GUIDE TO THE MODERN DATA STACK:

OPERATIONAL ANALYTICS AND REVERSE ETL

Census

SNOWPLOW

Table of Contents

CHAPTER 1

A NERVOUS SYSTEM OF DATA SITS AT THE CENTER OF THE MODERN COMPANY

We often liken the data warehouse --powered by cutting-edge modeling tools like dbt-- to the brain of your organization. It's your knowledge and decision-making hub, capable of running everything from the day-to-day vitals of your company to the high-cognition, blue-sky goals that unite all your teams.

To quote one of my favorite brain and thinking experts, Daniel Kahneman, author of Thinking, Fast and Slow: "The world makes much less sense than you think. The coherence comes mostly from the way your mind works."

Your data is the same way. The value isn't in the numbers themselves, but the way you can pull a story from them about the world of your product.

But, if you're neglecting the other parts of your data nervous system, all that work your data brain is doing is for nothing. And the only way your teams, and the tools they use, can reach it is through a convoluted, manual set of strings that play telephone with your source of truth.

This, as you can probably gather, isn't great. Processing between the body parts of your organization is slow, and by the time the information gets all the way out to the limbs, the data is probably stale. Even if the data going into your brain is great, your business can become dumb and laggy (no offense).

Companies--like living, breathing organisms--can only survive if they have consistent, accurate data to act on and a system that relays data to the correct endpoints.

Just like how you can't navigate a tricky line mountain biking with bad eyesight, your company's marketing won't make sense without a clear picture of your ICPs. Or, just as you'll singe yourself on a hot surface without real-time tactile feedback, your sales reps can burn your customer relationships without real-time contact records.

The data nervous system empowers companies to interpret information from the world of their customers and respond accordingly.

Why we liken your modern data stack to the nervous system

We get it, metaphors are tricky business and often pretty iffy because they fail to represent the true complexity of the subject they're describing. But we're pretty passionate about this one, and when you peel the layers back it's easy to see how your data systems drive vital actions for your company in the same way that your biological nervous system keeps you living and breathing.

To make organic and digital systems alike work, receptor cells take signals (data) back from an environment and send them as electrical signals to the brain (warehouse + modeling) via nerve fibers (the tools within your modern data stack). If any part of this network is broken, or if the brain itself isn't functioning with complete information, we can't respond to stimuli (customers) effectively.

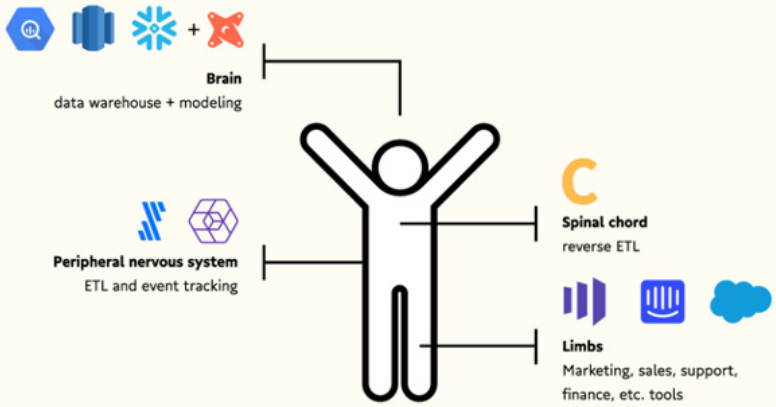


Image courtesy of Census

If we think of the modern data stack as the nervous system to our living, maturing business, we can see each layer take on a unique role:

- The **data warehouse and your data modeling tools act as the brain** of your nervous system, processing, and shaping all the data coming in and out from your body into a single source of truth.

- **Reverse ETL tools act as the spinal cord** of your organization, not just transmitting information but contextualizing it in the right format for your brain to act on.

- **ETL, event tracking, and behavioral data tools** represent the peripheral nervous system between your spinal cord and your limbs and are an integral relay system of sensory input throughout your business body.

- **Marketing, sales, and support teams (and their tools) represent the limbs** of your organization, interacting with the world and moving you forward.

Each part of the nervous system of your business builds on the data warehouse (and the modeling that helps it learn) as the brain to operationalize data. As your brain gets smarter and learns from the data fed into it, you can watch your business go from wobbly toddler to lanky teenager to sophisticated adult.

A brief history of the evolution of the business brain

We've touched on the concept of the data warehouse as the brain of your organization before, and argued that, as with the human body, the brain should sit at the center as your hub. From the hub, spokes (your applications) can connect directly to the freshest data available. This hub and spoke structure is your data nervous system.

But to understand the pivotal role of this system, we have to take a step back and recognize the brains that came before it.

Before we had the concept of the data warehouse as the brain, and the cloud computing technology that made this concept possible, companies leveraged a hive-mind of real, human knowledge to drive decisions. A single person (or a small collection of people) could act as the brain, and everyone downstream of them could execute on their knowledge.

However, as we tipped into the 21st century, companies began to rapidly grow their customer bases. The human mind couldn't hold all the data attached to millions of customers at once, so we started to outsource to our digital counterparts piece-by-piece.

We picked ad hoc tools to supplement our saturated brains, which has quickly avalanched into a mess of data stores connected point-to-point in a complex web of dependencies.

While the spirit is right--we should leverage digital means to improve our decision making--this approach effectively tied the shoelaces of our modern, digital businesses together, tripping them up at each step. If your stack looks anything like the example below, you probably know the feeling.

