



INTERNATIONAL
INSTITUTE FOR
ANALYTICS™

RESEARCH BRIEF

RESEARCH & ADVISORY NETWORK

Refactoring Analytics for the Cloud

DAVID MACDONALD

Executive Vice President and Chief Sales Officer, SAS Institute

ROBERT MORISON

IIA Lead Faculty

APRIL 2019



Introduction

More and more organizations are leveraging cloud computing in pursuit of tangible benefits of agility, scalability, and cost savings. Many analytics applications are natural candidates for migration to the cloud because they require very large amounts of data and computing power, but only temporarily while large-scale models run. The migration is on, the opportunities are great, and the landscape for analytics in the cloud continues to change. As organizations have gained collective experience moving analytics to the cloud, we have a clearer picture of migration benefits, options, and best practices.

Hybrid Approach

Many organizations are using a hybrid approach when deploying analytics in the cloud. Seventy-five percent of the organizations that we work with are either using or plan to use a public cloud by the end of 2020. These plans are based on quantified value creation at the application level or even at the workload level. Those in unregulated industries tend to lean more toward the public cloud, those in regulated industries more toward keeping analytics involving Personally Identifiable Information (PII) in-house.

CIOs and CEOs are thinking, “We can leverage what is available in the public cloud, but let’s be diligent about this. Let’s keep our options open, be flexible about what analytics are processed where, and avoid being locked into a single cloud vendor. Let’s determine what’s best for our organization.” For most, that means the flexibility of a hybrid platform for analytics.

Tangible Benefits

Agility is the number one driver of cloud migration. Organizations can access large-scale machine learning infrastructure in the cloud as needed on an intra-day basis. They have more choice and can take advantage of new hardware such as GPUs that are not readily available in-house. They have the agility to innovate – experiment fast, succeed fast, fail fast – by leveraging the right technological capabilities at the right times. With analytics in the cloud, organizations can learn, decide, and act faster.

A close second is scalability. Organizations can accomplish long-running compute-intensive analytic tasks, especially the ones that are run infrequently, in the cloud more efficiently. An auto parts supplier spins up over 8,000 containers in a couple of minutes and then shuts it all down 20 minutes later after completing a resource-intensive optimization routine that previously took over 24 hours to complete on-premises.

On-demand scalability reduces cost. Properly done, using the cloud for the right amount of infrastructure resources on demand can lower cumulative operating costs. One organization stores its data on the low-cost AWS and spins up the infrastructure required to analyze the data as needed. Once the models have run and the procedure is complete, the infrastructure is released and no longer charged for.

Managing the Data

A big variable for analytics in the cloud is the complexity of synchronizing the data needed for analytics. Matters may be simple with one-off data sources, perhaps including public or other external data. Gather it in cheap storage in the cloud, analyze it, and move on. Or if you’ve got large amounts of non-



sensitive data, which increasingly includes streaming data, the cloud can be the best place to run analytics.

But it's a different story if you need to synchronize with the organization's system-of-record data – not just extracting it from databases but posting it there or embedding it in the everyday information systems supporting employees' workflows. You need to look at the nature of data, how it's processed, and how it moves to determine whether and how the analytical models using it can migrate to the public cloud. Weigh the benefits of the cloud against the application and data risks you may be incurring.

For both functionality and flexibility, a hybrid private-public cloud platform may serve best. This may be familiar territory in organizations that are using HR and sales automation applications that are available only in the public cloud. They are gaining experience in wrestling with data movement and protection issues.

Financial Services in the Cloud

It says a lot when even the regulators and clearinghouses are moving to the cloud. With significant improvements in cloud security, both the U.S. Office of the Comptroller of the Currency (OCC) and the Depository Trust & Clearing Corporation (DTCC) are planning to move their operations fully in to the cloud in the next few years. At the same time, however, regulators are concerned about supervising global financial institutions that are progressively moving their customer data and their banking systems to the cloud. In its semi-annual risk perspective (Fall 2018), the OCC identified the top three innovation trends for banks as cloud computing, artificial intelligence/machine learning, and digitization of existing processes and products. These trends could enable financial institutions to reduce costs, increase efficiencies, and improve customer experiences – provided they can manage the risks to customer data.

Migration Paths

There are three basic options for migrating analytics to the cloud, and they have very different profiles in terms of effort and value.

