

Get more value from your analytics fast

Optimized AI and analytics in the cloud



In the cloud computing world, performance is literally money: cost control, revenue and time to value

A benchmarking study conducted by The Futurum Group shows that SAS® Viya® one of the most cost-performant and cost-conscious analytics platform available to enterprise cloud users. According to the study, the SAS AI and analytics platform is on average 30 times faster, 86% more cost-effective and has superior scalability compared to leading commercial and open source alternatives.

The conventional computing wisdom is that performance equals time. The ideal performance state is using your resources to their peak without affecting runtime.

On the right, you'll see an overly simplified view of the results. In Figure 1, runtime is cut in half when cores are doubled. In Figure 2, CPU is used to maximum capacity while the workload is executing. In a cloud environment, an analytics platform must operate within these extremes to be as cost-performant as possible.

Figure 1: Runtime vs. available cores

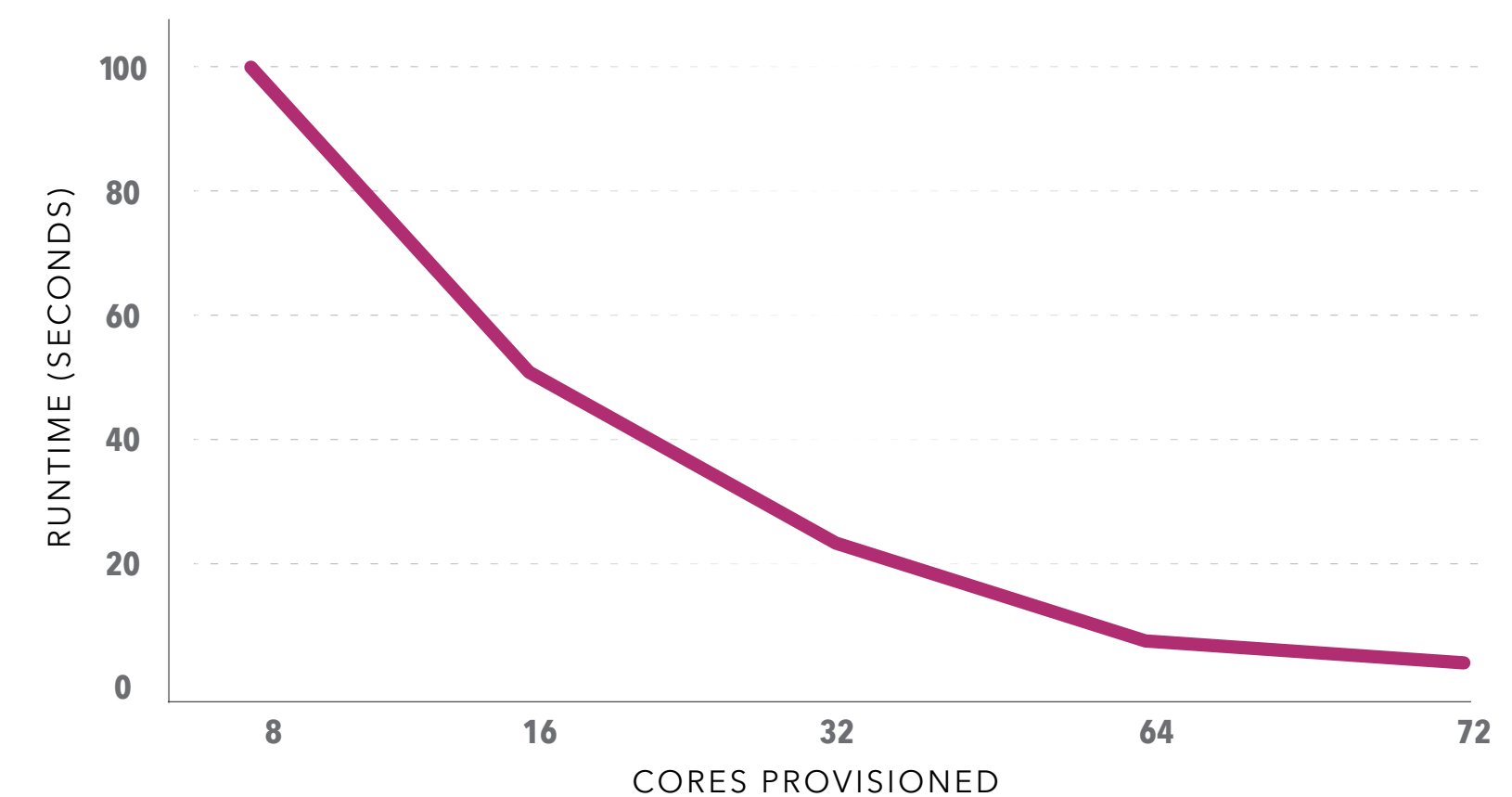
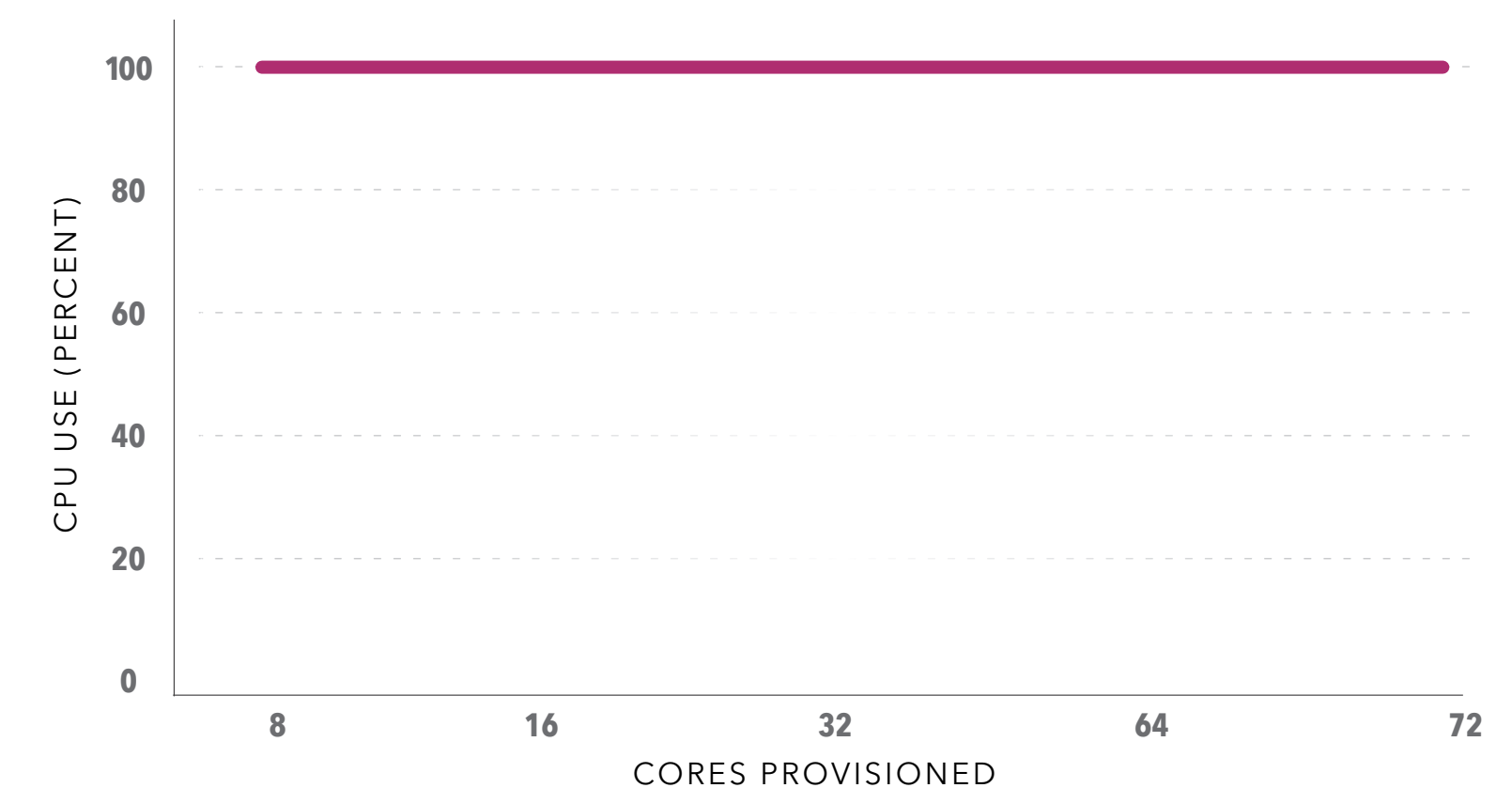


Figure 2: CPU use vs. available cores



The top reasons companies back away from the cloud

Eighty-two percent of enterprises surveyed by [Flexera](#) in late 2022 cited managing cloud spending as their top cloud challenge, the first time in over a decade of conducting the survey that cloud spending has ranked above cybersecurity.