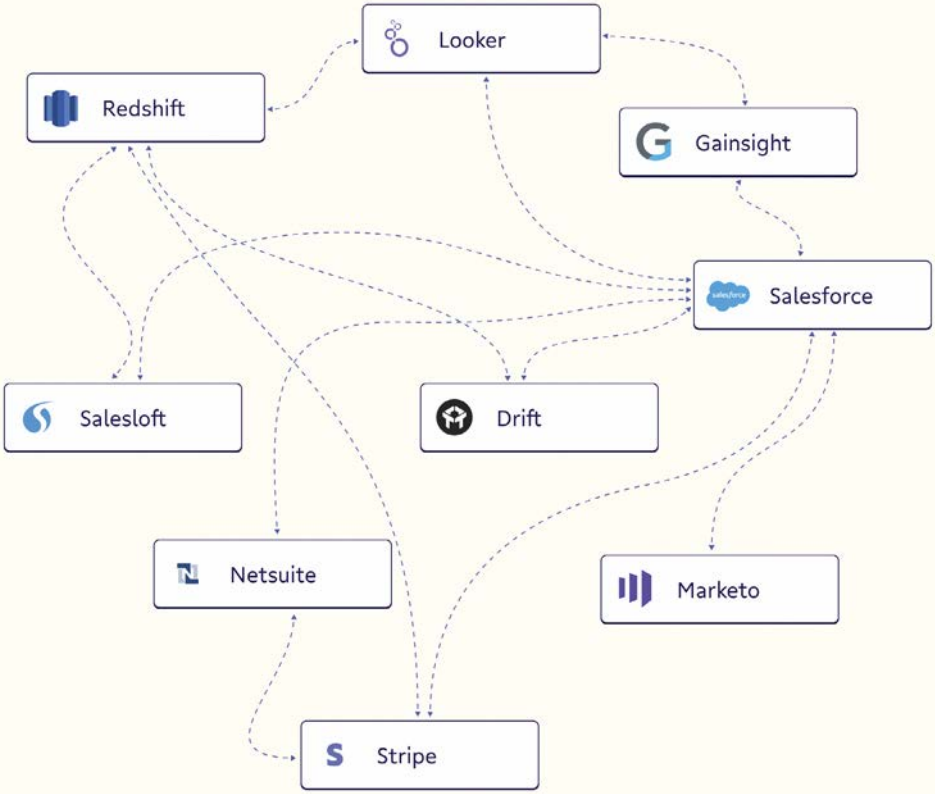
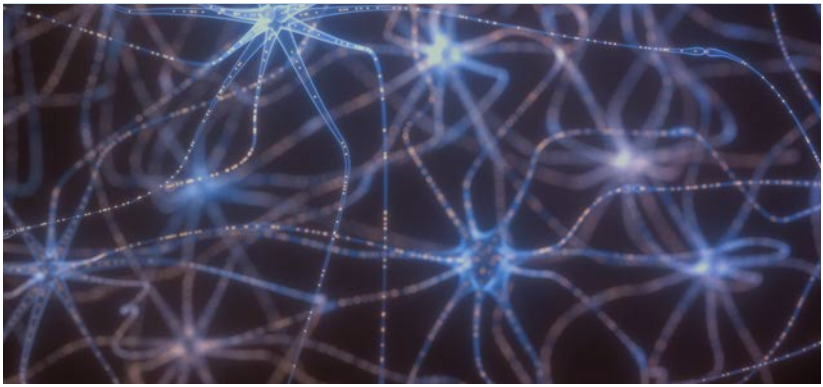


Image courtesy of Census



These point-to-point ties between different systems in the business body seem like a quick fix when the body is starting out, but quickly grow to cause problems. They don't grow with you, and each time one part of the system matures the dependent parts strain to keep up or break. The more connections--whether metaphorical strings or literal APIs and webhooks--you have between each limb, the more dependent and tangled the body becomes.

Essentially, it'd be like limiting your right hand by making it mirror the motions of your left. Sure, you can pick up some of the things you need immediately, but you'll lack the finesse to execute on anything truly impressive. In our business use case, this misfiring can look like sending the wrong emails to the wrong person at the wrong time.

Furthermore, by bypassing the brain as the hub of your nervous system, the entire system is vulnerable to data stomping, where the movement of one limb (app) can undo or override the movement of another.

Sounds like a hellish mess, doesn't it?

When you treat your business as an intelligent organism run by an evolving brain of data, your systems can grow with you as they learn the bespoke knowledge that's unique to your business (e.g. We send X email to Y users when they've completed Z action, or show them X promotion when they've done Y). Whereas a piecemeal system would equate 🔥= danger, an intelligent system with a learned brain can deduce that 🔥+🍔 is a good (even desirable) thing, not a threat, and send hunger (not flight) signals as a response. It's all about the context good data and smart processing provides. At the root of this intelligent system is a nervous system that looks less like the previous jumbled map and more like the streamlined system below:

