What does Segment do?

An explanation of Segment's technology and business

Justin ✔
Oct 12, 2020

♡ 36   💬 4   ⤴   🔖

The TL;DR

Segment is a **routing platform for your customer data** (what they call a "customer data platform"). Engineering and data teams use Segment to:

- **Fire events** whenever their customers take actions in the product (e.g. page viewed)
- **Move data** between standard sources, like Mixpanel to Salesforce
- **Manage customer profiles** via a single "persona" and **transform data** on the fly

If these bullet points seem to have nothing to do with each other, you're onto something - Segment is really hard to explain because it does a lot of things all together, and until you understand all of them separately, it won't make sense. And that's why, in general, very few non-developers do actually understand what Segment does.

Trials and tribulations of managing customer data

Companies revolve around data today (duh), and most of that data centers on **the customer**. Who are they? How are they interacting with the product? How did we acquire them and how much did it cost? In fact, arguably the entire point of data science and analytics is just to understand your customers. But like anything in the data space, it's hard and annoying to do!

1. **Generating standardized data patterns is hard**

Standard practice is to **fire an event** any time a customer does an action of interest on your site or product. A few examples:

- Visiting your website

- Signing up

- Adding more users to their account

- Upgrading their billing plan

These events all take place in your product, so you need to write custom code that takes the information (what just happened), wraps it into a data structure, and sends it to where it needs to go (e.g. Salesforce, your data warehouse, etc.).

One thing that makes this really difficult is **associating anonymous sessions with authenticated users**. When someone visits your site for the first time, you have no idea who they are - if they eventually sign up and give you their information, you need to *tie that back* to their pre-signup activity, so you can understand how you acquire your customers (e.g. which pages on your site convert best).

2. **Customer data comes from many sources**

Your app and website aren't the only places that generate customer data, i.e. data about your customer that's useful to your team. Stripe is generating your billing data, Salesforce is creating data of record for how your customers interact with the Sales team, and Google Ads is recording how your search ads perform.

For this data to be useful, it needs to be (a) centralized via [ETL](#) into a [data warehouse](#), and (b) pushed to other places where your team needs to use that data, both operationally (I need it to make this sale!) and analytically (I need it to answer this question!). Hence, problem number 3.

3. **Customer data needs to go to many sources**

Analytics teams need customer data to sit in a centralized warehouse where they can write really fast queries against it and answer annoying questions that the CEO asks. But they're not the only team that needs this data - it's also operational!

- **The Sales team** needs to know how often their leads are using the product (destination: [Salesforce](#))

- **The Marketing team** needs to target email campaigns based on which source they acquired users from (destination: [Marketo](#) / [Hubspot](#) / [Customer.io](#))

- **The Success team** needs to proactively monitor top accounts and see if their product usage is dwindling (destination: [Vitally](#))

Manually sending these events yourself and connecting these data sources together is a nightmare and a half. You need to schedule jobs to run regularly, maintain them as they (inevitably) break, each destination has a custom schema (e.g. "email" in Marketo vs. "customer_email" in Salesforce), etc. You don't want to be doing this. You will quit.

These are the 3 fundamental challenges to managing and getting value out of customer data. And they're the three things that Segment does really well.

The Segment product: a few things

Segment's purpose as a product (and how they market themselves to enterprise buyers) is as a **Customer Data Platform** (CDP). That might sound hand wavy, but now that you understand all of the problems engineering teams face managing their customer data, it starts to make sense - Segment provides a product that solves those problems, in two basic categories:

1. **A tracking library**

Segment's first "product" is aimed at solving the first problem - how difficult it is to fire events from your product and manage the identities of your customers (and your not-yet-customers). The package is called [Analytics.js](#), and it's a library you can use in your application to fire standardized, clean events that go straight into Segment's system (it's also available in other languages).

Segment has a rigid **philosophy** for how customer data should be structured, and they apply that in Analytics.js. Any action that a customer gets classified as one of [five different types](#), and you need to decide which you want to use for each event you fire. If you hear "identify calls" or "track calls" chances are someone is talking about Segment.

One of the killer features of Analytics.js is that Segment takes care of managing your anonymous traffic (i.e. site visitors who haven't signed up for your product yet). Segment handles the process of allocating IDs to them, and once they sign up, associating those previous "anonymous sessions" with the now authenticated user.

> 🔍 **Deeper Look** 🔍
>
> To get around ad blockers, some companies will create their own tracking libraries and *then* send *those* events to Segment. In that case, you need to build tracking from scratch, but you're still getting value out of Segment's internal mechanics for identify management (those ugly anonymous users from before) and their routing platform.
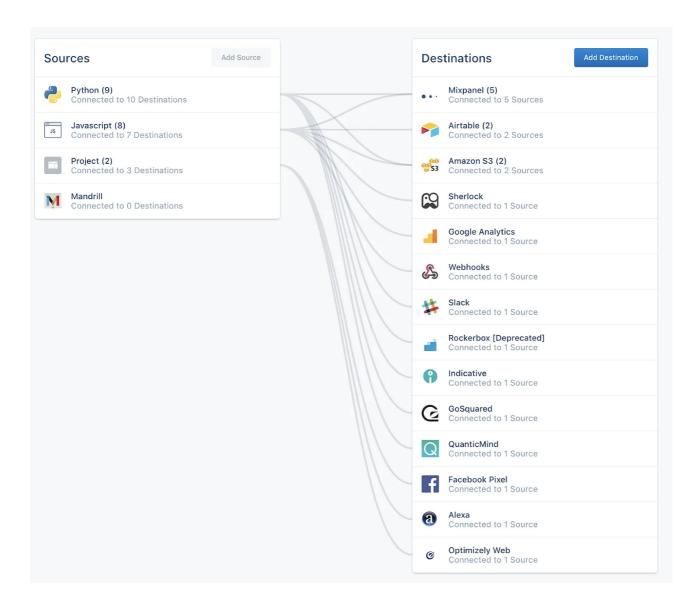>
> 🔍 **Deeper Look** 🔍

