HFS

Crack your cloud-data integration conundrum with co-innovation

Infosys and its banking client's "Juniper" solution could simplify cloud data integration for the entire banking industry work to be successful in the modern economy

April 2020

Reetika Fleming, Vice President, Research

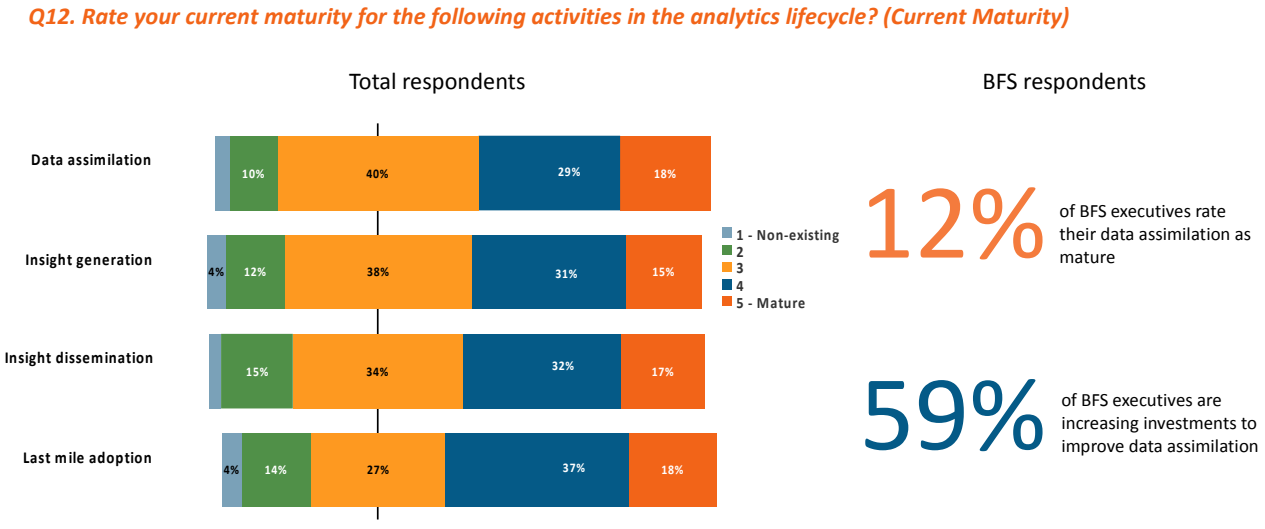
Defining Future Business Operations

Cloud data ingestion is the bane of every company trying to run large-scale analytics and ML programs

Every global business wants to have data-driven analytics built into the core of their organization. The intent is very clear: our research has found that 85% percent of enterprises view analytics as one of the top three strategic imperatives for market success, and a strong 81% have executive leadership teams aligned on an analytics strategy. Many have started down the path and invested in tools and platforms specifically for analytics, machine learning, and business intelligence. Parallel investments in foundational technology modernization also have a major impact, including cloud data migration, data lake creation, and core systems modernization.

Beneath the hood, however, not having quality data is a much larger underlying problem that can stall the most ambitious of analytics programs. The cloud is the way to go for data storage today, especially to run machine learning models that rely on large datasets. Data migration to the cloud is easier said than done, as every CIO has experienced over the last decade. In fact, as a recent HFS study on smart analytics found, data assimilation remains the biggest pain point across the entire data-to-analytics lifecycle (Exhibit 1).

Exhibit 1: **Data assimilation has the lowest maturity across the smart analytics lifecycle**



Source: HFS Research 2018, n= 262 Global 2000 Enterprise Leaders, 34 BFS respondents

Data assimilation—collecting, cleansing, and sorting data and making it available in the cloud—for analytics and ML is the least mature function in enterprises today. This holds true for banking and financial services

organizations as well, with only 12% of decision-makers rating their organization's data assimilation as mature. And yet, it is the most crucial and fundamental step for organizations on their way to becoming more analytically driven. Developing this capability will require organizations to make focused investments and undertake continual change management with the help of expert partners. In this document, HFS outlines the case example of Infosys and its banking client, which have done just that. The two companies have co-developed a product, Juniper, to solve cloud data ingestion and management once and for all—for the bank and, potentially, the rest of the banking industry, through an open-source offering in the works.