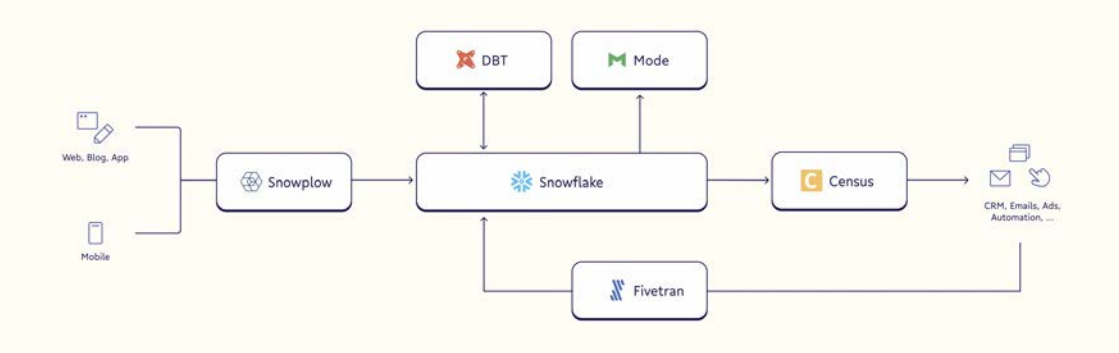


Image courtesy of Census

Now, let's look at the unique role each layer of the data nervous system plays in modern business.

Breaking down the layers of the data nervous system

Our nervous system helps us process the world around us via a series of biological ones and zeros. It lets us know when there's a critical failure or bug in our system, and quickly updates our brain with changes in information from the limbs via organic trigger syncs.

However, our receptor cells don't give us all the information raw from the world. They prioritize syncing new information and transform it along each step to meet the requirements of the next layer.

As data moves along the pipeline of our stack, it's encoded and transformed based on a set of computational rules set by each neuron (tool) and its relationship to the next connection. By the time the data reaches our brain (biological or binary) it's been transformed to fit a predetermined schema for future querying.

This knowledge can then be voluntarily (manually) or involuntarily (continuously/automatically) leveraged for each system within the business. Here's a breakdown of the role of each layer.

Warehousing and modelings: The brain of your data nervous system

In your modern data stack, the brain is made up of two parts:

1. The **data warehouse** (BigQuery, Snowflake, Redshift) acts as a clearinghouse for all your data.

2. The **modeling tools** (dbt) that contextualize all that data and allow you to learn from and use it in a bunch of different situations.

With these two layers of your stack working in tandem, each team downstream can reference a single source of truth for the most up-to-date and intelligent information. This can make the difference between your marketing team burning out a prospect that your salesperson has already talked to and your marketing team being able to effectively remarket information your warm leads need to convert.

This functioning brain enforces data consistency, syncs data across all the tools and limbs of your company, and can be easily queried when those downstream sources need assistance.

When you pair together with a competent data warehouse and modeling tool, your organization can process and learn from large amounts of data that the human hive mind wouldn't otherwise be able to get through fast enough. If your business were a human student, this pairing is what enables your business to go from solving basic addition to quickly parsing advanced calculus.

Reverse ETL:
The spinal cord of your organization

Reverse ETL tools represent a new stage of evolution in the modern data stack and the businesses that rely on it.

Prior to the emergence of reverse ETL--or operational analytics platforms, as we sometimes call it--there wasn't a great way for the well-armed data brain to get the information it learned back out to the limbs of an organization. Engineering teams within the business body had to build custom connectors every time a team at the edge wanted to improve how it showed up in the world. It was slow, tedious, and inexact all around.

With the addition of reverse ETL tools like Census, a whole new world of action and coordination is available to those teams on the front lines. Reverse ETL tools, when built like Census, integrate with modeling in the data brain and enable the company to sync clean, unified data back into the body via pre-built API integrations.

As a result, engineering and data teams behind the business body don't have to constantly keep an eye on the manual string connectors that link together the body. Instead, data can be validated automatically out of the warehouse and sent to the limbs that need it most.

Behavioral event tracking and ETL: the peripheral nervous system

Your peripheral nervous system is the series of nerves that connect your central nervous system (brain and spinal cord) to the rest of your body. In the modern data stack, ETL tools like Fivetran and behavioral event tracking tools like Snowplow relay vital data between the outside world and your data brain.

These tools, which combine together to gather sensory data from your teams and tools at the edge, relay and transform information for your brain to learn from.

Your behavioral event tracker (Snowplow) takes information from your web and mobile apps and sends them to your brain for further analysis. This layer of collection gives your data brain a clear picture of how users interact with your program in the wild so downstream recommendations are tied to the experiences in the real world.

Your ETL data loader lets you copy data from systems of records like Salesforce and Marketo and clean and transform it back into the common structure used by your warehouse for storage and later modeling.

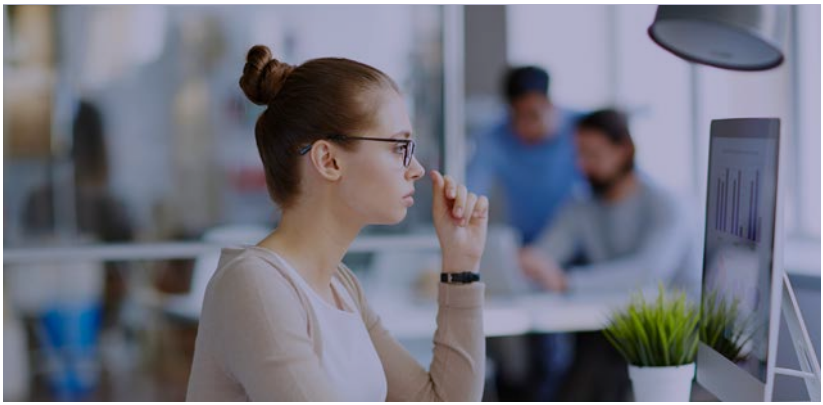
The combination of these two layers of tools helps your data brain contextualize raw data from the world and learn from it. This ensures each team at the edge of your business--whether sales, marketing, support, or beyond--works together in real-time to assist your customers and evangelize for your product.