The leading drivers of this retreat are:

82%

Managing cloud spending

79%

Security

78%

Lack of resources/expertise

Optimizing your cloud needs is step one

Organizations without a cloud optimization process tend to overspend by about 40% due to unmanaged costs, unexpected usage, suboptimal design and implementation, and wrong-sized production and waste in development and testing environments, among other reasons. SAS can help you by accurately assessing your analytics transformation before implementing your cloud migration strategy or supplementing an existing strategy.

By using ecosystem diagnostics, you can:

- See which workloads consume the most resources and have the longest runtimes.
- Generate cloud environment and workload sizing requirements.
- Determine business justifications for cloud and data strategies that consider interdependencies between data, locations, routines and users.

Having detailed knowledge about consumption metrics provides two powerful insights. The first is a reasonable forecast of cloud cost and performance expectations and where they can be improved. The second is creating a validated map of which workloads and data benefit from a cloud environment and which shouldn't be migrated.

Cloud spending is growing each year and becoming a more substantial portion of a company's costs. As a result, infrastructure expenses are increasingly important to senior executives. Estimates are that 80% of organizations will overspend their budgets for infrastructure as a direct result of poor cloud optimization and overspending on cloud commitments that are too small to support their analytical ecosystem's consumption.

Cloud spending is growing each year and becoming a more substantial portion of a company's costs. As a result, infrastructure expenses are increasingly important to senior executives.

Finding cost-performant harmony despite ever-changing requirements

Computations change as constraints are introduced.

The data that feeds the algorithms is constrained by the hardware and the distribution of the algorithm. Changes to the data and how that data is distributed to the computing chip can alter performance, cost and accuracy.

SAS Viya helps you move toward the perfect harmony of math, data and chip hardware in several ways.

SAS has proven to be a highly performant language. Our algorithms optimize CPU usage, runtime, memory usage and input/output speed, often eliminating the need to purchase additional infrastructure to compensate for

inefficient CPU usage or high memory requirements, which are frequent problems when running large jobs, translating to unnecessary additional expense. In this way, the cost of computing is married to performance for the best results.

Viya is inclusive; it embraces and improves open source algorithms. Many data scientists who write in Python, R or other languages can readily take advantage of the stability, performance, affordability and model governance. You can also use its ability to run any new algorithms as part of existing model tournaments.

SAS actively identifies leading open source algorithms and

incorporates them into Viya. If an algorithm is better at parallelizing quickly, maximizing core usage or offering new advances, SAS is ready to adopt and refine its approach for maximum results.

Users can reliably pursue optimal cost-performance by calling SAS Cloud Analytic Services (CAS) actions from Python. SAS assists you in choosing from all available options to help ensure the performance and repeatability of the analytics pipeline.

SAS® Viya® is inclusive; it embraces and improves open source algorithms. Many data scientists who write in Python, R or other languages can readily take advantage of the stability, performance, affordability and model governance.

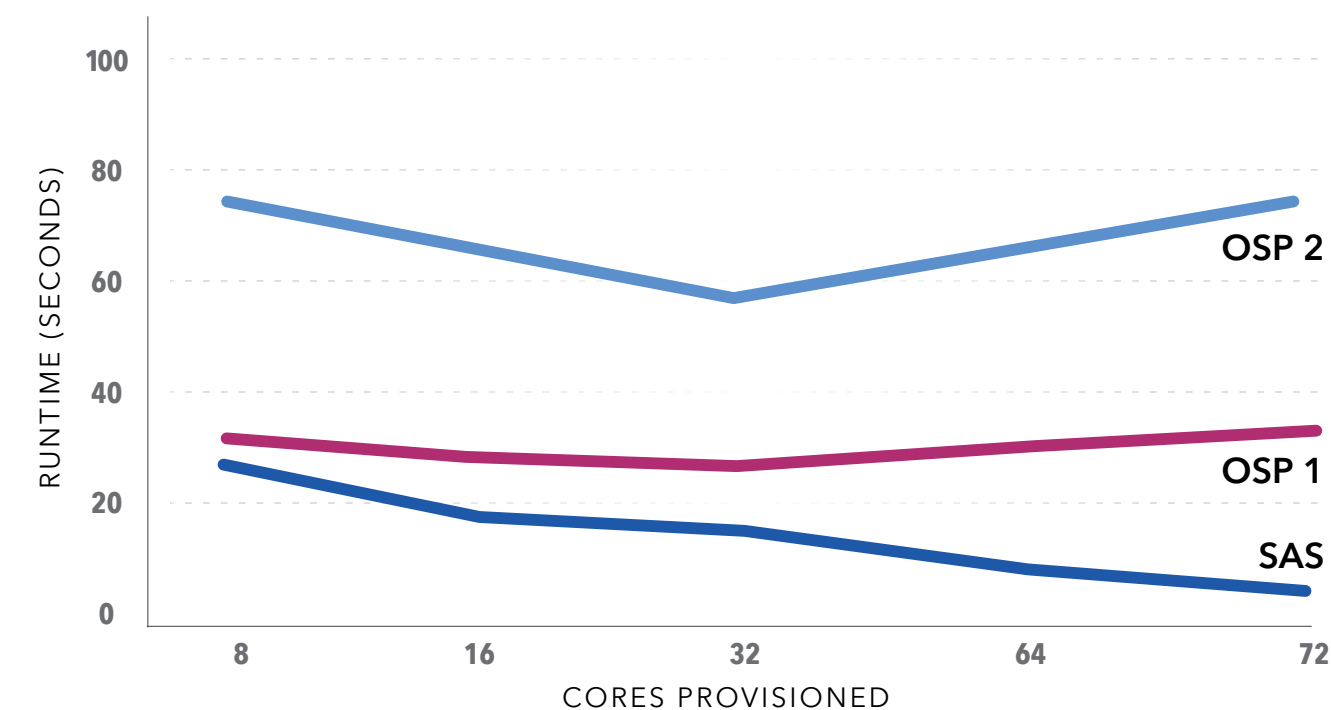
What does performance harmony look like?

