

CASE STUDY

How Acorns Created a Full View of the Customer with RudderStack and Databricks



Highlights

- Acorns adopted RudderStack as its CDP to capture tracked screen, container viewed, and custom events to capture and sync events from iOS, Android, and the web to multiple destinations.
- Acorns and RudderStack worked together to co-create a custom dbt connector, eliminating the bottleneck of routing data through Redshift.
- Acorns uses dbt's simplified seeds with RudderStack, reducing funnel analytics runtimes and code complexity and simplifying event and funnel mapping models.
- Acorns adopted RudderStack's ID Stitching module to create unified views of their customers across different platforms and front- and back-end events.

Key Stats

 Acorns' Data Analysts reduced query response times from 20 minutes to 30 seconds and funnel test runtimes from 3 hours to 15 minutes

Overview

Founded in 2012 and launched two years later, Acorns is a micro-investing/ robo-advisor platform that helps young people put aside money and build a financial portfolio. Users link their credit or debit cards to the app, and Acorns rounds every purchase to the nearest dollar, automatically investing the difference in diversified Exchange-Traded Funds (ETFs) comprising thousands of stocks and bonds. Acorns customers can also manually deposit funds or set recurring payments at the interval of their choice.

In 2020, Acorns partnered with ZipRecruiter to integrate job listings into its app. It offers online banking and a rewards program that earns investment funds for online and brick-and-mortar purchases. Finally, the company's educational materials teach young people financial literacy to help them manage their money and better understand how investments and investment markets work. Acorns has over 10 million users who have invested more than \$15 billion through its platform, and the company secured \$507 million in its most recent funding round in March 2022.

Challenge: Increase control over funnels and events by reducing runtimes and code complexity

Acorns is a subscription service. Customers pay \$3 a month for the personal investing and banking plan, including retirement savings, or \$5 for the plus plan that includes savings accounts for their children. To better understand what happened after people created an account, the company's analytics team needed insight into the user onboarding experience and conversions at every step of the funnel."

"Our data analysts asked our software team for help with the registration funnel," explains Acorns Staff Software Engineer Faizan Shabbir. "We monetize our product through subscriptions. Our analytics team needed to understand how and why users were creating accounts, whether they were subscribing, why they preferred the upper or lower tier, and how much money they invested. We also wanted to know whether changes we made to the funnel were causing users to drop off before signing up for an account. We strive to figure out what works, double down on it, and eliminate anything that doesn't. The more insight we can provide our analysts, the better."

The analysts' request presented several challenges for Acorns' software engineering team. The first was cross-browser and cross-device movement. Users often started their journeys on the company's website before moving to the iOS or Android app. Cohorting was also an issue, and the analysts needed a way to define which events to track and which ones to leave behind. Finally, event data resided in different data stores, and Acorns' existing solution had difficulty routing and processing it. "The best way to solve complex data analytics problems is to simplify your data pipelines. Acorns' data analysts and engineers were hobbled by our data architecture. We struggled to route data from our website and Android and iOS apps to front- and back-end destinations. We lacked visibility into our funnels, and SQL queries took forever. We used RudderStack and dbt to streamline our workflows and simplify event and funnel mapping models, resulting in deeper insights faster and increased collaboration between our analysts and engineers."

"The result was increased runtime and unnecessary code complexity," says Shabbir. "We had very long notebooks with SQL, Python, and Scala in them, and you had to be Keanu Reeves in The Matrix to figure out what was going on. It was hard to discern upstream tendencies, and the events and repeated logic that powered funnels were tightly coupled and hard-coded."

Shabbir and his team wanted a solution that gave Acorns' data analysts and engineers more control over funnels and events. They set out the following priorities:

- 1. A single interface that allowed data analysts to transparently query different data stores
- 2. Modularized code and data
- 3. A tool to find dependencies and downstream effects
- 4. A way for analysts to create new events and funnels and update funnel event ordering
- 5. A dynamic segment funnel analysis tool.

"Our analysts were writing notebooks in Databricks and querying Redshift, AWS S3, and Delta Lake," explains Senior Software Engineer David Garay. "We had the right tools, but we had to rethink our architecture because it was far from ideal with bottlenecks."