Sales, marketing, and support teams: The limbs of the company

At the end of the day, all the data processing work done upstream of your sales, marketing, and support teams is to enable these limbs to function better.

After all, it's these tools--such as Salesforce for sales, Hubspot and Marketo for marketing, Pendo for product data analysis, and more--that make your business and your customer relationships possible. And they can't do their jobs well without intelligent recommendations from your brain or the real-time relays on your connecting nerves.

When these limbs are powered via a healthy nervous system, you get amazing results. Your frontline teams can use segments of users to drive personalization, leverage account health score and reports, retarget Facebook and Google ads to the correct people, and receive notifications for account activities.



The secret to a long and happy life with data? Operational analytics.

To reach new heights, it's not enough to just act on what's happened in your life so far--you need real-time data to power in-the-moment decision-making to make better decisions moving forward.

This is only made possible if you're nervous system is healthy--and connected to the world around it with the strong spine of reverse ETL with Census and behavioral data collection with Snowplow.

If you want to learn more about how Census and Snowplow combine to help you and your team operationalize your data at scale, check out our upcoming series on operational analytics over on their blog this week and next. We'll also be hanging out with their team over on the Sequel Show later this month, and via webinar in July.

CHAPTER 2

WHY WE NEED TO LOOK BEYOND CDPs TO DELIVER EXCELLENT EXPERIENCES TO CUSTOMERS

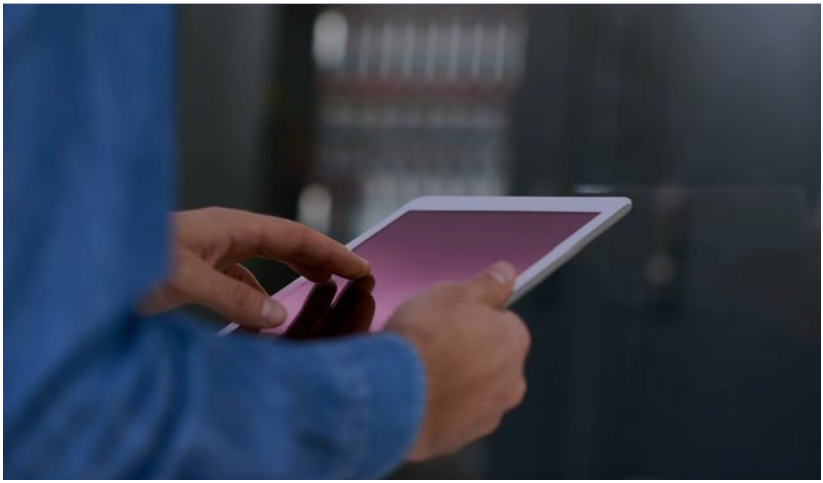
Building a single customer view is a key objective for many businesses today.

Users demand an excellent, increasingly-personalized experience across digital platforms, or, at the very least, that companies tailor marketing and recommendations to them. This not only requires effective user identification across platforms and over time, but also requires companies to activate user data in downstream systems. In this chapter, we will review the tooling that companies have been using to tackle this important topic, and how we see the landscape changing.

The emergence of CDPs

Developing a better understanding of customers came into focus as early as the 1970s, when Customer Relationship Management tools (CRMs) first emerged and companies could build customer profiles based on transactions and other interactions. With the rise of big data tools and the ability to collect granular behavioral data, a new category of customer data tool emerged, the Customer Data Platform, or CDP. CDPs promise to make working with data easy for non-technical marketers, bringing all customer data together and stitching it into a single customer view.

More specifically, the CDP Institute defines a CDP as “a packaged software that creates a persistent, unified customer database that is accessible to other systems.” Customer-facing teams can leverage this software in various marketing tools, such as advertising platforms, email marketing solutions, and more. However, they generally don’t collect behavioral data from owned applications (such as websites and mobile apps), or make that underlying event-level data available for querying in a data warehouse.



Today, there's more than 100 wildly different vendors claiming the title of CDP.

> In a report from January 2020, Gartner stated the *"Hype about customer data platforms (CDPs) as a panacea for customer-related problems is liable to confuse data and analytics leaders".*

In layman's terms, the wide promises of the CDP market can make too much of everything seem possible. Tools like Segment or Tealium, which originally started as tag managers, now allow companies to track behavioral data across all their platforms and channels and then send it to many third parties, as well as the data warehouse.

Key limitations in building a single customer view using CDPs

While the adoption of CDPs enables marketing teams to personalize and optimize campaigns, there are some downsides to embracing the CDP for all your customer use cases:

1. **Companies relinquish ownership of and control over some of their most important data:** how their users behave across all of their digital platforms and channels.

2. As these tools collect this data purely as a means to an end, **data quality is often an afterthought.**

3. **Leveraging these tools can lead to data silos.** Since CDPs primarily focus on the marketing use case, other teams within the business have to procure or build separate solutions for their needs.

4. **Companies are limited to sending raw data on to third parties,** often without the ability to aggregate over data, add business logic or derive insights to forward to those third parties instead.

All of these issues will be discussed in more detail on the following pages.

1. No ownership and control over data

CDPs typically require sending all of your user behavioral data to their systems. This means you do not own your data, nor do you have control over where the data is processed and stored. This makes it increasingly difficult to comply with data privacy regulations such as GDPR and CCPA, as well as "right to be forgotten" requests that can arise due to these regulations. This also locks you into a specific vendor and makes you vulnerable to their pricing changes over time.

