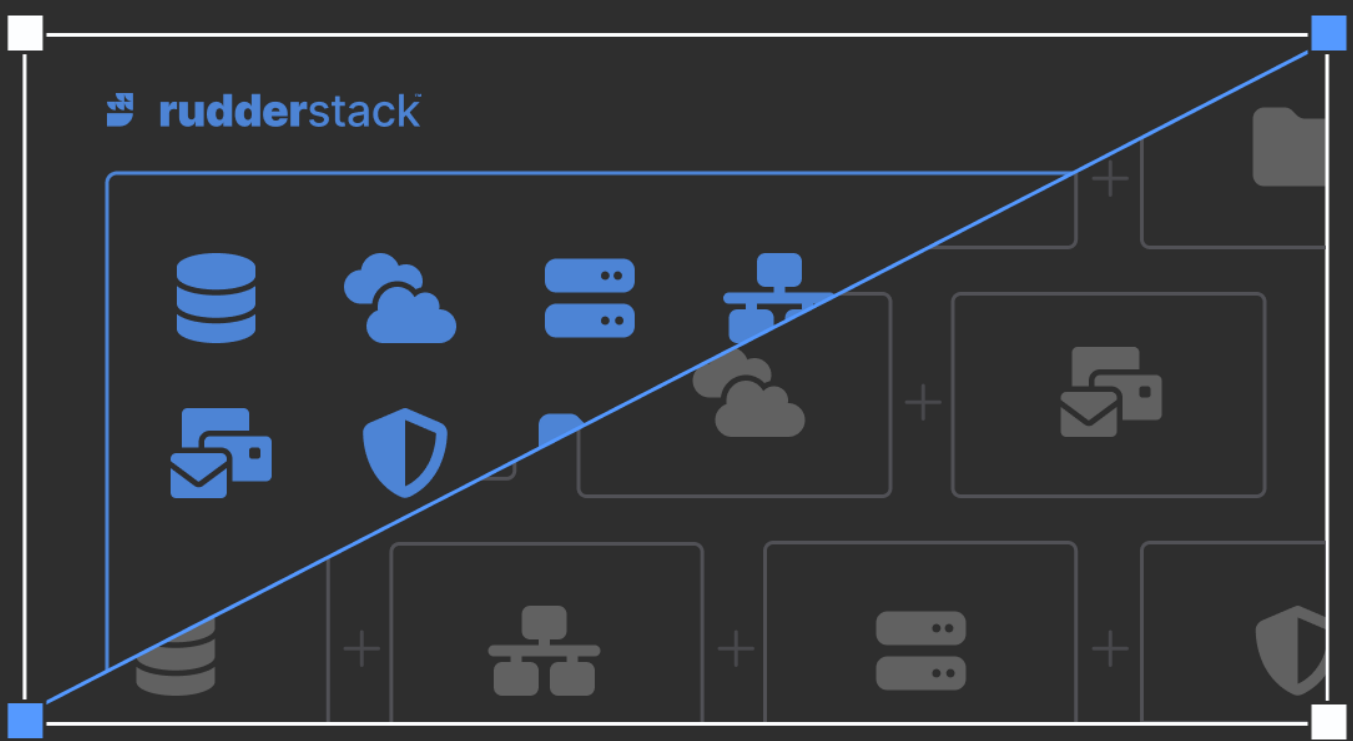
rudderstack™

The Definitive Guide

# Do more, spend less: 9 ways data teams are driving efficiency



# Do more, spend less: 9 ways data teams are driving efficiency

Table of contents



## Part 1: In 2023, the Data Stack gets a Reality Check

After a bull run decade, the markets in general have seen massive corrections. While the jury is still out on whether we will enter a recession and if it will be as bad as the 2000 bubble, companies from startups to large public enterprises are playing it safe and actively cutting down burn by laying off people.

There are still exciting technologies being developed that will shape the future of the data stack, but the hard reality is that both executives and individual contributors are thinking less about emerging technology and more about whether their jobs will be impacted. Data teams are no exception.

In response to uncertain times, data teams are giving their stacks a reality check and going back to the basics, making sure their work is mission-critical to their companies.

This often looks less like shiny new tech and more like maximizing the value of core infrastructure so you can do more with less.