Consider this comparison of workload executions for logistic regression (a common analytics algorithm) between Viya and open source packages (OSPs) when monitoring technology was deployed to the sessions to observe their performance.

View workload execution details

Viya provides the expected outcome: Adding cores showed that SAS runtime was distinctly faster than the OSPs across the worker nodes. OSP 1, which is closest to SAS in eight cores, didn't keep up with SAS runtimes as cores were added. This leads to higher costs with no improvement in runtime (or sometimes longer runtime).

Figure 3: Runtimes vs. available cores



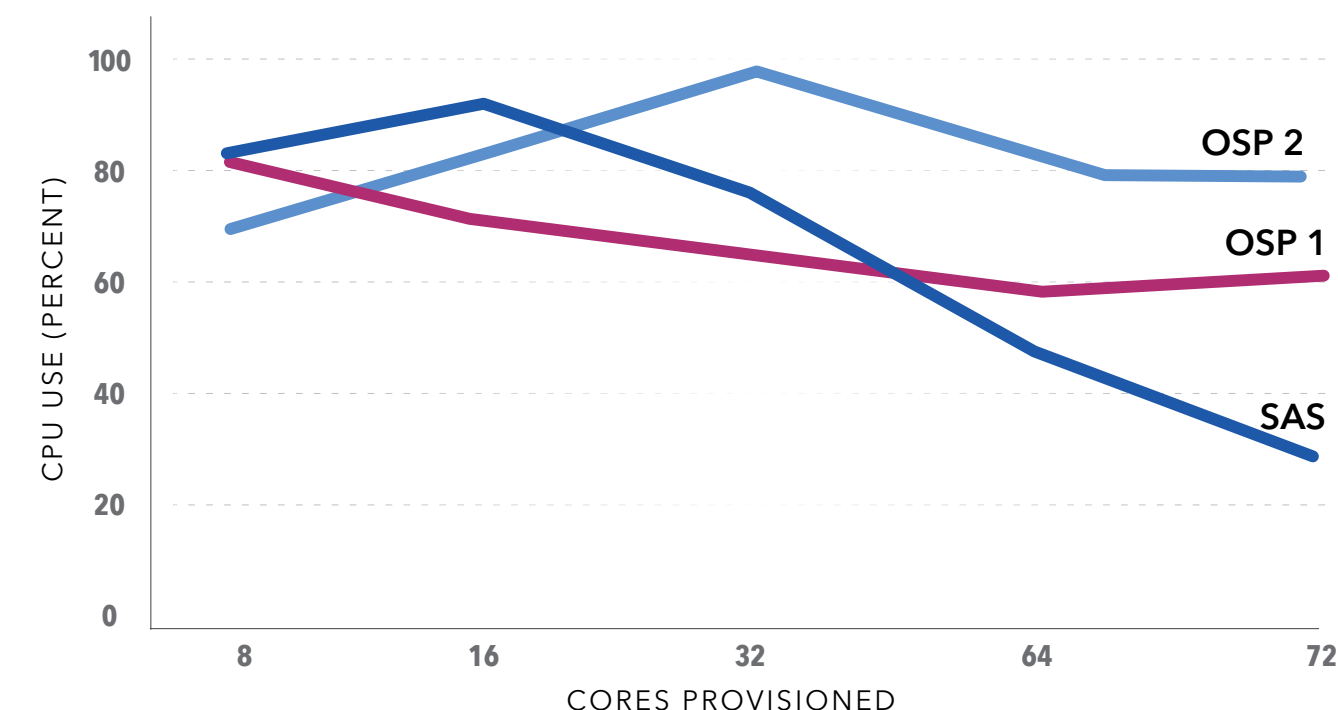
This result is confirmed by The Futurum Group's benchmarking findings, which showed:

- Viya is an average of 30 times faster than all alternative AI/ML solutions across all test scenarios.
- Viya is an average of 49 times faster than a Spark-based data and AI platform across all algorithms, data sizes and workloads. In some cases, Viya is up to 326 times faster.

