

#### CASE STUDY

### InfluxData uses RudderStack to Create a Single Source of Truth in Snowflake



#### Highlights

- InfluxData built a custom RudderStack connector to route siloed data from PostgreSQL to downstream destinations.
- RudderStack's pipelines and Snowflake's centralized data warehouse yield clean data and a single source of truth.
- InfluxData's analytics team uses RudderStack event stream data to optimize their marketing web pages and customize user experiences.

### **Key Stats**

- Creating a single source of truth eliminates 5-10 hours a week previously spent comparing and consolidating data from different applications.
- The data analytics team has end-to-end control of ETL pipelines, eliminating the need to hire a full-time senior data engineer to build and maintain them.
- RudderStack is helping InfluxDB minimize time to awesome for InfluxDB's 1,900 customers running 750,000 daily active instances.

#### **Overview**

InfluxData markets InfluxDB, a time series data platform for IoT, analytics, and cloud application developers. Comprising a robust API and toolset for real-time applications and a high-performance database engine, InfluxDB is optimized for time series data and features faster ingestion rates and custom retention policies purpose-built for handling real-time data.

It stores, processes, and manages high volumes of time-stamped data from IoT devices, sensors, applications, containers, VMs, and networks. InfluxDB supports all popular languages and frameworks and is backed by a massive open-source community. The platform is available in self-managed open-source and closed-source enterprise versions and as a fully managed cloud-based subscription service that scales to meet demand.

Launched in 2013, InfluxData now serves over 1,900 InfluxDB customers, from start-ups to blue-chip companies in sectors as diverse as fintech, retail sales, entertainment, manufacturing, and automotive. These include Capital One, Vonage, IBM, Nordstrom, Robinhood, RedHat, Walmart, and Volvo. InfluxDB provides scalable storage and fast access to continuous data for various use cases, including application performance and DevOps monitoring, predictive and real-time analytics, machine learning, network security, and IoT sensor monitoring.

With Telegraf, their open-source collection agent, InfluxDB can read, write, transform, and filter data from over 300 sources. InfluxData co-founder and CTO Paul Dix describes the platform as "improving the time to awesome" and making "the developer experience a priority." InfluxDB has over 750,000 daily active instances. With InfluxDB Cloud, powered by IOx, their new storage engine and recently-added SQL hooks for writing queries and BI tool support, the platform offers InfluxDB's growing user base real-time query speed, unlimited time series volumes, and the observability of metrics, logs, and traces in a single database.

## Challenge: Eliminate Data Silos and Create a Single Source of Truth

InfluxData had data residing in different systems, and none of it fit together. For example, the marketing and sales teams would pull the same data and end up with two sets of numbers. They were losing time assessing which figures were correct instead of focusing on improving sales and enhancing the customer experience. around as needed."

"We were wasting 10 minutes at the start of every meeting debating whose numbers tracked instead of talking about things that mattered," explains InfluxData Director of Data Analytics Mona Sami. "It was bad. We needed consistent metrics, guardrails around data definitions, and better data hygiene. In short, we wanted to eliminate silos, put all our data in one place to create a single source of truth, and move it around as needed." "Our data stack was inadequate. We didn't have a central data warehouse or a place where our customer data lived. Our PostgreSQL data was siloed, and we couldn't route it to downstream apps without difficulty. Our engineers were hand-coding ETL pipelines and fixing them when they broke. We used RudderStack to build a supercharged data stack, create a single source of truth, and easily move data between our analytics sources and destinations."

- Mona Sami, Director of Data Analytics, InfluxData

The company was pulling data from PostgreSQL as well as their own InfluxDB and feeding it to Domo (BI), Salesforce (CRM), and Marketo (marketing automation). They were also using the free version of Google Analytics for rudimentary insights into user funnels and journeys. "We were ingesting data from different sources and playing around with it, but it was really messy," adds Sami. "It was hard to scale, and we needed to fool around in Java if we wanted to implement any sort of data governance or version control."

InfluxData also lacked a central data warehouse. The company's engineering team had to build custom integrations to get event and product data from PostgreSQL into Salesforce and their other tools. The solution was impractical and time-consuming. "The engineering team wasn't happy, and neither was our sales team. Without a data warehouse, my analysts couldn't deploy an ETL tool that could route clean data to and from our downstream apps."