## How data teams can make 2023 count

With this context, we'll outline the top 6 things we see our customers doing to ensure they and their businesses make every team member, tool and data point count this year.

### 1. Find every cost efficiency in your data stack

**Vendor cost control**

Many companies are getting rid of tools that are underutilized or not renewing contracts, but we also see customers getting creative on how to maintain tooling with better cost control.

**Here are a few specific examples from our customers:**

- **Leveraging data lakes for cost savings at scale:** regardless of the warehouse/lake/lakehouse debate, most companies run both a data lake and warehouse. We've seen multiple high-scale companies move ingestion primarily to their lakehouse and run jobs to get the data into their warehouse downstream.
- **Controlling volume for specific integrations:** Khatabook collects hundreds of millions of events per day, not all of which need to be sent to their marketing automation tool. They use RudderStack's Transformations to filter events to that destination, significantly decreasing cost (check out the webinar to get more technical details).
- **Leveraging APIs instead of SaaS:** Here at RudderStack, we hit the Clearbit API for enriching leads in our event pipeline (as opposed to paying for the Salesforce app).

**Eliminating hidden costs of in-house builds**

We've seen many data teams planning to mitigate one of the most pernicious costs burdening them: engineering time spent on maintaining internal builds.

Ironically, these often start as cost-saving exercises, but at scale, and when engineering resources are more scarce than ever, it can make far more sense to leverage your talent on business-critical work as opposed to pipeline plumbing.

Two specific examples we've seen recently from our customers:

- **Deprecating in-house or forked SDKs:** it's not uncommon for companies to build their own web and mobile SDKs for various kinds of data capture, but at scale dealing with

cookies and schema management is a huge lift and easier for a dedicated vendor to manage.

- **Deprecating basic in-house pipelines:** we see in-house pipelines for things like reverse ETL all of the time. As a one-off connection, they aren't too hard to build and run. As things become more complex, though, dealing with maintenance and volume becomes a time suck, causing many data teams to look for out-of-the-box replacements that cost far less than a data engineer (and can give those engineers their valuable time back).

## 2. Mitigate security risks

Companies can't afford unforced data security issues in 2023. A PR nightmare, security review and massive cleanup project could mean competitors get the upper hand, or even worse, threatens survival.

We've seen customers tackle security on a small scale and large scale. Here are examples of each:

- Implementing a data governance plan (for real) - Data governance is a journey, not a destination, but we've seen multiple data teams cleaning up data governance debt by updating and implementing things like integration-specific PII policies.
- Deprecating risk-prone tooling like Google Analytics - Litigation related to 3rd-party storage of personal data is showing up more often and we've seen many of our customers deprecate tools like Google Analytics and work with their analytics teams to rebuild that functionality on their data warehouse with a visualization layer.

## 3. Talk to stakeholders on other teams

Data and analytics are only a means to an end, which is some sort of business outcome like improving website conversion rates or product engagement. In 2023 the teams responsible for those use cases will be under immense pressure.

While data teams power the data flows behind the projects, they can be disconnected from the end business outcomes themselves. 2023 is the time to build those relationships and it starts with getting in touch with marketing, product, and CS leaders and developing deeper understanding and empathy for their use cases.

Not all these conversations need to be at leadership level—if you are an individual contributor on a data team, proactively reach out to a peer analyst working on a marketing use case and figure out what's most important.

For example, a senior data engineer who works for one of our customers began spending time with the paid advertising team to understand their needs and identified opportunities where data could help improve targeting and decrease acquisition costs significantly (20-30%!).

## 4. Identify skill gaps on the data team and proactively fill them

In the face of layoffs and hiring freezes, data leaders will be under pressure to deliver with limited resources.

Ideally the team as a whole can come together and determine how to fill gaps, but individuals can make a huge difference on their own, even if it's taking a course like Andrew Ng's introduction to ML to better support data science projects.

We recently saw a pipeline engineer at a customer company work with the analytics team to scaffold a dbt model for analytics. Their knowledge of the schemas and data sources sped up the project significantly and they were able to level-up their SQL modeling skills through a real use case.

## 5. Ship projects that impact business outcomes

If companies can't afford to face unnecessary security problems in 2023, they can't afford NOT to ship the key data projects that will impact the core business.