2. Low data quality

Because most CDPs are offered as public SaaS, they collect data about your customers and users as a third party. As a result, they're often blocked by tracking restrictions introduced in the last year, including Apple's ITP and other browser tracking prevention methods, as well as the privacy controls introduced in iOS14. All of which make it difficult for companies to reliably identify all their users. Furthermore, ad blockers can easily pick up CDP tracker names, creating further gaps in the data.

To enable easy forwarding of the collected data to third parties, data must follow a predefined, inflexible format, with limited ability to add more information to the events. Moreover, preprocessing of the data often happens in a black box. For example, it's difficult to know how potential bot activity is treated, how different user identifiers are stitched together, or how much data is being excluded due to tracking issues.

In essence, companies often trade data richness and data quality for the ability to easily forward data to third parties. This tradeoff makes the use of data by other teams in the organization time consuming or outright impossible.

3. Data is siloed

The lack of flexibility around data collection methods, processing, and storage generally means that other teams across the organization end up procuring their own solutions for their use cases.

For example, the product team may use a dedicated product analytics tool, such as Pendo, Amplitude or Mixpanel, while marketing may use a packaged analytics tool such as Google Analytics 360 for their reporting and attribution combined with a CDP to manage marketing automation. Adding onto an already crowded stack, data engineering and data science teams often end up building their own solution to collect exactly the data they need for use cases such as personalization or recommendations.

Not only is this a significant investment for the business, different teams will also consume different data sets, which can lead to inconsistent decisions and distrust in data. The lack of a centralized data asset and holistic approach to understanding user behaviour makes it difficult to align on common strategies and build a truly data-informed culture in an organization.

4. Events, not insights

CDPs such as Segment focus on forwarding raw events to third parties. While this method gets raw data out to tools team use, there's three main downsides to this data flow:

1. **You can't add meaningful business context or aggregate data before sending it on.** For example, as an ecommerce site, you might send purchase events to Facebook to target new users. With a CDP, you wouldn't be able to reconcile purchases with returns ahead of time, wasting your ad budgets on customers already primed to return.

2. **You can't calculate the correlation between different events or recommended next actions in your own systems.** For example, this means Facebook would determine that users from a certain location are more likely to convert, but wouldn't necessarily share that information back with you.

3. **You're leaking valuable information about your users to third parties that could, in some ways, compete with you.** For example, a recruitment platform may share information about how users look for jobs with Facebook and Google, which also operate as job platforms.

Thus, companies are looking for new approaches to understanding their customer behaviour, and then leveraging that information to drive relevant marketing activities and meaningful user experiences across their platforms and channels.

The emergence of the modern data stack

Given the challenges discussed in the previous section, companies have started to move away from this approach toward a more modular setup, commonly referred to as the 'Modern Data Stack'. Behavioral data platforms like Snowplow collect user data centrally and load into the data warehouse, where it can easily be joined with other data sets such as transactional data, CRM data, and more. All of this data can then be modeled for different use cases across the business using a tool like dbt. This prepares the data for consumption in BI tools like Looker or Tableau, and positions companies well to send this high-quality customer data back out to the front line tools they use most with reverse ETL with Census.

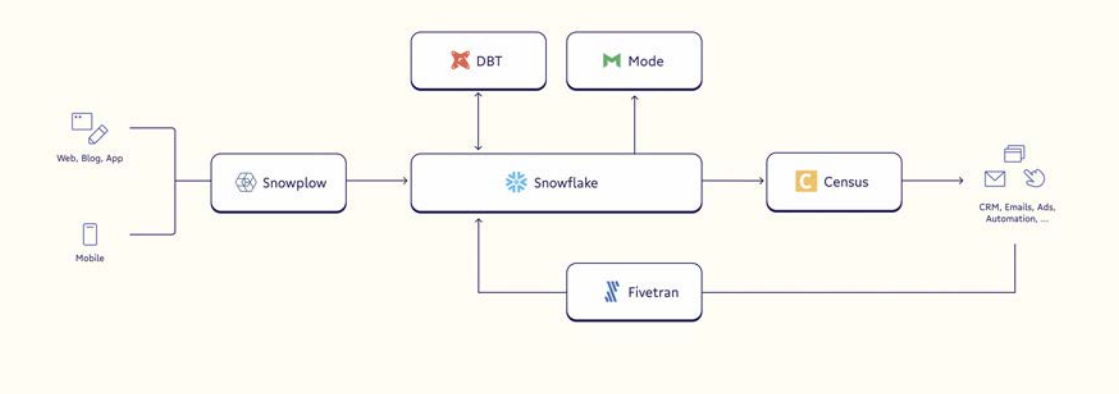


Image courtesy of Census

For example, an e-commerce company can build a segment of high-value customers in their data warehouse based on all the data they've collected throughout the sales journey, including browsing behavior, purchases and returns. Census can then easily forward this segment to all the relevant marketing tools, treating the data warehouse as the single source of truth for all connected apps across the company. Simultaneously, other teams across the organization can leverage this single source of truth too. For example, the product team can use the same data to build a product recommendation engine, or the buying team can leverage that data in dashboards that analyse trends and user preferences to inform their buying strategy.

For more information on using a CDP versus collecting data through a behavioral data platform, as well as using a reverse ETL tool to activate it in downstream systems, check out this excellent post by Census on the CDP that is already sitting in your data warehouse.