The company had to overcome these obstacles, free engineers to focus on products, and empower analysts to leverage customer and event data. So, they adopted Snowflake as their data warehouse and RudderStack as their CDP. These two products played well together and integrated with the company's technology stack, breaking down barriers and paving the way for new approaches to data.

#### Solution: A Supercharged Data Stack with End-to-End ETL Process Control

InfluxData's top priority was routing data from their PostgreSQL backend to Snowflake. "At the time, RudderStack didn't have a built-in integration," explains Sami, "But it was the perfect solution in every other way. We deployed it out-of-the-box in a day, and our engineers spent the next two weeks custom-coding a PostgreSQL connector." It was the first step in expanding the company's data stack.

"Before we deployed RudderStack as our CDP, data flowed in one direction," Continues Sami. "We were moving data for PostgreSQL into Domo and Salesforce, with the latter routing it into Marketo. But there was no consistency, and it took a lot of work to answer the simplest questions."

Adding RudderStack to the mix supercharged InfluxData's data stack. Using RudderStack's pipelines, the company extracts PostgreSQL, Salesforce, Marketo, and JavaScript API data and routes it to Snowflake, Google Cloud, and a custom Python reverse ETL application. From there, data is sent back to Salesforce and Marketo and downstream to Domo and Zuora. The company also routes data directly to and from Snowflake and its own InfluxDB platform.

"RudderStack gave my data analytics team end-to-end control of our data pipelines," adds Sami. "Instead of hiring and training a full-time senior engineer to build out reliable pipelines, we use RudderStack ETL tools out of the box to get the same results. We reduced our time to value, added new tools to our stack, and started exploring ways to glean deeper insights from our data."

"Building a PostgreSQL RudderStack connector was the first step in expanding our data stack. We now have clean, consistent data and the ability to answer complex questions without struggling to extract data from our backend."

— Mona Sami, Director of Data Analytics, InfluxData

# **Results: A Flexible Data Stack That Meets Users Where They Are**

Taking ownership of the company's ETL pipelines and flowing data to and from destinations in InfluxData's stack was only the start. The next step was working with RudderStack's event streaming features. Sami began by collecting event data about users logging into InfluxDB Cloud and then expanded the use of RudderStack to analyze the marketing pages on the InfluxData website.

"Our primary motivation in choosing RudderStack was building ETL pipelines," explains Sami, "But additional use cases quickly presented themselves. The first of these was adding RudderStack SDK to our product. I wanted to see what led people to sign up for an account. Where were they coming from? Had they read an online review? Searched for us on Google? Clicked on a banner ad? Next, I started looking at our company website. I wanted to establish the value of our marketing and support pages and see their direct impact on conversions."

RudderStack's event streaming features also helped InfluxData with identity reconciliation. "We were using the free version of Google Analytics," continues Sami. "Everything was anonymized, and we couldn't see what our users were doing on our website or how they were using our product. RudderStack has increased visibility into user behavior and user journeys. It has given us deeper insight into our funnel. We have real-time feedback on who's doing what and when and can run A/B tests that let us customize the user experience. All of this is incredibly valuable, and we hope to delve further into what RudderStack can do to help us improve our product and optimize our website."

Sami is also quick to praise RudderStack's support team. "One of the keys to choosing an enterprise data tool is getting good support, and RudderStack has been exemplary in that area. When something goes wrong , it gets resolved quickly. No matter how urgent or stressful the situation, RudderStack's team provides timely and consistent support, and our people always feel good about the resolution."

Thanks to RudderStack, InfluxData has built a robust stack that incorporates a centralized data warehouse and seamlessly routes data between destinations. "We're starting to think about real-time analytics and the other possibilities of our data stack," concludes Sami. "Instead of worrying about extracting data from our backend, maintaining ETL pipelines, and hand-coding connectors, we're using RudderStack to meet our users where they are and improve their time to awesome."

"Maintaining data pipelines, hand-coding connectors, and worrying about extracting data from our backend prevented us from improving InfluxDB users' time to awesome. RudderStack gave us ownership of our data pipelines, allowing us to meet developers where they are."

— Mona Sami, Director of Data Analytics, InfluxData