Cloud strategy underpins data and analytics dreams for this major financial institution

Infosys' client is a large banking and financial services institution. It has been on a technology transformation journey, initially moving to DevOps and agile from waterfall development a few years ago. Since 2017, the company has strategically partnered with Google to complement its multi-cloud environment already in place with AWS and Microsoft. The bank has an ambitious cloud strategy, and it is betting big on Google Cloud Platform (GCP). The company's technology leadership has made investments and plans around GCP to help grow its global business operations, letting its partners focus on infrastructure while it leads the charge on serving customers.

HFS had the opportunity to interview the bank's Program Director for Google Cloud Adoption and CTO for big data infrastructure and cloud migration. During our interview, the Program Director highlighted the key reasons for working with Google. Google's cloud capabilities give the bank the ability to scale, offer more elastic computing, and provide access to innovation capabilities. The company has also chosen Google as its preferred partner for data and analytics.

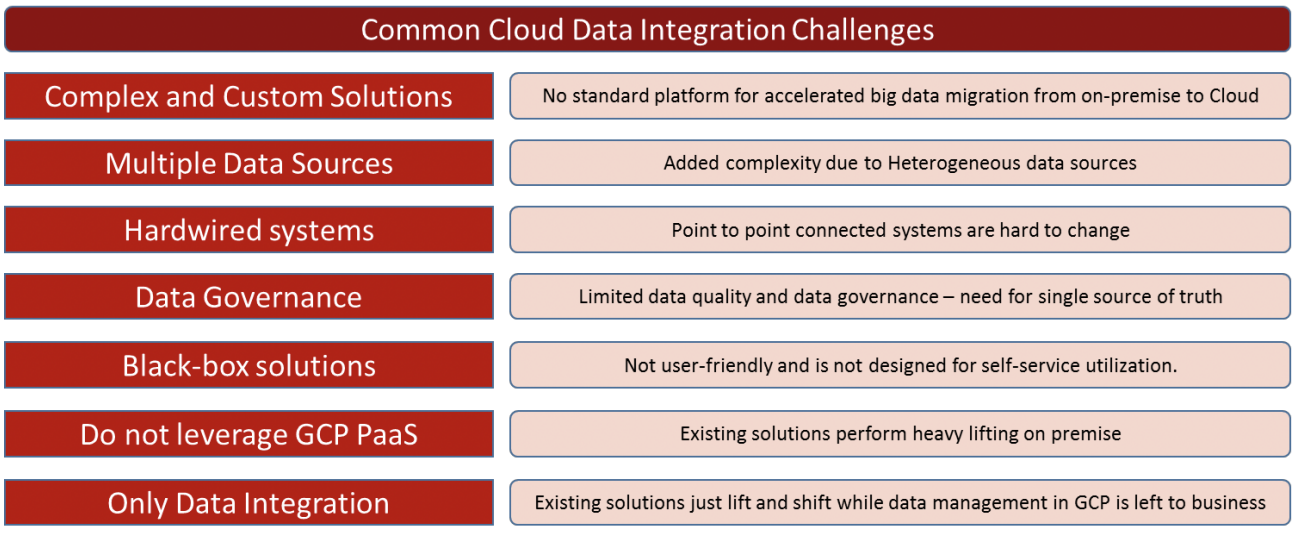
The bank is readying 300 projects for their move to the cloud. All of them, according to the Program Director, are based on a backbone of getting data from its raw form to a cleaned state into the cloud. Clean data in the cloud would form the basis for better analytics, shorter processing time frames, and forming conclusions from both large and small data sets. "Right now, we can't run queries on the last twenty years' worth of our data, but if the cloud lets us do that, we could find some really unique and interesting insights." Having the ability to ask questions on reliable, clean, and whole datasets would unlock further value from the bank's vast troves of historical data.