The real interesting piece of Segment, though, is the routing platform, and these two things combined pack a powerful punch.
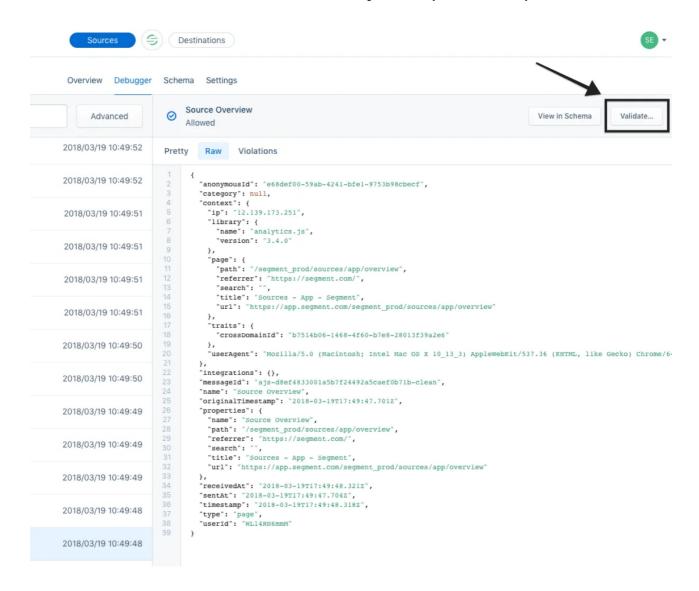
2. **A routing system**

Segment provides a service that lets you sync data between sources and destinations without writing any code. If you're generating events via Analytics.js, Segment ingests them and lets you send them to any source you want - Salesforce, Marketo, S3, whatever - on a regular schedule. They also provide really nice tooling for debugging when things go wrong, checking what kinds of events are getting through, and managing what exact data you want to sync and what you don't.

Your Segment dashboard shows the sources (data you're creating or ingesting) and what destinations you're sending them to:

You can click into any source or destination to get more details about how things are connected and what events are making it through:

I've used Segment at all 3 startups that I've worked at, and it's a *fantastic* product. I'm honestly not sure how we'd be able to keep things afloat without it - it probably saves time equivalent to an entire data engineering team.

New lines, pricing, and the end

→ New product lines

Over the past couple of years, Segment has started releasing new products on top of the core two above. I haven't used them, but here they are:

1. **Warehouses** automatically provision and manage a data warehouse for you, and syncs your customer data directly to it

2. **Personas**let you pull together all of your data for a single customer in Segment to get a centralized view / profile that you can export to other tools

3. [Functions](#)let you execute code on the fly as your data is moving between sources, so you can clean and prepare it without needing to build separate infrastructure

These make a lot of sense when you frame Segment as *the* place to manage your customer data - anything that helps you do that is in play, regardless of which part of the stack it sits in or which team might need to use it. But generally, they're in the early stages of adoption.

→ **Pricing**

Products that are valuable tend to be expensive, and Segment is no exception (not by a long shot). [They have 3 basic plans](#): free, team, and business. The general idea is that you pay **per MTU, or monthly tracked user**. Generally, you're going to pay $10-$12 per month per 1K tracked users, so if you've got 50K people visiting your site per month, you'll be paying at least $500/mo.
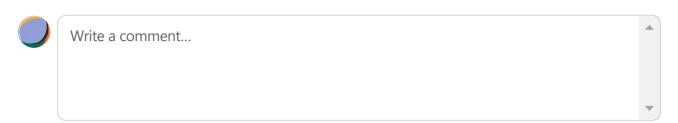
→ **The end**

Segment [got acquired this weekend](#) by [Twilio](#) (market cap: $45B) for around $3B. This was a ~2x markup over Segment's last valuation, but still confusingly low - general sentiment around Segment is pretty good, and a lot of people (myself included) expected a $10B+ outcome. Either way, this acquisition seems to make sense on paper, as Twilio is all about **contacting your customers**, and Segment is all about managing your data about them.

Further reading

- You know who does a good job explaining Segment? [Segment](#)!
- There are other tools that focus on moving data between standardized sources, like [Fivetran](#) (who, funny enough, has [a Segment integration](#))

4 Comments

Write a comment...

Brendan Keeler Writes Health API Guy   Jan 17, 2021

Great write-up. For that last comment, my understanding is Fivetran is more focused/specialized on ELT/ETL - getting data to a storage solution like a warehouse. Segment does a bit of that, but is much more focused/better at on real-time streaming and unifying into user profiles for marketing teams. Hence the integration between the two (allowing them to stay in their own lane and have mutual customers)

♡ 1   Reply   Gift a subscription   Collapse   •••

Feno Oct 14, 2021

One question, how does a CDP like Segment differ from what Mulesoft does? Are they solving the same problem but in different ways? Any color on this would be appreciated. Sent the same question on twitter DM but thought it is better to share your knowledge publicly so I post it here also :)

♡   Reply   Gift a subscription   Collapse   •••

2 replies

2 more comments...

© 2023 Justin · [Privacy](#) · [Terms](#) · [Collection notice](#)
[Substack](#) is the home for great writing