to a wide range of open source products, including heritage applications.
- This will include development of managed open source stacks, backed by rational, robust methods for evaluation, a single strategic road map and governance methods for adoption, development and usage.

Every aspect of business management today depends on analytics, so the decisions made in the near future about sourcing, procurement, implementation, management and continuous evolution of analytics tools will be critical.

The task for enterprise IT management is to deliver the speed and local autonomy that business units require, while still making strategic decisions and maximizing the potential of end-to-end data flows.

This article is excerpted from *Evolution or Disruption: The New World of Analytics? An Informed View From Inside Enterprise Business*, a Quantum Insights independent research report commissioned by SAS.



Analytics for all, all in one place

Govern best-of-breed analytics
with an analytics platform

Let me introduce you to some smart, analytically minded employees at a growing national bank who are each seeing positive results with analytics in their own divisions:



James

James is a data scientist in the business development group who analyzes data to create customized offers for the bank's premium customers.



Susan

Susan is a digital marketer who tracks and influences the customer journey for prospective mortgage customers.



Christina

Christina is a risk analyst who builds risk models for the bank's loan portfolios.



Marvin

Marvin is a citizen data scientist in a top retail banking location who takes it upon himself to analyze data about local customers at his branch.

Between them, these enterprising employees are using a dozen different packages for analytics and data management. From Informatica and SAS to Python and R, some of these applications are open source, some are cloud-based and some are hosted as enterprise applications.

How can the IT department at this bank make sure all of these projects are using trustworthy data, accurate models and a rigorous process of analysis that will guarantee useful results? And whose responsibility is it to stitch all of these disparate code bases and business scenarios together to trace the customer journey or find other opportunities for analytics?

A single ecosystem for analytics

What this bank needs is one place to combine data for analysis, consolidate all of its analytics endeavors for tracking and monitoring, and provide appropriate access to models across divisions. But how can one platform corral all that analytics activity, provide easy access for all levels of users, and provide model governance across all types of analytics projects?

Sarah Gates, Analytics Platform Marketing Manager at SAS, recommends a centralized analytics platform that is inclusive, scalable and resilient. "A unified platform that works with your IT requirements to govern the entire analytics life cycle gives you the ability to quickly adapt to the full spectrum of analytic opportunities in your organization, foster innovation and streamline deployment of the results."

The platform should be accessible to different types of skill sets, including providing data management capabilities, analytics functionality and model governance for data scientists.

"It needs to govern all analytics, not just some," says Gates, especially open source packages that don't come with their own governance functionality.

"You need an analytics platform that unifies and governs your complex and diverse analytic ecosystem," says Gates. "That is key for driving business value from your data while mitigating risk."

How an analytics governance platform works

So what is an analytics governance platform? It's software that allows you to connect all the parts of your analytics ecosystem. It helps the business quickly get value from analytics. It helps IT from a governance and compliance perspective by providing information about model versioning, audit trails, model lineage and source data.

And how does an analytics platform support model governance? It provides data scientists with a single platform to leverage throughout the entire analytics life cycle. The platform provides them with a centralized repository of data, an inventory of analytic methods and a standardized code base that can be incorporated into any deployment activity or action. They are able to build,

"An analytics platform positions you to deliver business value from analytics, from concept to innovation and implementation to ROI."

Sarah Gates

Analytics Platform Marketing Manager

SAS

compare, and select models regardless of the language in which they were created. Model results are easy to interpret, ensuring they are meeting the business need. Code is portable, so it can be defined once, run anywhere and scaled to solve any size problem. Finally, they can register models for deployment. Business analysts apply business rules to describe the conditions in which a model should trigger. Everything is documented and tested for deployment purposes.

As a result, IT understands the business context associated with every model developed. Models don't have to be recoded into another language before deployment. In all these ways the platform ensures effective model governance, as well as accurate results through monitoring of model inventory, performance and deployment.

"It's not unusual for data scientists to recode models in a different language and wonder why the results don't match," says Oliver Schabenberger, Executive Vice President and Chief Technology Officer at SAS. "With analytics governance in place, you can request analytics on the same data from different languages, and you don't have to worry about idiosyncrasies and differences of algorithm implementation, options and data models. Consistency and standardization on the back end remove worries about differences on the front end."

An open, inclusive analytics platform allows all data users to stand behind the best algorithm. It doesn't matter what language you write it in, or where you drive it from - you know it will be accurate.

"With the platform in place, it's no longer an argument between an algorithm in an R package or a Python library," says Schabenberger. "The languages execute the same analytic code on the same data model, no matter how you trigger it. And all languages reap the benefits of multithreading, distributed computing, common data access and security models."

Schabenberger compares the model governance capabilities of an analytics platform to web governance, explaining that you can pull up a webpage from multiple different laptops, tablets or smartphones and still see the same content. Web governance standards allow that to happen.

Analytics model governance works the same way. You can pull in data and build models from different systems throughout the organization, and the analytics governance platform ensures consistency, accuracy and performance.

Open and enterprise analytics together in one platform

Let's get back to the bank from the beginning of this article. An analytics governance platform lets this organization be open enough that James, Susan, Christina and Marvin can all use their preferred tools, and the bank can guarantee consistent enterprise-class results across all the different packages. IT now governs analytics implementation and the data to make sure everyone is working against the same data model.

When so many different users are accessing data in different ways, an analytics platform is the foundation for becoming a data driven organization. "An analytics platform positions you to deliver business value from analytics, from concept to innovation and implementation to ROI," says Gates.

Open your mind to SAS and open source working together with [SAS® Viya®](#).

Follow us:



To contact your local SAS office, please visit: sas.com/offices