Solution: Reign in Complexity and Rethink Data Pipelines with RudderStack and dbt

Shabbir and his team made two significant changes to optimize Acorns' architecture. First, they dropped Databricks built-in notebooks and adopted dbt, a SQL-first transformation workflow tool that allows data analysts and engineers to collaborate on analytics code that is portable and modular. Second, they chose RudderStack as their customer data platform, allowing the company to capture and track event stream data across multiple platforms—including the web, Android, and iOS—and route it to multiple destinations.

"I looked at our previous architecture and stripped away everything we didn't need," says Shabbir. "When I saw what was left, I realized I'd removed all the complexity and added incomprehensibility. We had the data and the tools to analyze it, but we had trouble ingesting it, routing it downstream, and authoring queries we could easily share and modify. RudderStack and dbt allow us to log and clean event data, route it to our S3 and Delta Lake stores, and move it in and out of Databricks. Our analyst and engineers now work together to make sense of it in Databricks and visualize it in Tableau with fewer details and delays."

Acorns worked with RudderStack to architect the perfect solution through an iterative design process and landed on a design that wrote directly to Databricks. During this process, Garay and Shabbir realized that the initial version lacked partition pruning and proposed a solution to RudderStack.

"We sent RudderStack a sample query to show them how it works in Delta Lake. It's an optimization technique that limits the number of partitions inspected by SQL query," explains Garay. "Instead of looking through all the data to find a record, you specify the partition containing the needed information. If you're not specific, a MERGE INTO query can take 10 or 20 minutes to provide an answer, but if you limit the range—for example, to a day or two—the answer can take 20 seconds. Since we're merging massive quantities

of event stream data in real-time, partition pruning massively accelerates the process. I truly appreciate how RudderStack's team collaborated with us to fine-tune a solution through two iterations instead of sticking to good enough."

RudderStack's Databricks connector was the final piece of the puzzle. Acorns finalized its architecture and started streaming event data from iOS, Android, and the web to its new data architecture.

Results: Enhanced Pipelines, ID Stitching, and Modular Models Drive Experimentation Leading to Meaningful Insights

Acorns' new architecture greatly simplifies the company's data pipelines, eliminating bottlenecks while consolidating and speeding up event stream data processing. Data analysts and engineers now collaborate on analytics workflows, and dbt's modular, reusable code has made confusing long notebooks a thing of the past.

"It all starts with RudderStack," explains Shabbir. "We use it to capture container viewed and custom events and tracked screens across iOS, Android, and the web. We then sync everything to multiple destinations. RudderStack provides rich metadata and flattens data from various sources into a consistent format," he adds before joking, "We then use dbt to put the fun back into our funnel."

Acorns' analysts use RudderStack's ID Stitching module to create unified views of their customers across multiple platforms and channels before routing that data to dbt for further analysis. "Our marketing team is motivated by empathy," explains Garay. "They understand how challenging it can be to open an account and set up our app, so they try to alleviate issues with our registration funnel. RudderStack associates anonymous actions on our site, our apps, and across the web with actions taken once a user creates an account and logs in. This added information helps us identify drop-off points and resolve leaky bucket issues across our funnel in dbt."

Acorns also uses RudderStack and dbt to map logged actions to semantic events, defining and segmenting funnels based on user attributes. The company consolidates front- and back-end event stream data from RudderStack using an event mapping model in dbt and feeds this single stream to a funnel mapping model. The result is a dynamically generated funnel that outputs only the metrics Acorns' analysts need to test their hypotheses.

"Defining new events and funnel mappings is easy," explains Shabbir. "You change the text in the seed file, and you're done. Because dbt automates version control, you can see past changes and revert to a previous version if you don't like the new output— or you can try something else. Our analysts are free to experiment without worrying about breaking our models.

"We also created a macro that automates funnel making," he continues. "Our analysts can call arguments to create funnels and column groups to segment them. They can test end-to-end funnels without writing a line of code in a development environment, without affecting our production funnels. It makes the developer experience much nicer. And we've reduced runtimes from three hours to 15 minutes because dbt modularizes our mapping models, and we no longer have to do joins across data stores thanks to RudderStack."

Acorns' enhanced data pipelines provide the speed and flexibility to dig deeper into event stream data, user behaviors, and the company's funnels to extract faster insights. "One of our data managers said it best," concludes Shabbir. "RudderStack and dbt have made funnel reporting more flexible, efficient, and robust. We are experimenting, uncovering areas of high importance and impact, and finding new opportunities more easily."