What's Reverse ETL?

Getting your data OUT of your warehouse?

Justin
Feb 1

The TL;DR

Reverse ETL is the process of syncing data **from your data warehouse to your business tools** like Salesforce and Hubspot.

- Data teams spend their time organizing data into a **centralized warehouse**
- It's operational teams, though, like sales and marketing, who **need to use that data in their day to day SaaS tools**
- Reverse ETL is the process of **moving data from the warehouse into those tools** so operational teams can take action on it
- It's called reverse ETL because the data is ***leaving* the warehouse**

As the so-called "modern data stack" has consolidated around centralized cloud warehouses like Snowflake and BigQuery, the process of reverse ETL-ing that data into SaaS tools has gotten easier, and tools like Census offer it as a simple SaaS service.

What kind of data is in the warehouse?

One of the confusing pieces of reverse ETL is understanding what, exactly, is *in* the data warehouse and where it comes from. For a refresher on what data warehouses are, check out the original Technically post here. Every warehouse setup is different, but you should expect to see some combination of:

- **User data**

Every app has a `users` table in their production database, and that usually forms the foundation for the kind of data you'll have about your users in the warehouse. Common attributes include when the user signed up, their email address, their name, the company they work at, and sometimes high level activity data like the last time they signed in.

- **Product usage data**

What users are doing in your tool is some of the most valuable analytics data you can capture. Teams tend to track this in terms of **events**. The code powering your app records an "event" every time a user clicks on a specific button, navigates to a particular page, or does anything of note in the product; those events then get sent to the data warehouse for future analysis. You can roll up this data to answer questions like what percentage of your users tried a specific feature.

- **Marketing and attribution data**

In the same vein as product events, data teams collect information about which website pages users visit, what source referred them to the site (a podcast sponsorship? Google Search?), and the "journey" they took across marketing materials before signing up. This helps marketing teams understand which channels are performing well and where to allocate spend.

- **Payment and billing data**

Data on which plans users and teams are on – as well as how much they're paying on a monthly basis – helps teams understand where revenue is coming from and what activity it might be tied to. Some of this comes from your own systems, and some comes from payment APIs like Stripe.

No matter what kind of data we're talking about, all of it gets *into* the warehouse via ETL (or ELT), a process that Technically has covered extensively here. Data teams spend their time taking source data, modeling it to make it useful for analytics, and depositing it in the data warehouse. Getting data into the

warehouse in a useful format is one of the key responsibilities of a data team!

But **reverse ETL** is the opposite: getting it *out* of the warehouse and *into* your team's SaaS tools.

Data needs to be in context

Companies run operationally on some hacked together combination of SaaS tools. Each team uses different tools, but the theme is the same: they all need user and activity data to be useful. Let's start with a few examples:

→ Sales

Most sales teams use Salesforce as their system of record and central hub for operations. Salesforce tracks all of your open opportunities and accounts, lets you move companies between stages (e.g. prospect → customer), automatically pulls in email data, allows teams to build pipeline reports...there's a lot you can do in there.

→ Marketing

Marketing teams use a CRM, too: usually Hubspot or Marketo. These tools allow teams to track all of their leads, organize them into groups, and send targeted email campaigns at specific times based on triggers. If you've signed up for a product and gotten a welcome email a day later, that was probably sent using one of these tools.

→ Support

Support teams use tools like Intercom and Zendesk to track all open tickets and requests, as well as communicate with customers through email or chat. These tools let teams triage and organize open tickets into statuses, collaborate with other team members (e.g. tag a coworker who might know the answer), and send CSAT surveys among other things.

Even though all of these tools generate and manage their own data, they become a lot more useful when you can pull in external data about your users from the data warehouse. Back to our examples:

- **Sales** – product usage data
    - E.g. how much time has this user spent in the product? What features have they used? How many team members are active?
    - Helps qualify prospects, allocate time more effectively
- **Marketing** – demographic data
    - E.g. how many employees work at this company? When they signed up, what did they say they were going to use the product for?
    - Helps build useful groups and target messaging / drip campaigns
- **Support** – payment data
    - E.g. what plan is this customer on? How long have they been paying? How large is the account?
    - Helps prioritize support for more mission critical accounts

These are illustrative: teams get creative and you never know what kinds of data are going to be valuable in different contexts. The important part is that the data needs to be *in* the tool that the team uses to take **action**.

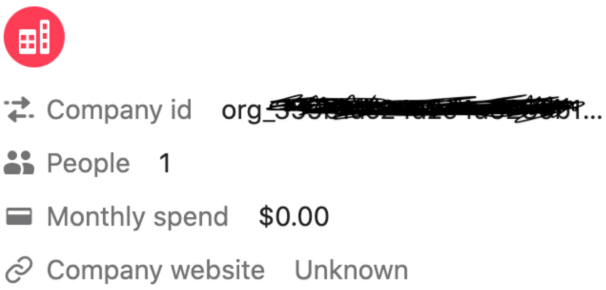
How Reverse ETL works

Now you know what kind of data is usually in a warehouse, and that it's important to be able to get that data out of the warehouse and into your team's operational SaaS tools. **Reverse ETL** is the process of making that happen. Another way of thinking about it is that Reverse ETL lets you operationalize your data, which is just a fancy way of saying you can actually use your data for day-to-day actions and decisions vs. sending it into a dashboard that'll get stale in a couple weeks.

