StreamSets
A SOFTWARE AG COMPANY

Multi-cloud Matters
A Business Case Study for a Multi-cloud First Approach

" To expect the unexpected shows a thoroughly modern intellect. "
– Oscar Wilde

This wisdom still holds as companies continually evolve their modern data ecosystems. Those with the quickest minds are prepared to pivot away from the same absolutes they planned for only one year before.

Take, for example, a telecommunications giant we started working with a few years ago who told us in no uncertain terms that they were **not** interested in the cloud. It simply wasn't secure enough. They were building an on-premise data lake, and that's all there was to it.

Twelve months later, everything had changed. Faced with a looming deadline for a critical business initiative, the team realized they had to modernize their data architecture. The on-premise decision they had wholeheartedly committed to one year ago couldn't support the new initiative. They were now "all-in" on Amazon and asked StreamSets to help get their data to AWS from the on-prem data lake and various other on-premises databases too.

Within a few months, the migration was complete, thousands of StreamSets smart data pipelines were humming along, and they were delivering data to their new 'one true data source' – an AWS S3 data lake. Flexible architecture allowed them to make this unexpected change, but that's not where the story ends. The company then struck a strategic enterprise deal with a second Cloud Solution Provider, and a new mandate came down to move the analytics platform to Google.

But wait, there's more…

If this was even five years ago – or if they were using a different data integration solution, or the team wasn't thinking three steps ahead – they would likely have swapped AWS for GCP and been done with it. Instead, the team was ready to take advantage of multi-cloud, to use the right cloud for the right workloads, because they 'expected the unexpected.'

Even when GCP offered seven figures to move all the telco's pipelines to its native cloud technology, the team chose flexibility. The money was tempting, of course. But thinking about how quickly today's market changes (after all, just over a year before, they weren't even considering cloud), they knew they needed the ability to reuse existing investments and respond quickly to major strategy changes without financial penalty. Still, they had to justify turning down a seemingly great deal. So they did the cost analysis, and voilá! It became a simple decision. Rewriting existing pipelines would be 3-5x more expensive than leaving them with StreamSets, and timelines would slip significantly. And so began the journey to multi-cloud.

Our telco customer is just one example of many. Today, organizations must make decisions at the speed of need to improve innovation, reduce costs, and increase overall performance, and it's critical to have flexible infrastructure in place that can support those unexpected changes.

This executive brief summarizes a more in-depth white paper. Download now to learn more and get example pipelines for each stage of the cloud journey.

StreamSets
A SOFTWARE AG COMPANY
www.streamsets.com

Why Take a Multi-cloud First Approach?

Multi-cloud, which is "the deliberate use of the same type of cloud services from multiple public cloud providers," according to Gartner analyst David Smith[1], allows organizations to take advantage of the best features and benefits of different clouds. An essential component of that definition is the word 'deliberate.'

It's common for different parts of an organization to use different clouds for different projects and workloads. But when organizations use multiple public clouds without planning, that aren't integrated, they can run into the same data silo issues encountered with on-premise systems. And integrated data is not the only reason to go with a multi-cloud first approach.

The fervent desire to **avoid vendor lock-in** is a common refrain when implementing a multi-cloud strategy. Many IT leaders have realized that what was once perceived as a lower cost option than building and running your own data center can actually run up huge bills if left unchecked. It's critical to maintain negotiating leverage by building in the ability to move workloads across clouds upfront rather than risk being at the mercy of a single vendor with a stranglehold on your operations. If you architect for multi-cloud, you don't have to spend months redeveloping pipelines; you're already there.

Cost and performance optimization are the next big motivators. All CSPs are not the same. Not only do they have different strengths technically for different workloads, but they also have different economic levers. One may be lower cost for storage,

while another may be more cost-effective for compute. At a macro level, you want to be able to optimize costs by:

- having different vendors
- architecting the right mix of cloud environments
- and knowing which cloud to use for what workloads.

You can run into hidden costs when you're in a single cloud. For example, ingestion to raw landing zones is close to free in many clouds. But what about once your data lands? Suddenly, the data is not only distinctly not free, but it may be a lot more expensive than it needs to be. If you architect for multi-cloud, you can send data to whichever cloud makes the most sense from a cost or technical perspective. And to send it there affordably, you need a modern data integration tool that allows your team to scale.

Despite its enterprise status, the staff required to support pipeline creation and replatforming twice in a short period was beyond our telco example. But with a data integration solution that allows a single data engineer to support five or more developers (aka StreamSets), it was able to do just that.