Many companies already know what these projects are, but have struggled, for a variety of reasons, to ship them.

In 2023, it will be critical for data teams to demonstrate business outcomes from their projects.

It's not enough to ship the recommendations project—it needs to demonstrably move the needle or be shut down quickly. Did the model actually increase repeat purchases? If so, by how much and what is the estimated revenue impact?

This is ultimately about turning the data team into a revenue center, not a cost center.

Whether data teams are building churn models, attribution models, full-funnel analytics or personalization, this is where we are seeing the most creativity relative to the cost-cutting measures mentioned above.

For example, one of our retail customers put off buying expensive enterprise software to drive recommendations and had their engineering and marketing teams work together to leverage existing infrastructure like Redis and their headless CMS to power personalization use cases.

### 6. Be vocal about the impact you are having

Taking on a personal marketing project might not sound appealing to data engineers, but the current environment means it's critical to create visibility around the impact you are having, both in terms of cutting costs and in driving results.

This is true for both leaders and individual contributors.

Present at an all-hands meeting, write internal (and external!) blog posts and look for opportunities to highlight your work at conferences (like our customer Acorns recently did at Coalesce).

---

## Part 2: Three architectures to make your data stack more efficient in 2023

In 2023, data teams are working hard to make every project and every piece of data count in the face of uncertain economic times.

Companies are accelerating projects that leverage data to become more revenue efficient, like personalization and churn reduction, but they are also scrutinizing expenses and ensuring that every piece of technology is pulling its weight in driving value.

Over the past several months we've worked with many of our customers to evaluate their stacks and find opportunities for efficiency.

Here are three effective approaches we've seen data teams use to leverage RudderStack for making their function more cost-efficient in terms of both time and technology.

### Time is money: eliminate low-value integrations engineering

The number of tools used by every team in modern companies is exploding (the average company has 75 tools!).

This often translates into custom integrations work for data teams across the stack, from ingesting data to creating custom pipelines that send that data to whatever downstream tools need it.

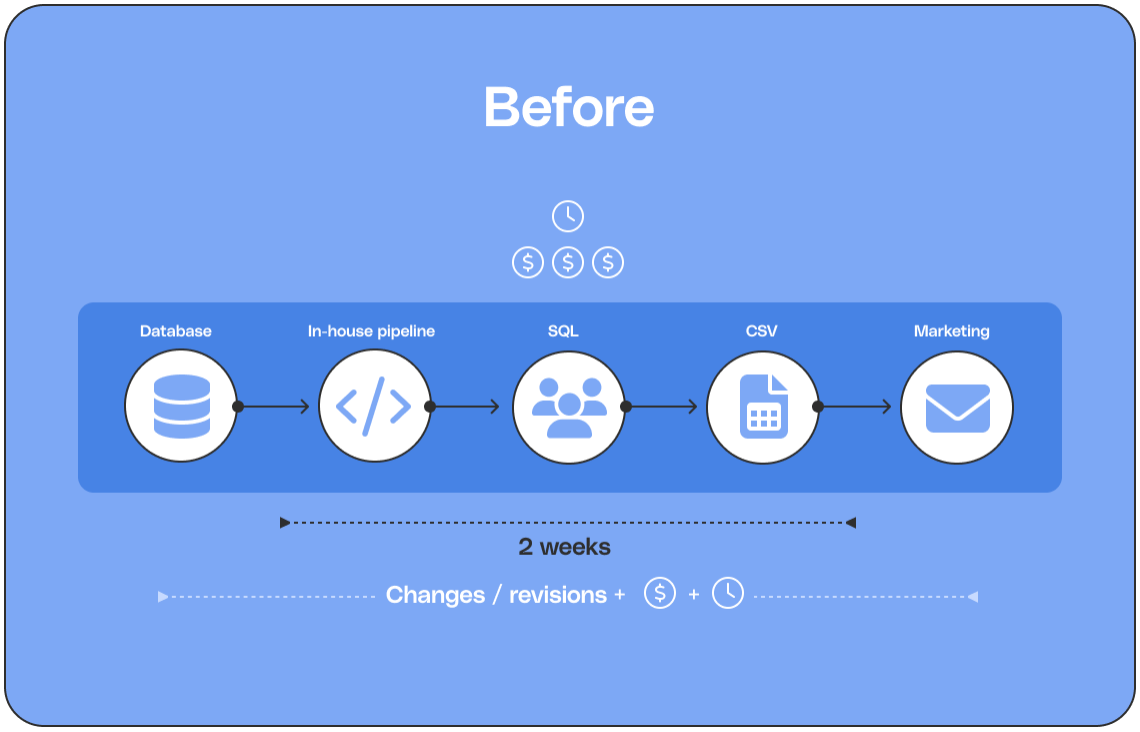
Here's a specific example from one of our customers (an international furniture retailer):