→ SaaS tools have databases you can use

All of the aforementioned tools – Salesforce, Hubspot, etc. – allow you to send data to them that they'll store for you and surface in the tool's UI. They usually expose this functionality via a set of APIs – like this one from Salesforce. Once you get data in, you'll see it in context when you're using the tool. Here's an

example from Intercom that you'd see next to a support conversation with a user:



Whoever set this up is pulling data on the company's internal ID, their team size, their monthly spend (they're not spending anything!), and their website, which is also unknown. They used Intercom's API to get this data into Intercom's systems, and now they can use it side by side with the tool. The data you're seeing here got to where it is via Reverse ETL.

→ Wave 1: build reverse ETL from scratch

Until recently, state of the art was to do reverse ETL manually. To logistically get data from your warehouse into SaaS tools, you need to set up some sort of recurring job that pulls data out and uses the SaaS tool's APIs to get it into their systems. The APIs will require that data fits a specific format, so you'll likely need to transform it in flight. You'd also need to set up reporting and a retry scheme for when errors happen, because errors will happen.

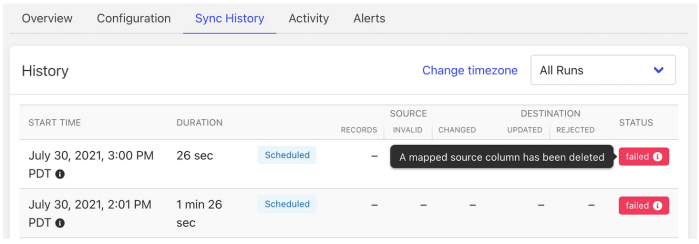
> 🔍 **Deeper Look** 🔍
>
> The process of building reverse ETL pipelines from scratch actually very closely mirrors building normal ETL pipelines from scratch to get data *into* your warehouse. The doubly annoying thing here is that each destination – i.e. the APIs of your SaaS tools – require data in slightly different formats and configurations, which means each pipeline needs to be customized.
>
> 🔍 **Deeper Look** 🔍

In case my phrasing didn't make it obvious, this is kind of a pain in the ass, especially since APIs and formats change over time. A few years ago, purpose built tools started popping up to make things easier.

→ Wave 2: purpose built SaaS for reverse ETL

Today, you can use something like Census or Hightouch to do reverse ETL for you pretty easily – you just need to do a little up front configuration. These tools connect to your **sources** (data warehouse and others), your **destinations** (Salesforce, Hubspot, etc.) and let you easily move data between them using SQL. You can also create models that organize your data and prep it for the format your tools want to intake it in.

A common use case might be to sync your marketing attribution data from your warehouse to Hubspot once a day. Census will let you schedule that sync, transform the data if you need to in advance, and show you a history of how each sync went:



Behind the scenes, they're taking care of running queries against your warehouse, transforming that data on the fly, and hitting the APIs of your SaaS tools to regularly insert and update the data you want.

Further reading

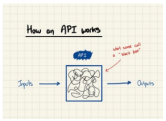
- Census's explainer on Reverse ETL
- A roundup of Reverse ETL tools

- A roundup of reverse ETL tools
- Benn Stancil's post about the warehouse as a platform

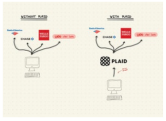
♡ 19 Comment Share

Write a comment...

Top New Community

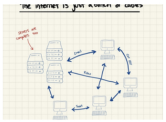


What's an API?
What McDonalds and Lyft have in common
Justin Jan 9, 2020 ♡ 187

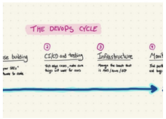


What does Plaid do?
Technically begrudgingly tackles Fintech
Justin Jan 14, 2021 ♡ 34 3

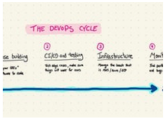
What does New Relic do?
Keeping an eye on your servers and apps
Justin Jan 11 ♡ 23 2



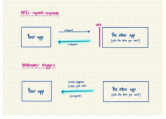
What happened to Facebook?
A basic explainer of what that outage was all about
Justin Oct 5, 2021 ♡ 29 4



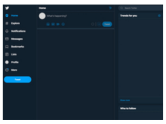
What does GitLab do?
The TL;DR GitLab is a somewhat contrarian take on DevOps: it's basically one giant tool for literally anything you'd want to do relating to building and...
Justin Jan 4 ♡ 8



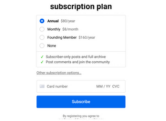
What's DevOps?
IT has a cool new name
Justin Jan 5, 2021 ♡ 34



What are webhooks?
Triggered
Justin Sep 13, 2021 ♡ 28 5



What's Headless E-Commerce?
We may be running out of names
Justin Nov 2, 2021 ♡ 20 7



I built a (basic) Substack clone in a month
This was probably a waste of my time
Justin Sep 2, 2020 ♡ 50 4

What's Kafka and what does Confluent do?
Help with solving Kafka-esque data problems
Justin Jun 15, 2021 ♡ 29 4

See all ›