Coming next

In this chapter we have reviewed CDPs and the role they have played in helping companies build a single customer view. We also discussed their limitations, and what new approaches we see to help companies deliver excellent experiences to their customers and users. In the next chapter, we will dive deeper into how you can leverage Snowplow and Census to deliver behavioral data for actionable outcomes.

CHAPTER 3

SNOWPLOW AND CENSUS: DELIVERING BEHAVIORAL DATA FOR ACTIONABLE OUTCOMES

It is difficult to overstate the potential of behavioral data to an organization. At its essence, behavioral data is truth. As Zach Wilson, Tech Lead at Airbnb succinctly put it:

> *Data, when correctly generated and processed, is truth. Giving your company truth to correct biases is critical for a fairer, more efficient, less error-prone world...When data reveals something counterintuitive, that's really when data shines.*

Behavioral data is so special because it reveals the truth about how your users and customers interact with your product. The "truth" can be used within a business context in all sorts of ways, with hugely powerful implications. With it, you can improve the recommendations your customers see on your website; you can discover the types of customers that are more likely to convert. Better still, you can gain a better understanding of your customers in general, how they behave within your product domain and why they might do what they do.

Armed with this information, you can then set a course for building competitive advantage. Not all your competitors are so equipped with the means to understand and serve their customers. This means doing behavioral data – namely the challenge of capturing, managing, modeling and delivering behavioral data to key decision makers – is a frontier on which there are winners and losers in the modern business world. The winners, Spotify, Airbnb, Netflix and the like, are steaming ahead. For the rest of us, it's a battle just to get in the race.

Why driving value with behavioral data is such a challenge

Capturing and managing behavioral data to drive competitive advantage sounds great, but it's far from easy. There are a number of complex challenges involved with driving value from behavioral data.

These range from the technical challenges:

- Building (or buying), managing and maintaining a robust data stack;
- Aggregating data sets from disparate sources into a cohesive data model that means something to the data consumers who use it;
- Handling large volumes of behavioral data.

To the people or operational challenges:

- Proficiency with data and analytical languages like SQL means that data literacy is not evenly distributed throughout the organization;
- Often data is broken up into 'silos' because individual teams prefer to work with their own packaged tools;
- Trust in the data can easily break down when data quality issues arise, which can negatively impact the whole data culture;
- Without a shared language around data, communication can break down between different stakeholders.

This is not even an exhaustive list, but a hint of the challenges organizations face when building their behavioral data set. At the heart of the issue is the idea of data democratization – ”How can I provide the right data, to the right people, where they need it?”. While some companies such as Strava have made great strides towards a culture around democratized data, many organizations struggle to fulfil this ambition.

As the company size increases, the scale of the challenge becomes magnified. There is more data flooding in, more disparate teams to serve with their own individual needs and requirements, and the tangled knots of data supply lines only get messier.

The role of behavioral data management

Tooling, processes and frameworks around behavioral data management play an important role in addressing these challenges. With the arrival of methodologies like DataOps, which borrow the popular principles of DevOps, we have seen new thinking around how organizations can handle behavioral data more efficiently to empower key decision makers in, say, product and marketing teams.

We have seen innovations in this area, from different ways to structure data teams and their placement within organizations to new tools in the landscape, both left and right of the centralized data warehouse – the modern "brain" of the business. Better warehousing and modeling, more robust pipeline infrastructure and more intuitive BI tools have all played a part in easing the pain of distributing data to those who need it.

To summarize a few areas, we have seen:

- Better data management tooling that helps data teams send behavioral data to the data warehouse, where it can be aggregated and filtered into BI tools like Looker, Tableau and PowerBI where it's more accessible to data consumers;

- Infrastructure and processes that help organizations build assurance in the quality and robustness of their data;

- A move towards total ownership of behavioral data, enabling data teams to apply their own logic to their data, leading to more meaningful data for the data user;

- Shifts toward the centralization of the data team in modern organizational structure, which allows data teams to sit under every part of the business;

- Warehousing data from multiple sources means there can be a single source of truth to drive richer insights and act as fuel for countless use cases.

These are all great steps forward. Even five years ago, we did not have access to the plethora of data tools available today. Nor were there thriving communities of data practitioners sharing and showcasing exciting ways they drive value with behavioral data in their industries.

The data warehouse has arguably been the most transformative influence in the recent data revolution. As a consequence, new tools and solutions have emerged to complement and capitalize on it.

But data warehouses have not solved one final step required on the data journey. We have fixed the “How do I get the right data, for the right people”, but not the “where they need it most” part of the equation. For that, we need to look at the next innovation in the data landscape, where data is not only captured and stored, but ‘operationalized’ for front-line business teams.

Operational analytics: Freeing up the data from the data warehouse

Operational analytics may seem like another opaque data phrase or buzz word, but it carries huge significance in this last aspect of delivering value from behavioral data. Making data 'operational' means putting it in front of the teams who need it most – the marketing, product and sales teams who drive business growth.

Or, as Boris Jabes, Census CEO and early pioneer of operational analytics, explains:

> *"At its core, operational analytics is about putting an organization's data to work so everyone can make smart decisions about your business. You might've heard such a promise before from other technologies or platforms, but operational analytics is the only way to accomplish this at scale, because it introduces a set of fundamentals for leveraging data across your organization." - Boris Jabes, Census CEO*

If we consider the day-to-day schedule of a Product Manager, we can bring this problem to life. Let's imagine a Product Manager, Rita, logs into a product analytics tool first thing in the morning to see the results of an A/B test the team has conducted.