**The problem**

The data team spent a significant amount of time responding to marketing requests for adding new dimensions to user profiles for email campaigns, which required:

- Collecting data that was fragmented across the stack
- Working with analysts to join data and produce the desired data points
- Managing and running custom-built pipelines to get that data into the marketing team's email automation tool

Because of time delays, the marketing team would often rely on CSV exports, which they cleaned and loaded directly into their tool, which introduced constant challenges related to data discrepancies and file size limitations.

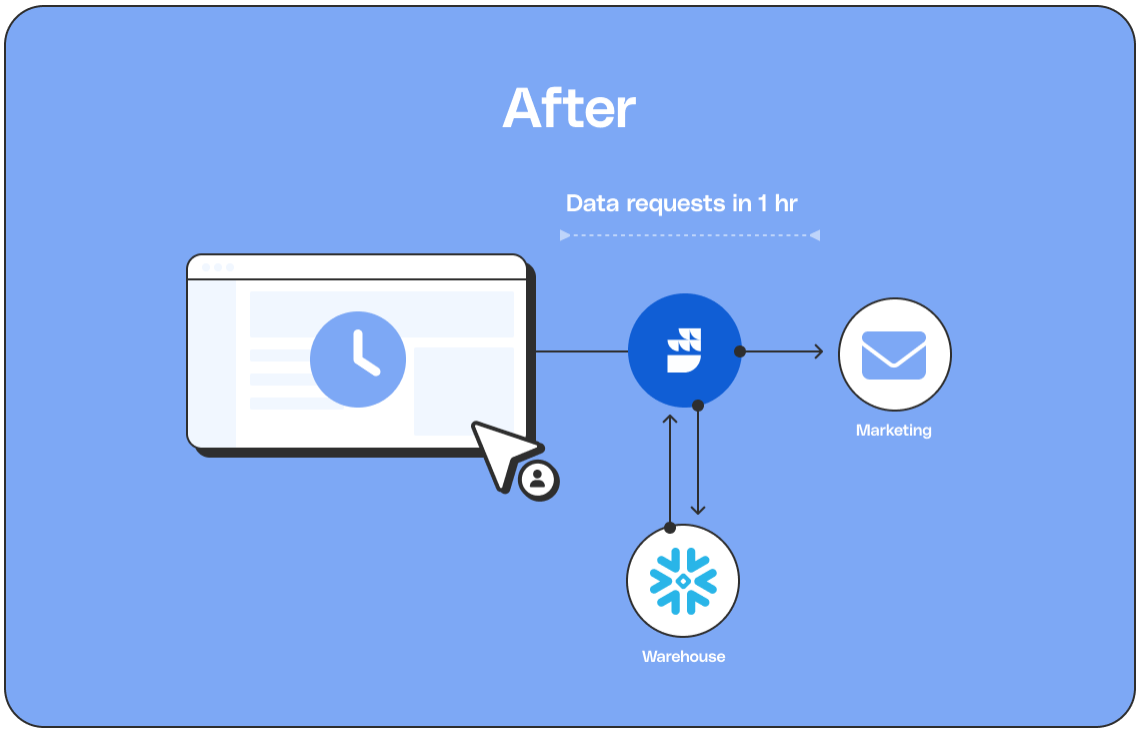



Old architecture:

- Brittle in-house pipelines
- Complex queries to join basic data
- 1-2 weeks to get new data points into marketing's platform
- Manual, error-prone data cleaning in spreadsheets

**The solution:**

This particular data team implemented a data integration layer using RudderStack's out-of-the-box integrations, eliminating all of their custom in-house solutions and manual data wrangling. The data team is now focused on helping marketing implement use cases key to thriving in 2023.