The example presented in Figure 4 shows how well the CPU was used during the workload execution. Because Viya has efficient

execution times, the CPU requirements across all cores are highest for eight and 16 cores in this test. The pink line declining indicates that the workload is about to complete, and the free cores can now be reallocated for other work as needed. The OSPs remain high in CPU usage across worker nodes but are not vastly improving the runtimes, like driving a car in second gear and redlining the engine without going any faster.

Figure 4: CPU use vs. available cores



“SAS Viya produced results in minutes, while competitors ran for hours or simply failed.”

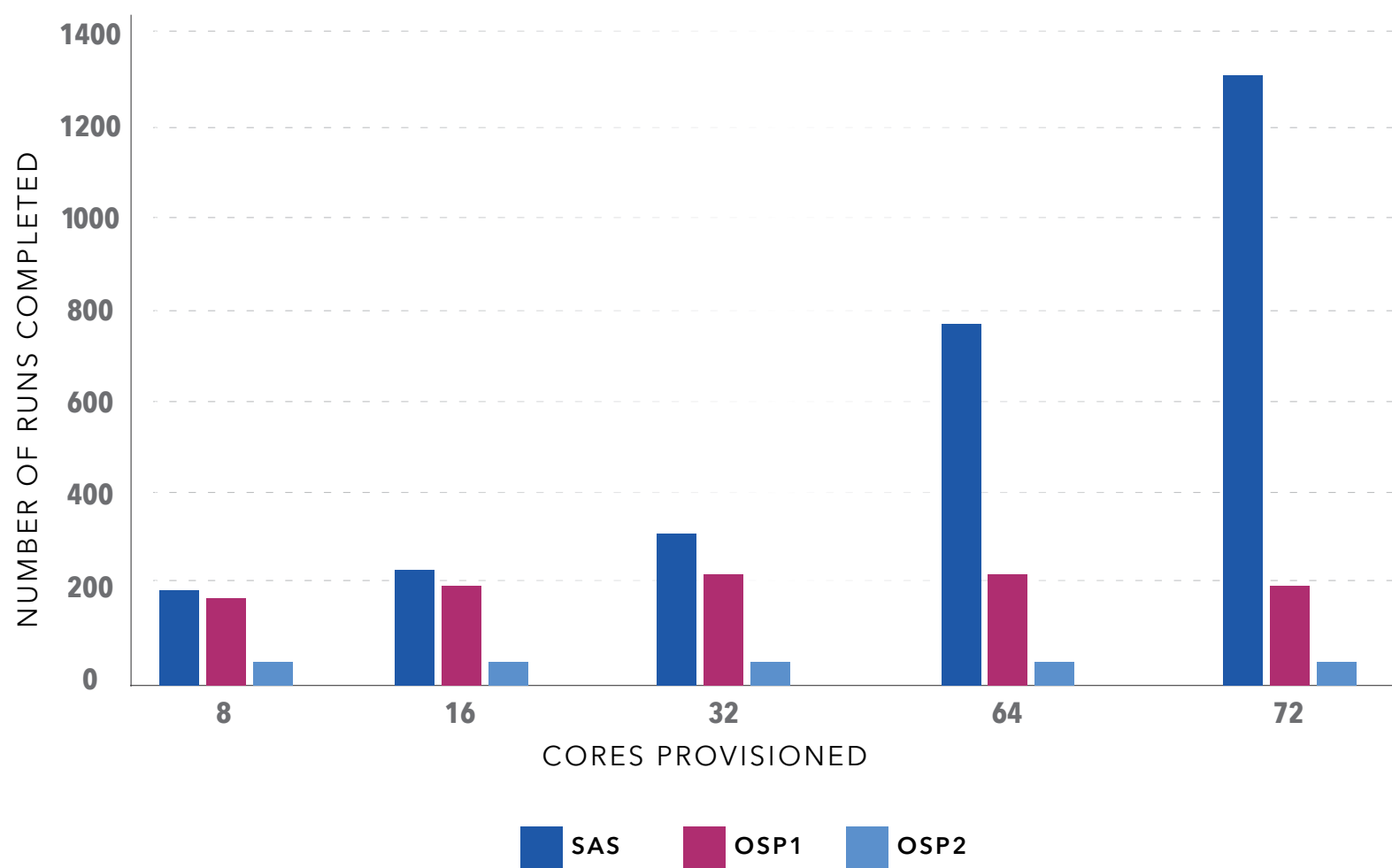
Russ Fellows, Senior Partner and Analyst at The Futurum Group