> We have valuable data sets, such as global trade flows, which we could process at scale to provide a lot of value while still meeting data privacy requirements. The potential outcomes include better products to customers, better financial crime detection, and more personalization to millions of customers around the world."
>
> **—Program Director for Google Cloud Adoption at the bank**

Moving complex datasets to the cloud emerged as a major bottleneck

Like many global enterprises, the bank's data landscape was spread out globally. This data existed in many source systems and technologies, resided in multiple databases such as Oracle and Db2, and was run on multiple technology platforms. Compounding the data complexity was the company's new cloud strategy. With the gradual expansion of Google cloud adoption within the bank, the company found the top challenge was on-premise-to-cloud data integration. Coupled with integration, other related constraints persisted, including data security and metadata-based governance to comply with regulations like GDPR.

Exhibit 2: **Cloud data integration poses many challenges for enterprises**



Source: Infosys Ltd., 2019

“We needed a product that would enable us to pull in data, regardless of source, and verify that we had indeed taken all of the data. We needed guaranteed data delivery at scale,” shared the Program Director. In the past, the bank had solved these problems at the project level using various methods, which created cloud silos. For each project, different project teams had approached data ingestion in different ways without consolidating best practices or investing to formally round out the solutions. Finally, the team decided to approach this challenge head-on by investing in a strategic platform for the entire company. “If you get data ‘almost’ right, it’s not good enough,” he noted from past experience.

Collaborating on a potential solution—the making of Juniper

The bank started ideating in December 2017 for a way forward to its cloud data challenge. At the time, Infosys, a global IT services provider, had a good working relationship with the data group at the bank. It had worked with the company for six years, delivering various services in data management. The Program Director’s team felt that Infosys would be a good partner to co-create the solution it was seeking. “It was clear from their experience that they could help us in our new venture,” he remembered. He added that Infosys had demonstrated the capabilities needed (cloud, data streaming, migration, etc.), and had a flexible working style that would suit this next venture, where the plans weren’t exactly laid out concretely from A-Z. The company identified the key issues as learnings for the bank, drawing from past experience (see Exhibit 2).

“We had to design the product by reverse engineering and learning from all our previous challenges. Infosys worked in a flexible manner with a good staff, which reduced the risk of going with a new vendor,” recounted the Program Director.

At the outset, the bank discussed the key goal with Infosys: to develop a reliable and scalable way of getting its information from on-premise to cloud storage. Infosys came back very quickly with an MVP, which helped fast-track the requirements development in an interactive manner. The bank highlights the fact that the two companies didn’t use a traditional waterfall development method of building for months and then reviewing. Instead, two-week sprints allowed the teams to collaborate and iterate faster on areas that needed more refinement. Working in this manner evolved a product that, in the Program Director’s view, is proving to be very useful for his organization. Infosys and the bank have both invested in resources to pull this solution together over the last year, christening it “Juniper.”

Juniper brings repeatable best practices for automated cloud data ingestion and management