New architecture:

- Data is sent directly to the email automation tool
- Data is sent directly to the data warehouse
- Reverse ETL enables fulfillment of new data requests in 1 hour

## Money is money: replace legacy CDPs with data infrastructure

10 years ago, legacy CDPs that stored your data made sense because data warehouses were still relatively new and cost-prohibitive for many companies.

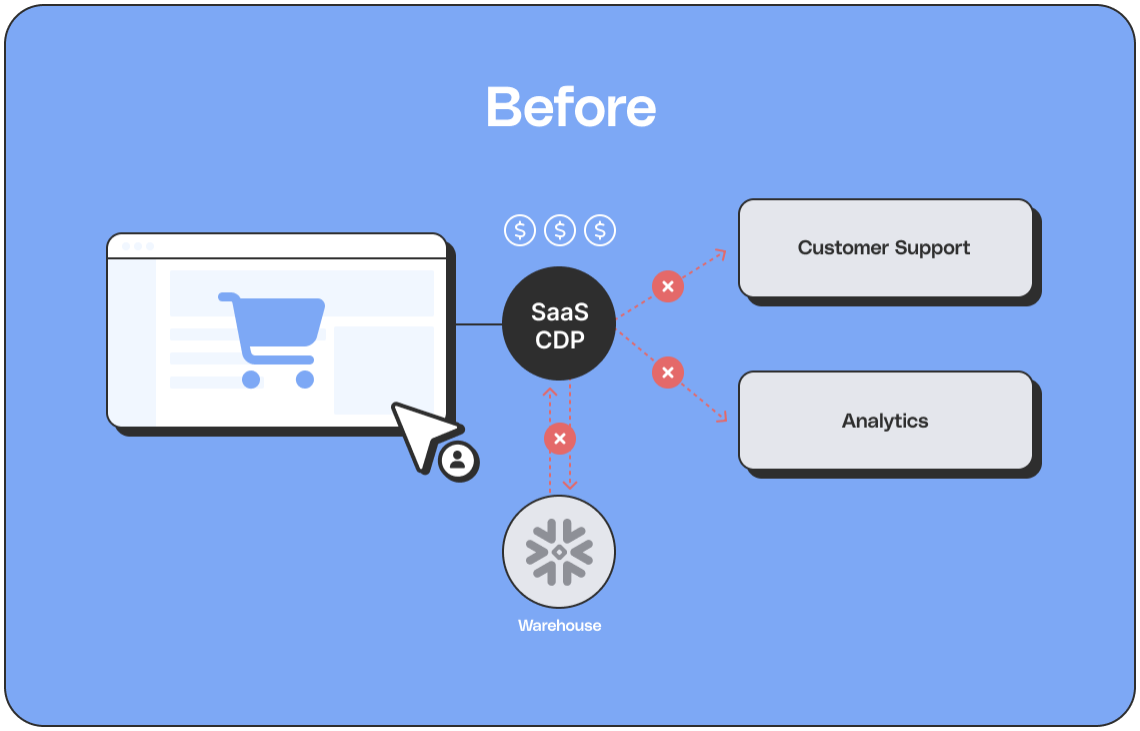
Today, the cloud data warehouse has become the most scalable, efficient, and secure place to centralize your data and ship complex projects like ID resolution and multi-channel attribution—all without creating another silo with an expensive CDP.

Let's take a look at an example:

A national shoe brand wanted to grow their eCommerce business with improved customer experiences, and knew that ID resolution was a crucial step to personalization and other use cases.

**The problem:**

Their legacy CDP vendor was costly, inflexible, and created a separate data set that kept teams siloed.