In the A/B test, a small cohort of users are shown a banner in the mobile app, inviting them to check out the latest promotional offers on sale. The test is set to determine whether users who are shown the banner are more likely to add an item to the basket than those who don't. Rita looks at the numbers on a continual basis to see how the A/B test is progressing, and determine what it can teach her team.

Within the product analytics tool, Rita can see the users and gain information about how they behave within the app. She may be able to see what devices they're using, what region they live in and the lifetime value of purchases they've made. These are all useful insights.

However, what if Rita could drill into far richer information about those users and their behavior? She knows, for instance, that the company captures a wealth of data about their users, and that it's possible – if she signs into specific dashboards in another BI tool – to discover every search entry, every button click and product selection those customers have made on an individual basis.

If these insights were available in Rita's product analytics tool in real time, her team could orchestrate more powerful A/B tests, allowing them to answer questions like:

- What impact does a banner ad have on these types of users?

- What happens when people who search for these queries are shown these products?

If Rita could connect the dots, she could build a deeper understanding of her users to inform the company's product roadmap. She would be able to use that information to truly enhance the user experience.

Closing the loop on the data lifecycle with Snowplow and Census

The above scenario gives us just one example of how operationalizing data can open doors to richer insights. Yet the potential for operational analytics goes even further – to equip sales, marketing, product and other teams with behavioral data in the platforms they use to drive growth. To truly operationalize data by syncing it into apps like Marketo and Salesforce with "last-mile" models, companies need to look to a new breed of data tooling: reverse ETL

While tools like Snowplow are ideal for capturing and delivering behavioral data to the data warehouse, reverse ETL solutions like Census allow product and marketing teams to unleash data into their platforms of choice. With this powerful combination, the organization retains total ownership and control over their valuable data set and can better ensure data freshness and quality throughout their pipeline, all without facing the challenges caused by relying on Customer Data Platforms (CPDs).

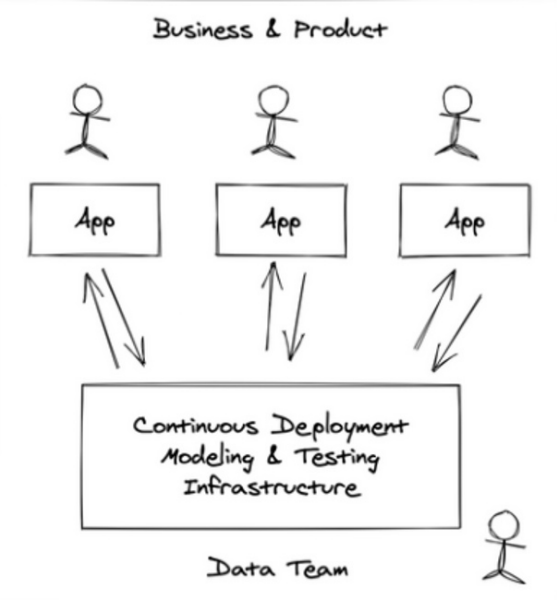
With data ready and available in CRMs, marketing platforms and product analytics tools, front-line teams can unlock behavioral data that would otherwise sit in the warehouse. There are a few key advantages to this:

- **The data is where front-line teams need it most,** where they can 'action' it to drive real user outcomes. For instance, a marketing team equipped with behavioral data can improve their email campaigns from their marketing platform.

- **Data teams and analytics are no longer a bottleneck** for data productivity because individual teams can self-serve.

- **The single source of truth persists across the data journey** – data silos are broken down as access to the data itself is democratized across teams.

There's another key advantage to operational analytics, in that the data streams can be bi-directional. This means that, unlike with CDPs, behavioral data can flow both from the data warehouse into a marketing or sales tool via reverse ETL tools like Census, and back into the data warehouse again via ETL tools. In this way, there is a continuous loop of data flowing from the warehouse into operational platforms – as well as a feedback cycle where the centralized data asset can be enhanced by operational data (as seen below in a rough sketch of Boris's).



The business intelligence of the whole company is therefore augmented by this two-way street (we'll explore this more in our next chapter).

CHAPTER 4

ENHANCING THE SNOWPLOW TRIAL EXPERIENCE WITH CENSUS

At Snowplow, we had a problem with product discovery. We were confident in our tooling – we knew our behavioral data engine was robust, arguably the most expansive, reliable solution for collecting rich behavioral data. The problem was (and sometimes still is) our prospects and customers didn't know this. Worse, they had no easy ways to find out exactly how Snowplow could help them and their business. We could wax lyrical about the benefits of great data quality and rigorous processes for validation, but so what? It's difficult to truly understand such abstract concepts without seeing them for yourself. Without experiencing the advantages of Snowplow data, all our most passionate marketing and storytelling would leave our prospects cold.

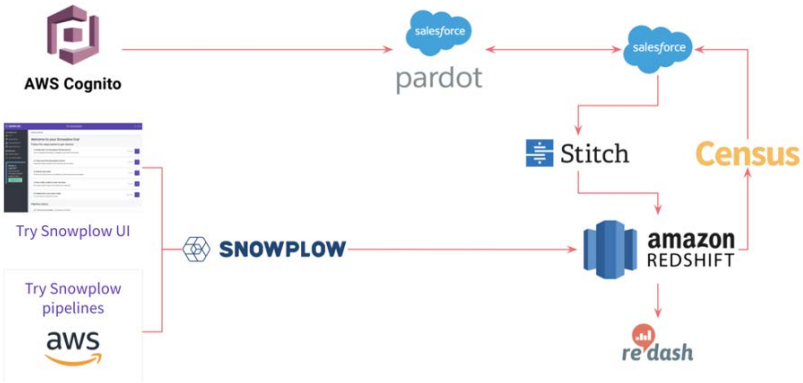
We realized we needed a way for our prospective users and customers to try Snowplow for themselves. Before you invest in an expensive car, you want to take it for a test drive, after all. The same is true for an investment in software, especially when the tool handles your company's most important asset: your behavioral data.