Finally, **a centralized, top-down approach to IT is no longer realistic.** With the easy accessibility and affordability of the public cloud, shadow IT has become inevitable. In the past, it was common for organizations to have a locked down EDW and a governance model that kept tight controls over data usage. In a modern data stack, these constraints hinder innovation. Companies can now have multiple synchronized and reliable sources of truth with a multi-cloud approach. They can have agility with their compute and decouple that from the pipelines they build. And they can still design a governance model that keeps data usage on the right track while maintaining oversight of rogue workloads that were already outside their control. From our perspective, you want to "let 1000 flowers bloom" to encourage innovation (within a governance framework through which your organization must abide, of course).

[1] Multi-cloud strategy: Pros, cons and tips, CIO

What Makes Multi-cloud Work?

With the help of their implementation partner, our telco customer leveraged the **following distinct strategies to develop a data platform across multiple clouds and on-premises in under 12 months.** With these strategies, our customer architected an environment ready to run as the multi-cloud operational environment it is today – one that will be flexible enough for the unexpected changes that may occur tomorrow.

1. **Go cloud first and pick consistent technologies across public cloud and on-premises.** Rather than taking what you already have on-premise and deploying that in the cloud, take the simpler approach of selecting a technology that does the same thing. For example, Cloudera maps with EMR, HDFS maps to Amazon S3, etc. Compatibility saves time in implementation. And, since it can be difficult to build teams that know all the technologies required for data engineering, implementing cloud native technologies that are more self-managed allows you to add cloud to the mix without adding headcount.

2. **Go schemaless.** In traditional data integration, you need to know ahead of time exactly what data you will need, and in what shape it needs to be delivered for reporting and analytics. It can take weeks or months to map it all out. And just when you do, somebody has a brilliant idea and changes or adds a column – and your entire migration effort breaks. Our telco likens it to walking up to Walmart to buy bread but having to account to a bouncer at the front door who makes you explain exactly what bread you want, what the package looks like, and what aisle and shelf it's on before he lets you in. With a schemaless data integration platform, you don't

need to know anything about the data to move it. All you need is access to the data and to know the business meaning of it. The telco team points to StreamSets' schemaless design as essential to how fast they got three massive databases worth of information to the cloud, and subsequently multi-cloud.

3. **Modularize, decouple, and automate, automate, automate.** Automation allowed our telecommunications example to generate tens of 1000s of pipelines for the migration from on-premises to AWS and Google Cloud using StreamSets. The secret? With StreamSets, pipeline logic is decoupled from the specific source or destination. This means you can build modularized reusable components (aka fragments), then automate how the pipelines are composed of these different fragments and how they are deployed. The pipeline logic stays exactly the same; only the origins and destinations get swapped. And with StreamSets SDK for Python, the team took automation to a whole new level, setting up programmatic bulk process amendments.

4. **Make sure you have a single pane view of all of your pipeline and data engineering assets across all environments.** Our telco used StreamSets as the backbone of their multi-cloud data integration strategy to eliminate blind spots and control gaps. This allowed continuous observability, full transparency, and control across hybrid, on-premise, and multi-cloud environments.

Prepare Yourself for Change

Our customer's journey showed that change is the only constant in technology today. It's more important than ever for data architecture teams to embrace modern sensibilities by 'expecting the unexpected.' Where possible, avoiding CSP-specific technologies ensures:

- Agility to adapt to sudden changes in end-user preferences, business strategy, or technology partnerships
- Ability to maintain, leverage, and optimize existing investments
- Insulation from risk due to price fluctuations and ecosystem shifts by providing the ability to move analytical workloads to the cloud provider that best aligns with your requirements today and in the future.

Using a modern data integration platform like StreamSets to get to a hybrid/multicloud architecture gives you a way to move fast and realize ROI.
Our telco conservatively estimates they saved:

- Over $1M and 50 months of staff upskilling (2 months/data engineer)
- Over $350K in platform-specific tooling and 6 months of data pipeline re-engineering
- At least 5.5 months to launch, with the agility of cloud and StreamSets enabling them to reduce project launch time from 6 months to 2 weeks.

If you want to understand more about how StreamSets supported this telco – as well as large insurers, banks, and other enterprise organizations – in its on-premise to multi-cloud journey, book a session with your local advisor.

StreamSets can help insulate you from the inevitability of change.

About StreamSets

At StreamSets, our mission is to make data engineering teams wildly successful. The StreamSets DataOps Platform empowers engineers to build and run the smart data pipelines needed to power DataOps across hybrid and multi-cloud architectures. That's why the largest companies in the world trust StreamSets to power millions of data pipelines for modern analytics, AI/ML and smart applications. With StreamSets, data engineers spend less time fixing and more time doing. To learn more, visit **www.streamsets.com** and follow us on **LinkedIn**.

START YOUR FREE TRIAL

StreamSets
A SOFTWARE AG COMPANY