The value of optimized performance

Performance is money in the cloud computing world. To save on unnecessary costs, most companies choose to turn off cloud resources when they complete analytics workloads or during off-peak hours for nonproduction environments.

To achieve cost-efficiency and maximize productivity, you must account for the time that cloud resources are used, because this translates directly to cloud spending, and the number of workloads you can execute when your cloud resources are up and running. Another important consideration is the efficient use of the available CPU.

Figure 5: Number of runs in one hour at 100% CPU utilization



The theoretical example illustrated in Figure 5 shows that an organization that uses all the available computing power with SAS Viya runs significantly more workloads compared to both OSPs. As a result, organizations benefit in multiple ways by either:

- Reducing the computing power needed to minimize costs while still achieving the desired business outcome.
- Or by using the same computing power to:
 - o Run more analytics workloads in less time to achieve better results due to the iterative nature of analytical processes (tune, run, examine the results and repeat).
 - o Turn off the cloud resources faster to save on costs.

It's worth mentioning that it's possible to improve runtimes in all three cases by further optimizing the code (the default settings were used for Viya and OSPs), but this would require significant levels of expertise, time and effort that all translate to a large increase in costs.

- Our analysis shows that the result is a significant reduction in the cost of computing and an increase in productivity that can be realized by organizations when performance is optimized. The Futurum Group compared Viya with alternatives in the market and validates that users can lower costs by 86% on average because of the speed and scalability of Viya in developing and implementing machine learning models.

The development community of open source is vast, but being the best depends on providing confidence and trust

The cloud and distributed computing introduce new constraints that affect performance, computing costs, and the repeatability and accuracy of results. Comparing SAS and OSP workloads is just one example of how SAS is evaluating leading open source and commercial alternatives against SAS Analytics.

By embracing open source algorithms, Viya enables your analysts to pursue the best outcome every time regardless of algorithm. However, when it comes to putting analytics into production, you need certain attributes, including the ability to generalize to different data set sizes, types and shapes while managing model drift.

A SAS algorithm must be:

- Efficient.
- Effective.
- Cost-conscious.
- Repeatable.
- Accurate.
- Ethical and responsible.

SAS also continues to test open source algorithms to evaluate their ability to provide the same attributes. If an open source algorithm delivers on most of those attributes, SAS will improve it to the point where it can be adopted on the Viya platform. This enables users of Viya algorithms to spend less time coding, testing, interpreting and trying to achieve better harmony between the algorithm, data and hardware. This frees them to focus on delivering results.

Help has arrived for cloud financial education

A growing concern about unmanaged cloud costs led to the creation of the FinOps Foundation. It's a cloud finance best-practices hub and certification provider that provides guidance on cloud financial management and helps influence the cultural shift precipitated by widespread cloud adoption.

The organization's goal is to maximize the business value of cloud spending that, according to the FinOps Foundation, requires "engineering, finance and business teams to collaborate on data-driven spending decisions." It's a new world in which business units share responsibility for cloud usage, keep cloud costs in mind and collectively make the best use of the various cost advantages offered by cloud computing.

These financial considerations highlight the need for analytics-driven enterprises to improve accuracy and repeatability. When the harmony of algorithm, data and hardware is less than perfect, errors and inaccuracies proliferate.

One processor can easily generate a single result with accuracy. But with distributed data, the ability to produce reliable results is limited by the amount of data that can fit on a specific chip. Unless a central organizing principle is applied, the distributed results will produce an incorrect macro result. It might be fast, but it is neither accurate nor repeatable.

Each time you distribute the algorithm to the computing chip in a different way, it will produce a different answer. Python manages this variability by requiring appropriate configuration of the algorithm to the computing chip. While this can be effective, it requires hand-coding by a user highly adept at such mapping – reducing productivity, increasing cloud costs and putting the analytical results at risk.

SAS helps ensure reliable results

SAS Viya notifies users that an inaccurate result is likely if the input is distributed, indicating that the best option is a single-threaded process to execute that workload.