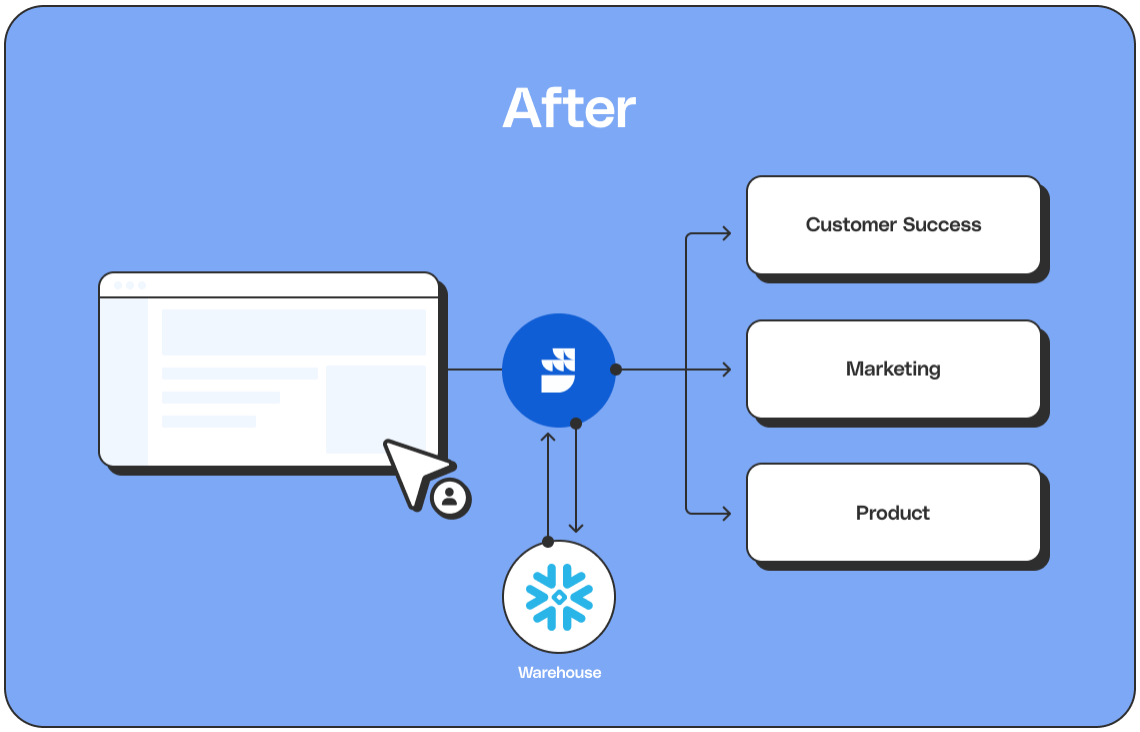

Old architecture:

- No ID resolution
- Limited use cases and data integrations
- Limitations in syncing internal data into vendor platform
- no Reverse ETL
- Trigger mechanisms mostly based on order data

**The solution:**

They replaced their legacy CDP with their own data warehouse (Snowflake), and RudderStack's Event Stream and Reverse ETL pipelines, giving them far more control, scalability, and flexibility at a fraction of the price.

Plus, they improved security and compliance because the legacy CDP no longer kept a copy of their data.




New architecture:

- Warehouse is the single source of truth
- ID resolution with all warehouse data and compute power
- Pricing based on usage of owned infrastructure, low to no maintenance cost
- Full ownership of data - secure and compliant
- Advanced features for data sharing, APIs and code-based transformations

By replacing their expensive legacy CDP vendor, the data team was able to extend their Snowflake investment and help the business deepen customer relationships with consistent data across their entire stack, and personalized experiences across the customer journey.

## Flexibility is money: move faster with adaptable infrastructure

In 2023, taking advantage of opportunities quickly will be key to staying ahead of the competition. For data teams, that means fueling ideas and initiatives with data and helping teams get them live as quickly as possible.

New projects often create new requirements, which often means either some sort of custom engineering work or buying a new tool to do the job, both of which will be extremely hard to justify in the current environment.

We've seen our customers use RudderStack's flexible infrastructure to build creative solutions that help other teams move fast while avoiding unnecessary costs.

Here are two examples.

## Using Transformations to enrich data payloads with 3rd-party geography and account information

There are entire companies focused on data enrichment and their services can be pricey, especially as add-ons for CRMs and marketing tools.

There are plenty of powerful, cost-effective APIs out there, though, and our customers use RudderStack Transformations to hit those APIs and enrich event payloads in the pipeline before they are delivered to downstream tools.