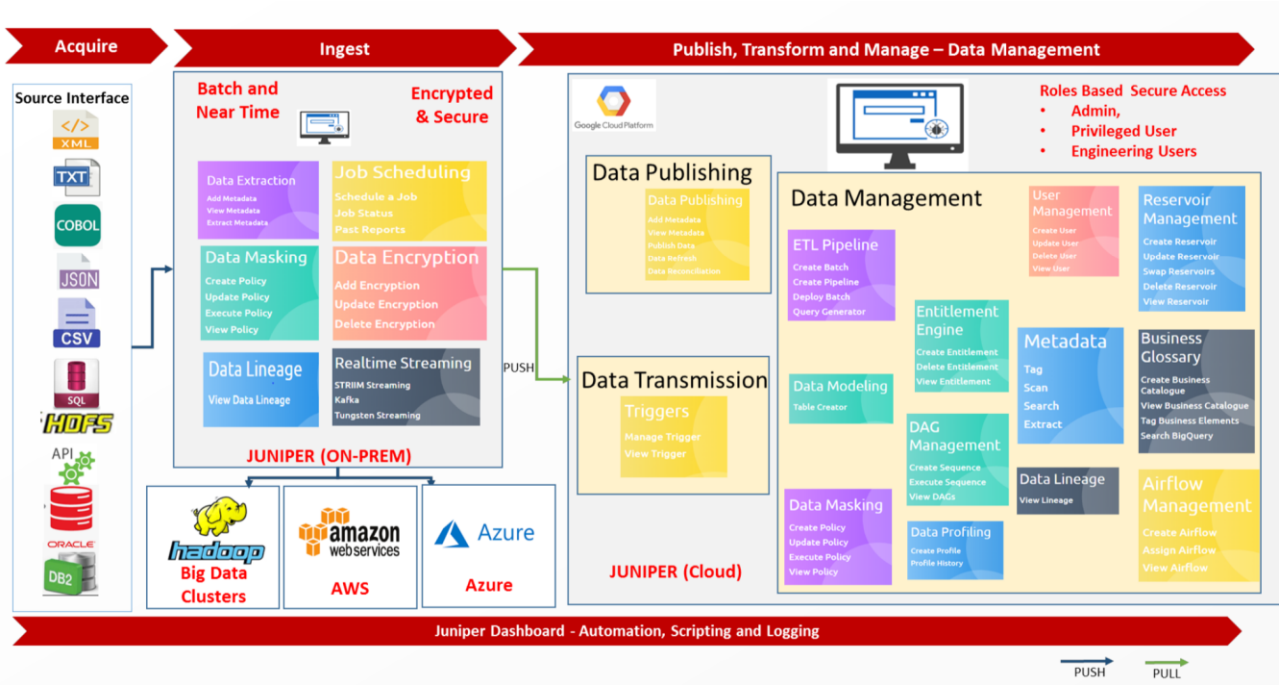
Juniper's design delivers on three key capabilities:

- Platform for open-source, scalable, multi-cloud ingestion and data management;
- Ability to replicate data from on-premise to a multi-cloud environment;
- Supports both batch and near real-time movement from on-premise to cloud, with a less-than-five-second lag.

Juniper has already gone live at the bank and is set to become a Tier 1 application, vital to the company's global cloud adoption. Faizan Mohammad, Business Client Partner at Infosys, described the thought process that led to its development with the bank, remarking, "We tried to see what it would take for a bank the size of [the bank] to start using GCP capabilities. We figured out that in banking, you cannot pull data from the cloud; rather, there needs to be a utility to push data to the cloud. We could automate that from different apps going from files to database to APIs." Accordingly, the teams created these UI-driven configurable connectors to open-source technologies to move data securely to GCP storage.

Once data had transitioned to GCP, Infosys and the bank's teams then evaluated how to materialize cloud data and make it usable and accessible while building in control frameworks. Juniper's capabilities were further enhanced using GCP's Cloud Dataflow, BigQuery, BigTable, and other components under the hood. Juniper features a user-friendly interface to orchestrate these elements and create control frameworks for how data is moved on the platform. Data lineage was another gap that Juniper filled through a solution to move from cloud storage to BigQuery. Faizan explained that Juniper's development team evaluated capabilities already present in GCP and the bank's tech landscape, and created complementary feature sets to best achieve the bank's goals.

Exhibit 3: **Juniper in action—acquiring, ingesting, publishing, and managing data**



Source: Infosys Ltd and the bank., 2019

Since going live, Juniper has already contributed significant benefits to the bank. According to the Program Director, the number of feeds that it is putting into production has increased dramatically. Compared to other systems and methods in the past, Juniper is much easier to use and is impacting faster turnaround times for data ingestion. Notably, the way the product deals with data challenges is curbing the number of failures that get put into production, ensuring that only high-quality data gets published on the cloud. The bank can also have 10 or so teams working on Juniper at the same time, which is usually a bottleneck process with other tools.

Key ways in which Juniper has impacted cloud data ingestion and management at the bank include:

- Guaranteeing data delivery at scale;
- Providing full reconciliation and dashboards;
- Data cataloging, helping with data linage and traceability, and improving the overall data quality;
- Integrating key management functions—store, scheduler, and alerting— enabling ingestion in hours rather than weeks.

Juniper's biggest impact, according to the bank, is the "velocity in time to market" that the product offers the company. "The value of Juniper is in what you can do with it," explained the Program Director, indicating that the product has great value as a foundational capability that impacts multiple digital and cloud initiatives for the bank.