Using all of these capabilities, Viya creates a solid foundation for the cornerstones of predictive modeling: cost-performance, productivity, accuracy and repeatability. Viya helps eliminate common risks faced by users who adopt open source to take advantage of perceived lower startup costs.

For example, using a less effective analytics package and execution engine while paying for bulk cloud commitments routinely costs more and can produce inferior answers. Whether migrating from traditional on-premises computing or reassessing your existing cloud economics, Viya can help you balance how cloud commitments are applied to variable workloads with greater flexibility and speed. We believe this is the ideal state of analytics performance in the cloud.

It's a new world in which there is a collective responsibility for cloud usage, staying up to date on cloud costs and making the best use of the various cost advantages offered by cloud computing.

Does cost constrain access and innovation?

Performance is more than saving time. Better performance means your organization can produce more. Automated machine learning (AutoML) results in many models that can generate good answers, but paying for each model iteration of a large project is far less than ideal. This is especially true if the technology is slow. The unintended result is that data scientists will limit their tournament to fewer models to optimize cost – to the detriment of the project.

In examples like the one shared in Figure 3, SAS' research demonstrated that SAS can run much faster than leading OSPs that use more infrastructure. This enables data scientists to run more models without inflating costs.

In short, bigger tournaments equal better answers. But the cloud cost structure requires an additional framework that has not been a consideration in prior computing ecosystems: achieving the desired goals while ensuring value within cost constraints.

All the steps required in developing machine learning models happen while the meter is running in the cloud. This offers a new dynamic for users who are now responsible for the cost of their analytic design time in addition to runtime costs.

Citizen data scientists and business analysts need to be able to perform higher-level analytics without being overburdened by cost constraints or writing a large amount of code. This is why no-code/low-code interfaces using SAS algorithms can be incredibly powerful. They avoid the coding and overhead. They lower time to code models, offer approachability to new

“The larger and more complex the data set, the bigger the advantage for Viya. Additionally, SAS was able to produce the same model results orders of magnitude faster, running in seconds or minutes, versus hours or failing completely for some competing alternatives.”

The Futurum Group, Performance at Scale - Comparing AI/ML Performance of SAS Viya vs. Alternatives, April 2023

users and eliminate the need to ensure the data, math algorithm and chip are as close as possible to perfect harmony.

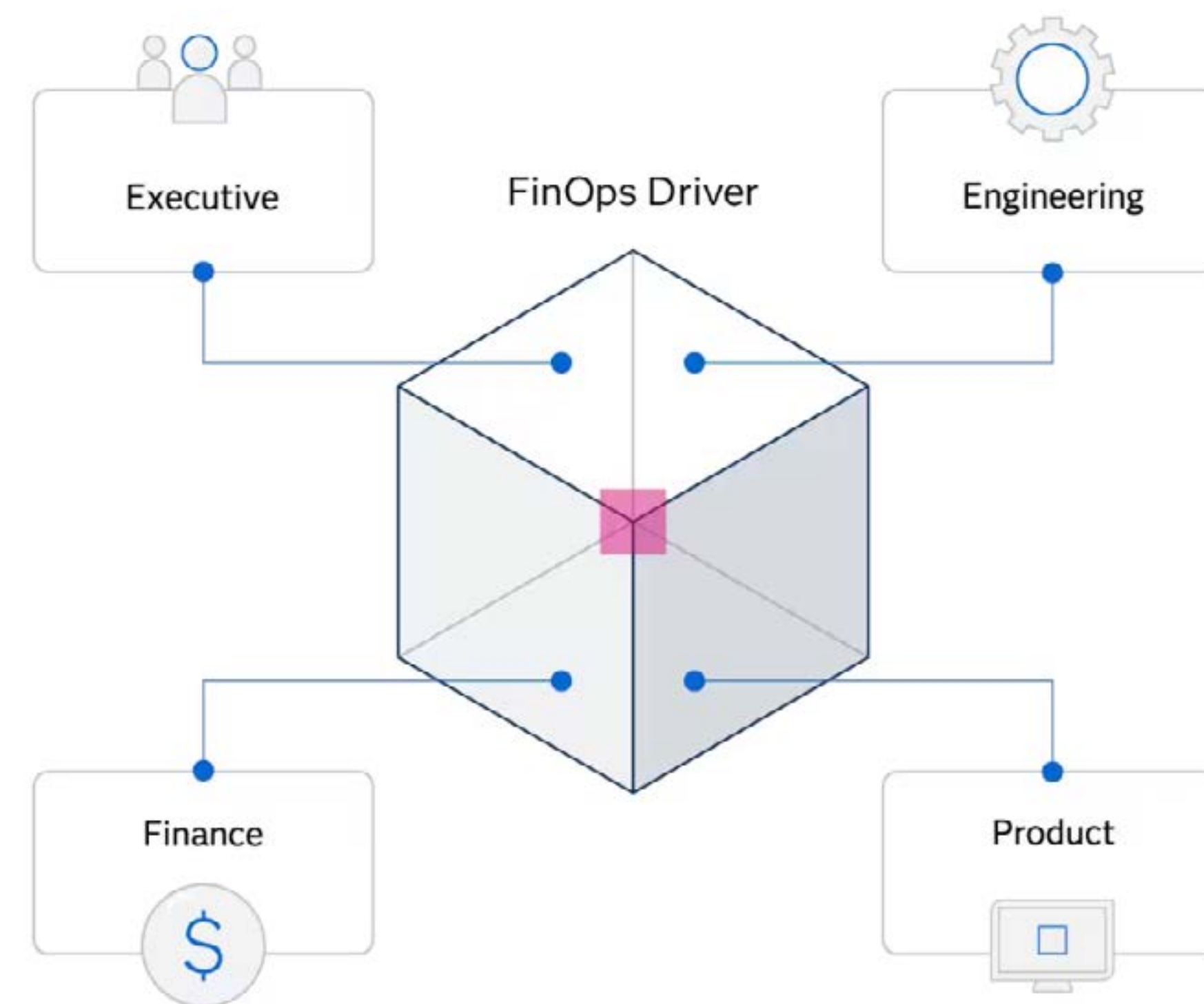
This new dynamic highlights the importance of cost-performance and why SAS is highly attuned to this issue. In the on-premises world, constraints are addressed by buying more cores and capitalizing the expense – a long-term commitment with predictable IT spending. In the cloud world, IT spending happens at the edge, with resources scaled according to need. This creates challenges for forecasting IT spending and hampers the organization's ability to benefit from heavily discounted long-term cloud commitments. This change in buying behavior increases the risk of overpaying for cloud computing.