**Geolocation based on IP**

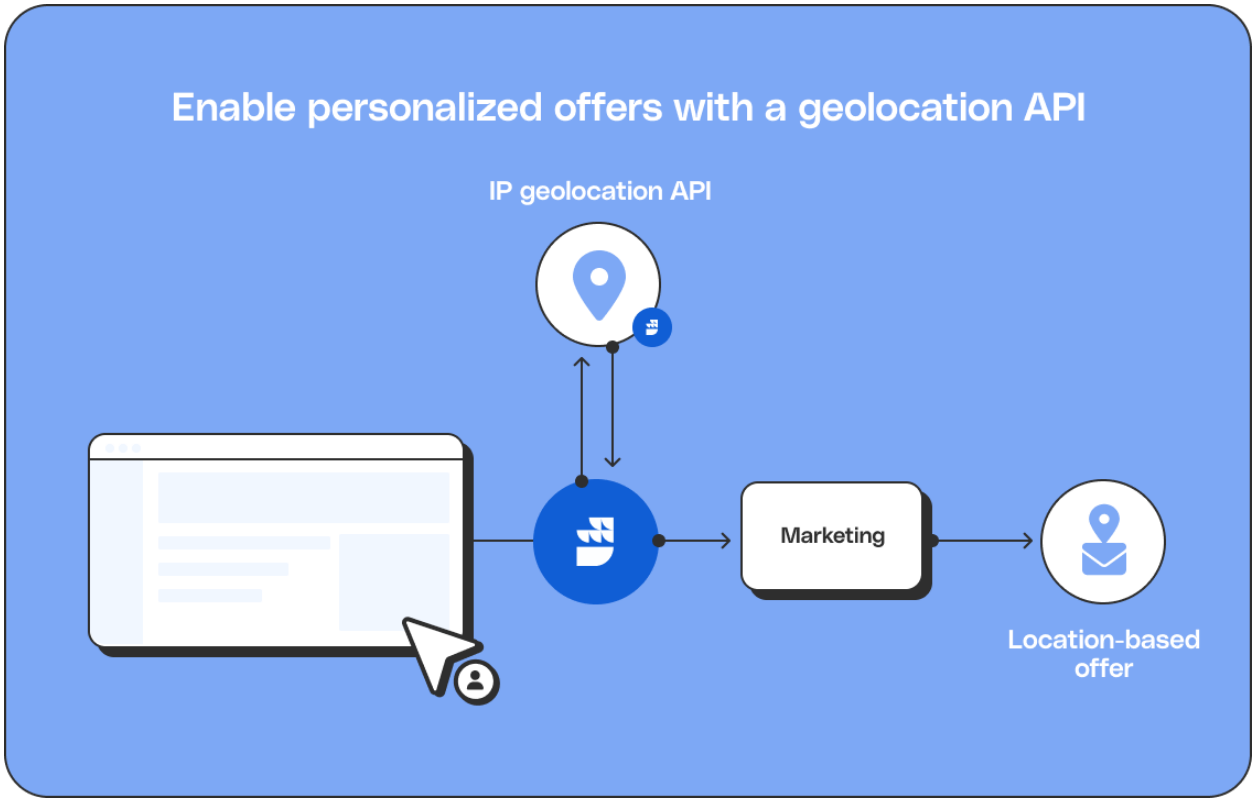
*Problem*

A customer marketing team figured out that delivering localized offers by region significantly increased conversion rate, but they didn't have addresses for many of their users.

*Solution*

Their data team leveraged a cost-effective geolocation API to solve the problem. In a RudderStack Transformation, they passed the user's IP address to the service and appended the returned region to the payload, which was passed into a custom field in the marketing platform.

The marketing team was then able to automatically segment users into regional lists and trigger location-based offers in real-time.