- Lowest effort and value come with *rehosting*, or “lifting and shifting” analytics applications to a cloud provider. You're simply continuing what you do today and taking advantage of the economics of a different data center. Rehosting doesn't maximize agility and scalability benefits. It's more of a pure cost play, and savings can be reduced by the added costs of data movement.
- The second approach is *recoding*, where you recreate applications in a different tool to take advantage of the cloud. Recoding brings the benefits of cloud, but often at a steep price. A major insurer has over 300,000 analytics programs and estimated that recoding them would cost over \$50 million and a prohibitive amount of time and effort. Recoding runs the risk that the new model may operate differently, and the organization may be unnecessarily abandoning a great deal of intellectual property in the original code.
- The best alternative is *refactoring*, or repackaging existing production analytics routines to run on cloud platforms. Analytics applications are now containerized to be portable and scale on demand. With this architecture, existing applications can run seamlessly in cloud-native environments. Analytics code is largely unchanged, so the IP is protected and the models continue to function as designed. All the cloud benefits accrue.



In practice, organizations can combine the three approaches. Stable, commoditized analytics that run frequently and have clear runtime definitions are good candidates to be rehosted. Selected applications might be recoded, especially if they are due for a functional or performance overhaul anyway. And the best bang-for-the-buck approach for most cloud-bound analytics is likely to be refactoring, especially high-value compute-intensive workloads that are run infrequently and require substantial infrastructure.

Dramatic Benefits of Refactoring

A major provider of data and services to the financial services industry works with vast amounts of data – hundreds of millions of records with thousands of attributes and multiple years of history. They provide access to all this data along with a private analytic environment for each of their clients. Prior to refactoring in the cloud, this service was delivered on static infrastructure with slow data transfers that undermined the client experience. And it was too expensive. The platform was not performing as expected, and was constrained in bringing in new clients or adding new data.

By refactoring their workloads, including taking advantage of cheap storage, smart scaling and distributed in-memory capabilities, they improved data transfer speeds and analytics performance tenfold, with a substantial reduction in operating cost. Much of the cost benefit was passed along to their customers, who also gained great productivity benefits.

Migration in Stages

We recommend starting by profiling your analytics workloads. Which are cloud candidates to begin with? What is the best migration approach for each

candidate? Apply the 80/20 rule – which workloads will provide the greatest value from refactoring?

We do not recommend refactoring directly into a public cloud in an all-at-once fashion. Instead, refactor first to on-premises cloud-ready infrastructure. That way, you can demonstrate the viability and value of the refactoring approach for the workload, test the use cases in your own on-premises containerized environment, and anticipate what specific cloud benefits you're after. Then you can choose when and how to migrate workloads to a public cloud, and you can deploy them there with less effort, less risk, and greater confidence. Actual migration will then be simplified, and you retain the flexibility to move the workload back in-house or to a different cloud provider.

Organizations are recognizing that it's a sensible strategy to get analytics containerized and cloud-ready on-premises first. In the process, many are also finding that they may be able to push more workloads to the cloud than they initially thought.

Inventory and Triage

What analytics applications belong in the cloud? The answer starts at the endpoint – what are your most important goals? Are they more about reducing costs? Or more about agility and new kinds of value? Where are your organization's greatest opportunities and imperatives with respect to analytics? Decisions about workloads happen individually, but it helps to have this broader context in which to profile and triage the applications.

Some organizations have developed a large analytics footprint over time. They are very focused on rationalizing or consolidating applications and reducing the costs of running all their models. They



should conduct an inventory and assessment of workloads because they're likely to find that a small fraction of them are consuming the preponderance of computing resources. Refactoring and selectively recoding those workloads can reduce consumption dramatically, and then gain more savings by migrating to a less expensive public cloud.

Periodic inventory and assessment are valuable independent of cloud migration plans. You'll find simple fixes and applications to retire. Also look at datasets created and saved – are they really necessary, and where are they best stored? Be sure to profile applications in terms of business value, functionality, cost, and risk.

Analytics Governance

This profiling process and preparation for refactoring is really a form of analytics governance. You're getting a consolidated view of analytics assets. You're better able to control applications and data, as well as orchestrate the people and skills behind them. You have solid information for triaging and prioritizing prior to any large migration to the cloud.

Mature organizations recognize the importance and value of governing and managing the complete lifecycle of analytics – from experiments that yield knowledge but don't make it to production, to moving models into production and monitoring their performance, to as-needed refreshing of models, to their eventual retirement or replacement, and to documenting them along the way to maintain manageability. Lifecycle management facilitates the entire process of analytics inventory, assessment, triage, and migration.

Pitfall to Avoid

The biggest pitfall in cloud migration may be the temptation to recode, or the misconception that you have to recode and use more open source technologies to maximize the performance and flexibility of analytics in the cloud. As mentioned, recoding can be extremely expensive, and organizations underestimate the value of the IP in their existing models and analytic processes. The real question is, “How can we leverage our IP *and* gain the benefits of cloud and avoid exposing the business to unnecessary cost and risk that rework involves?”