The other element in this dynamic is that the cost decisions are now in the hands of the teams running the models – not with the IT or purchasing departments. In the cloud, system engineers procure the infrastructure, choose the algorithm, select the amount of data – and now carry the fiscal responsibility when they run their workloads.

As a result, organizations are trying to find FinOps unicorns to guide executives, finance, system engineers and product owners toward more fiscally responsible engineering in the cloud.

Analytics users want to spend their time finding better business solutions. Instead, they're spending time writing code, ensuring it is tailored to be cost-performant, or finding themselves limited to building and testing fewer algorithms and models to stay within cost constraints. Some users may even be excluded entirely from pursuing analytics due to these considerations.

Figure 6: A FinOps unicorn coordinates cloud finance between stakeholders



The cost-performant approach

Achieving cost-performance in the cloud provides organizations with the freedom to choose how they optimize their analytics strategy to drive down costs and maximize productivity. SAS suggests an approach of continuous improvement.

Be flexible and innovative by empowering users to manage their cloud consumption.

The cultural empowerment of cloud financial management is essential to prevent excessive cloud costs and unnecessary workload repatriation. SAS offers a solution for ecosystem diagnostics to identify and improve the cost-performance of resource-hungry code.

Consistently assess performance to control cloud costs.

The research shows that similar approaches can have vastly different outcomes in the cloud. Being able to effortlessly achieve better performance at scale allows cloud costs to be brought into line with expectations. Tight harmony between algorithm, data and computing chip allows algorithms to run faster and reduces cost considerations by several orders of magnitude.

The shift to cloud has increased the ability to mine untapped value from data, but at the risk of driving costs into uncharted territory. Changing the behavior of data scientists, analysts and engineers and the code they run is a new challenge and an evolving process.

[The Futurum Group study](#) shows that “SAS Viya’s ability to deliver the same results for 86% lower cost on average or deliver 30X more results for the same resource utilization provides a high degree of flexibility to companies looking to optimize operational costs and AI/ML results,” says Russ Fellows, Senior Partner and Analyst at The Futurum Group.

This equates to a continued, compounding return on investment. It is now in your hands to develop an effective strategy that focuses on maximizing productivity and driving down cloud costs by capitalizing on the power of SAS Viya.

“The performance of SAS Viya was impressive. They didn’t just outperform competing AI/ML libraries, they crushed the competition.”

Russ Fellows, Senior Partner and Analyst at The Futurum Group

SAS® Viya®

**Get more information about the SAS Viya benchmark study by
The Futurum Group**

Download the executive summary

[“The Futurum Group: Faster AI & Analytics: SAS Viya Outperforms the Competition”](#)



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