Next on the horizon—analytics, ML, and a potential industry utility

The bank is very optimistic about Juniper's future within the company and anticipates continued investment for further expansion. Right now, the bank is using Juniper's data ingestion capabilities and starting to explore its data management features. Next on the horizon is analytics. The Program Director sets a broad mission for Juniper's impact, stating, "If Juniper is wildly successful and all data ingestion at the bank would go through it, our data would all be controlled, tagged, reconciled, and enable us to build analytics on that data with confidence."

After starting with four developers in October 2018, the bank now has over 200 developers working on Juniper-based data ingestion and management. Infosys continues to support the initiative, making up 25% to 40% of the resources allocated to Juniper. The intention for the future is for the bank to have multiple internal teams using and managing Juniper initiatives independently as its capabilities gradually increase and its adoption grows within the company.

The bank will continue to work with Infosys to chart Juniper's product direction and add to its source code along the way. The platform will eventually be available as open source, so Infosys intends to approach potential new clients to offer them support and managed services with Juniper as a utility. Working with multiple clients will make the product more robust, and the benefits will be available to all who use the open-source technology.

Juniper has already planned its product roadmap, including ways to enhance its capabilities. Infosys emphasizes three key features for the product vision:

1. The ability to automate data ingestions across multiple cloud environments, including on-premise;
2. End-to-end data management features, including data quality, data lineage, and data security;
3. A unified platform to enable data-driven analytics.

Infosys is working in parallel on another product called KEPLER, which will pick up where Juniper leaves off in data ingestion and help enterprises to better undertake data-driven analytics. KEPLER is part of the vision for leveraging Juniper's data management ecosystem. The analytics platform will prioritize self-service analytics for different business users. Infosys is currently working with the bank on KEPLER on the back of Juniper's success.

The Bottom Line: You need quality data in the cloud to become a data-driven organization with the ability to do real-time predictive analytics and ML. Find ways to solve this problem now to avoid data swamps and cloud silos in your future.

As businesses become more digital, the need for quality data will only become more pronounced. Businesses need to pick up the speed and get smarter about moving data to the cloud for analytics to avoid the problems of the past. With Infosys' help, the bank is on the path to having this superpower, which will undoubtedly help it make better use of its data assets. Both companies are invested in refining Juniper, the product of their continued collaboration. What is most compelling is that they are contributing this knowledge not just for themselves, but for other organizations facing persistent challenges in cloud data integration. HFS sees great promise in the utility model for a foundational capability such as Juniper and will continue to observe how Infosys shapes this platform as it takes on new clients.

A few learnings emerge from this bank's case example for peer organizations on a similar journey:

- **Don't replicate poor processes.** In this case, the bank identified how several project teams kept trying to solve the same problem in different ways, with no real consolidation of knowledge and significant resources spent in re-inventing the wheel.

- **Look for partners that can co-invest in something sustainable.** Infosys, in this case seems, to be a flexible partner that was willing to go at the bank's pace and solve their unique challenges, all with an eye on releasing something broad-based for the larger market.

- **Look out for Juniper as a potential solution for your cloud data initiatives.** As Juniper makes its way to Github, evaluate if it could potentially solve your challenges with cloud date ingestion.

HFS Research author

Reetika Fleming | Research Vice President

Reetika Fleming leads coverage for smart analytics, insurance, and finance & accounting at HFS Research. She studies the broad use of data and analytics within enterprises, with a research focus on emerging strategies to institutionalize machine learning and other AI techniques. Her research extends into the impact of digital business models, IoT, Smart Analytics, and AI on business process services for insurance specifically, and finance and accounting broadly.

About HFS Research: Defining future business operations

The HFS mission is to provide visionary insight into major innovations impacting business operations, including: automation, artificial intelligence, blockchain, Internet of things, digital business models, and smart analytics.

HFS defines and visualizes the future of business operations across key industries with our Digital OneOffice™ Framework.

HFS influences the strategies of enterprise customers to help them develop OneOffice backbones to be competitive and to partner with capable services providers, technology suppliers, and third-party advisors.

Read more about HFS and our initiatives on www.HFSresearch.com or follow @HFSResearch.

HFS