For example, an organization was rewriting code to take advantage of the Spark in-memory framework for scoring, when they could have simply published their SAS models in Hadoop and executed in Spark without any code changes. We recommend working with your software, tools, and infrastructure providers to stay up-to-date on the options and methods for analytics in the cloud, and to plan your migration accordingly. If the providers aren't informed about the options and flexible in incorporating them, that may be a warning sign.

The philosophy here is to always be additive. Create new value in the cloud without having to go back to the drawing board or change what is already working well.

Getting Started

If your organization is thinking about migrating analytics to the cloud, first clarify your objectives. What mix of agility, scalability, and cost benefits is your organization pursuing? Does it want to innovate and do things it can't do today? Or consolidate assets and lower costs? Remember to ask, “What's the cost of



not doing anything?” That’s often an unanswered question.

Then start planning. Do the inventory and assessment and triage. What analytics applications and data can migrate easily? What applications benefit most from refactoring? Does anything really need to be recoded – is it worth doing the rework? What is definitely left in place on-premises, at least for the present? What will it take to get workloads cloud-ready before actual migration? And will my vendors make the journey easier for me?

Course Corrections

Many organizations are midstream or have at least made a start with analytics in the cloud. There has perhaps been a fair amount of trial and error involved, and they are doing their best to follow their roadmaps. However, if those plans were developed a year or more ago, chances are they entail too much recoding and expense. Technologies and methods have advanced significantly in that space of time on three fronts: migration methods, hybrid platforms, and analytics capabilities – including machine learning – in the cloud.

It's wise for organizations to pause and reevaluate periodically. If you're midstream, it's worth taking a checkpoint. Review your broad objectives and destination. Take a fresh look at the landscape and what's possible. You're unlikely to say “stop the presses” and change direction entirely. But you may find that the market has moved much faster than your migration. There are likely course corrections to make, maybe even on basics like who is the best cloud provider for your needs.

Conclusion

The days of trial and error in cloud migration should be over. You can fast-track selected analytics into the cloud by leveraging expertise, using the latest methods, and focusing on the end state. You can find ways to create more business value with analytics in the cloud – and ways to get there faster – than you initially thought possible.

Additional Information

To learn more about this topic, please visit www.sas.com/cloud.

This piece was preceded by: “*Migrating Analytics to the Cloud: It’s About Time.*”

About the Authors



DAVID MACDONALD

EXECUTIVE VICE PRESIDENT AND CHIEF SALES OFFICER

David Macdonald is responsible for SAS' global sales in 59 countries. A sales leader, Macdonald has more than 23 years of sales experience focusing on financial services and technology across a variety of industries. He believes in partnering with customers to provide them with the right solutions that align with their business strategies. Having worked with many Fortune 500 companies, Macdonald understands how analytics can uniquely position businesses and organizations for success in their respective markets. He is passionate about empowering sales teams, arming them with trust and the resources necessary to serve their customers.

Prior to his current position, Macdonald was the Vice President and General Manager for SAS' Financial Services team. In that role, he led a highly skilled sales and pre-sales technical team providing financial services companies with the analytics they needed to stay ahead of customer demands. With an established career in technology, Macdonald is adept at identifying opportunities and working with customers and partners to tap into exciting areas such as artificial intelligence, cloud, big data and analytics. He has helped financial institutions implement SAS® analytics that can scale throughout the organization to achieve better performance and competitive differentiation.

“One of my goals as the CSO is to make it easier for organizations to harness SAS in their digital transformation journeys,” said Macdonald. “I want the global SAS customer-facing teams to help our customers set a vision and play a material role in achieving it. It is very rewarding to be a part of the changes we see our customers going through and giving them the technology and expertise that contributes to their ultimate business success.”

Macdonald has more than 30 years of business and technology sales experience with companies like IBM and Intrinsic. He received his bachelor's degree in engineering from the University of Dundee in Scotland. He is an active Victory Ride committee member for the Jimmy V Foundation for cancer research.



ROBERT MORISON

Robert Morison serves as Lead Faculty for IIA's Enterprise Research Subscription. An accomplished business researcher, writer, discussion leader, and management consultant, he has been leading breakthrough research at the intersection of business, technology, and human asset management for more than 25 years.

He is co-author of *Analytics at Work: Smarter Decisions, Better Results* (Harvard Business Press, 2010), *Workforce Crisis: How to Beat the Coming Shortage of Skills and Talent* (Harvard Business Press, 2006), and three Harvard Business Review articles, one of which received a McKinsey Award as best article of 2004. He holds an A.B. from Dartmouth College and an M.A. from Boston University.

IIANALYTICS.COM

Copyright © 2019 International Institute for Analytics. Proprietary to subscribers. IIA research is intended for IIA members only and should not be distributed without permission from IIA. All inquiries should be directed to membership@iianalytics.com.