We got together and launched a trial version of Snowplow, quaintly named "Try Snowplow." To give people a taste of the software, users could sign up to a 'Try Snowplow' account, spin up a small version of the Snowplow pipeline in their own cloud account and fire off events into their data warehouse, all in 30 minutes or less.

Naturally, we were curious to see how the trial experience would be received. But beyond that, we wanted a deep understanding of those using the product, and how we could make it better. We wanted to know what challenges they would face, the bottlenecks they might encounter, and whether or not the experience would help them discover the benefits of the Snowplow product.

How we measure Snowplow trial engagement

In order to better understand our users, we assembled a stack to capture their interactions at each stage of the trial experience (without, of course, seeing what data they capture for themselves). In a nutshell, here's what our data stack for the "Try Snowplow" experience looked like.



We deployed tracking across the web UI for Try Snowplow as well as Scala tracking in the Try Snowplow pipelines themselves. Snowplow sends this data into our data warehouse (Redshift), where we're able to use SQL Runner to aggregate important pieces of information together to build a picture of our users. It also means we can qualify potential leads early in the process, before alerting our sales team to new activity (more on that later).

Census delivers the behavioral data we have in Redshift into Salesforce, giving our sales and marketing teams a clear view of our user interaction. Although there are a lot of moving parts, there are clear advantages to building a stack that can deliver insights into our user behavior in both our data warehouse and our CRM.

Delivering richer insights with Census

Product analytics can be full of red herrings. For example, if a user spins up a Snowplow pipeline but quickly abandons it without taking any other action, we could hardly consider them a highly engaged user. Conversely, another user might never deploy a pipeline, yet spends a lot of time browsing documentation, asking questions in our Discourse forum and reading articles on our website, that user might be inferred as a highly engaged prospect.

These nuances are only perceptible when we have access to rich behavioral data. Thankfully with Snowplow, we can capture all sorts of information about how trial users interact with our channels, mapping out their journey in detail from the buttons they click to the time they spend reading documentation. This is great for analytics, but what about helping our sales team make informed decisions?

As we mentioned previously, operational analytics is not just about capturing granular data, it's about actionable insights that front-line teams can use to drive success. With Census, we're able to deliver this data into Salesforce, which allows our sales team to build meaningful relationships with prospects. By seeing how trial users are interacting with Try Snowplow (for example, by determining which use cases they're most interested in), our sales team can generate richer, more engaging conversations with them. This visibility enables us to provide better support to prospects, and gain a deeper understanding of the bottlenecks in our product experience.

> One reason we love Census is that it gives us that layer of intelligence, where we can add business logic in the data warehouse and send it on to our sales and marketing platforms. – Cara Baestlein, Product Manager at Snowplow

Having Census relay insights directly to sales reps has an immediate impact on their success. It keeps the sales team informed about the prospect's needs and ambitions, which makes dialogue smoother for both parties. As for the product team, understanding the user journey enables them to identify potential drop-off spots, or stages where people get stuck.

Operational cohesion: better together

Being able to send behavioral data to and from Salesforce has not only helped us improve the trial experience for users, but also to streamline internal workflows. We don't have to rely on engineers to constantly implement new tracking, or take up an analyst's time building dashboards. Better still, our sales team can continue working within the platforms they're most comfortable with – which makes them far more inclined to put this data to good use.

One of the best outcomes is that our sales and marketing teams can work with our modeled, meaningful data in Redshift, which means we can apply our unique logic to the data and set strict criteria to those users we would classify as "a prospect". By specifying which actions the individual user needs to take before a sales representative reaches out, we ensure that prospects are only contacted when they're ready.

Census has been instrumental in allowing us to close the loop on the data lifecycle – bringing sales, marketing and product together to learn, iterate and improve. Because the data is bi-directional, traveling both from and to the data warehouse into Salesforce where it can be put into action, there is a constant flow of intelligence for product managers and sales reps to work with.

While we're still working on making Snowplow to be easier to discover, trial and experience, the flow of behavioral data into our front-line platforms empower us to make incremental gains along this journey.

SNOWPLOW

Snowplow is a behavioral data management platform that collects and operationalizes behavioral data, at scale. We empower you and your team to rise above the difficulties of data delivery and organization, enabling you to focus on your data journey.

Each company has its own data use cases and bespoke data requirements. We provide you with the flexible tools that enable you to collect the data that's most important to you.

Snowplow empowers you to collect high-quality data in an accurate and efficient manner, giving you more time to analyze and drive value from your data.

Trusted by data-driven companies, Census is the operational analytics platform that syncs your data warehouse with all your favorite apps. We empower everyone in your organization to take action with data, no engineering favors or custom scripts required. We take care of the integrations with your existing infrastructure so your team can focus on driving your business forward.

Learn how Snowplow and Census make for a powerful combination

BOOK A SNOWPLOW DEMO

BOOK A CENSUS DEMO

getcensus.com

snowplowanalytics.com