*Bonus: the data team was also able to scrub the actual IP address from the payload before delivery to the marketing tool to maintain PII compliance.*

## Using Webhooks to hydrate internal APIs and tooling

*Problem*

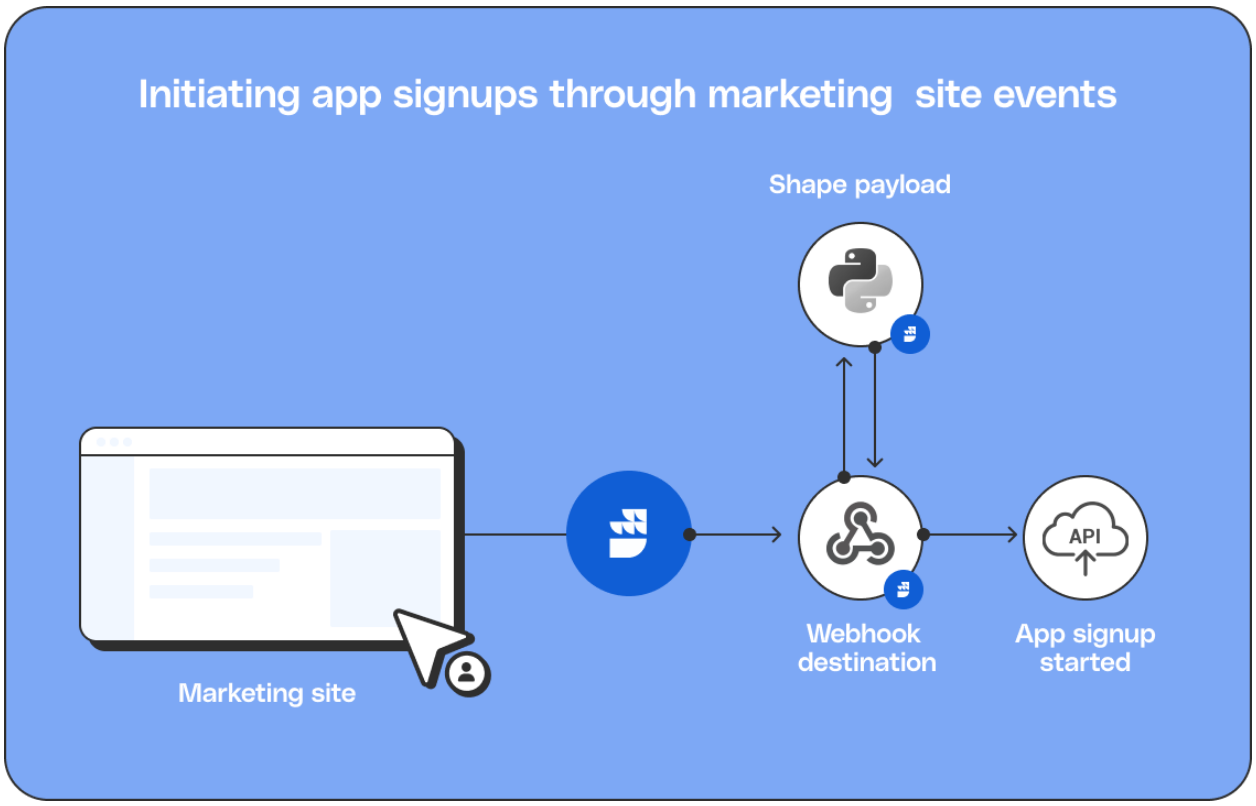
A customer product team was experimenting with various sign-up experiences and learned through A/B testing that there were certain pages on the marketing site and in their documentation that were likely to convert very high.

The problem for the product team was engineering resources: there were higher priorities than building a custom signup flow into the marketing site itself.

*Solution*

The product team brought this problem to the data engineering team, who had recently implemented RudderStack. The data team also knew that the app exposed an API for account creation.

Their solution was to use RudderStack's Webhook destination combined with a Transformation to initiate the signup process without requiring any work from the engineering team, which was a huge win for everyone.




Here's how it works:

- Payload comes into RudderStack via the marketing site source
- The payload hits a Transformation to shape it correctly for the internal signup API (this team uses Python, but you can also write Transformations in JavaScript)
- The Webhook destination forwards the payload to the signup endpoint
- The signup process is initiated and the user gets an email from the app

---

Ready to make your data stack more efficient and ship more powerful use cases? Reach out to our team to get a demo of RudderStack.

RudderStack is the warehouse-first, customer data platform built for developers. We take a new approach to building and operating your customer data infrastructure, making it easy to collect, unify, transform, and store customer data as well as securely route it to a wide range of marketing, analytics, sales, and product tools. Over 18,000 sites and apps run RudderStack including Crate & Barrel, Acorns, Hinge, Stripe, Allbirds, and more.

www.rudderstack.com

@rudderstack