



The Composable CDP: Powering Customer 360 with Snowflake



Introduction

Companies are more desperate than ever to access a single unified profile of their customers, but with the explosion of SaaS tools, ever increasing data silos, death of third party cookies and increasing privacy regulation it's never been harder than ever.

Five years ago, Customer Data Platforms (CDPs) started gaining popularity by marketing themselves as an all-in-one solution to collect, unify, and activate customer data. However customers who adopted off-the-shelf CDPs have long since struggled with their rigid data models, long onboarding times, and redundancies across analytics and marketing tools with [only 1% of companies actually meeting their current and future needs with CDPs.](#)

Now, the rise of the [modern data stack](#) and data activation means that companies already have access to best-in-class solutions for each component of an off-the-shelf CDP as part of their existing data platform. [Snowflake's](#) Data Cloud has emerged as the most complete source of truth for customer data inside the organization. And with other best-in-class components like [Fivetran](#) for ELT, [dbt](#) for data transformation and [Census](#) for data activation, the result is an ideal solution for data-forward companies looking to maximize their existing data investments: **The composable CDP.**

Who is this whitepaper for?

This whitepaper is for data teams who work closely with marketing teams and are looking to understand what CDPs are, as well as learn more about the composable CDP alternative to buying an off-the-shelf CDP. Perhaps your marketing team is looking at purchasing an off-the-shelf CDP, or you're looking to enable more self-service access to data for your marketing teams. In this whitepaper we'll discuss off-the-shelf CDPs, the rise of the composable CDP, and how to turn your data warehouse into a composable CDP to start getting value in your organization today.

What is a CDP and why do people buy them?

A Customer Data Platform (CDP) is an all-in-one platform built for marketing teams to collect, transform and activate customer data. The number one reason why marketing teams are rushing out to buy CDPs is to get a unified view of their customer, with [63% of marketers](#) saying that unifying data is their key challenge.



There's a lot of variety within the CDP category, but generally, all CDPs have (at least) these four components.

The Four Components of a CDP

1 Data Storage

CDPs have their own proprietary database that enables them to aggregate and stitch the customer data they are collecting. For packaged CDPs, this database does not match the structure of an analytics application and therefore can't be queried directly.

2 Data Collection

To power their customer profiles, CDPs need a way to collect customer data. For event streaming data they provide APIs for developers to track user traits and events across websites and mobile applications. They also provide basic native integrations to ingest data from operational sources or other databases.

3 Data Transformation

CDPs usually have out-of-the-box identity stitching functionality to power their "unified customer profiles". Generally, this is done with deterministic matching.

4 Data Activation

The core value prop of CDPs are helping marketing teams take action on data. So they provide audience-building capabilities and integrations enabling you to sync customer profiles, events, and audience segments to marketing tools.

You might be thinking, *"Hey, these components sound mighty familiar to those of the modern data stack"* — and you'd be right. The CDP provides a lot of overlapping functionality to that of the company's existing data platform. So... Why do data teams want them?

Typically, there are two scenarios where data teams are looking at CDPs:

For their event collection features. CDPs like [Twilio Segment](#) started as event collection and tag management tools. As a result, they offer incredibly developer-friendly APIs that make it really easy to track new events. The biggest downside to this is that because they are an all-in-one solution and have a volume-based pricing model based on Monthly Tracked Users, they quickly become very expensive for companies with many customers. This makes

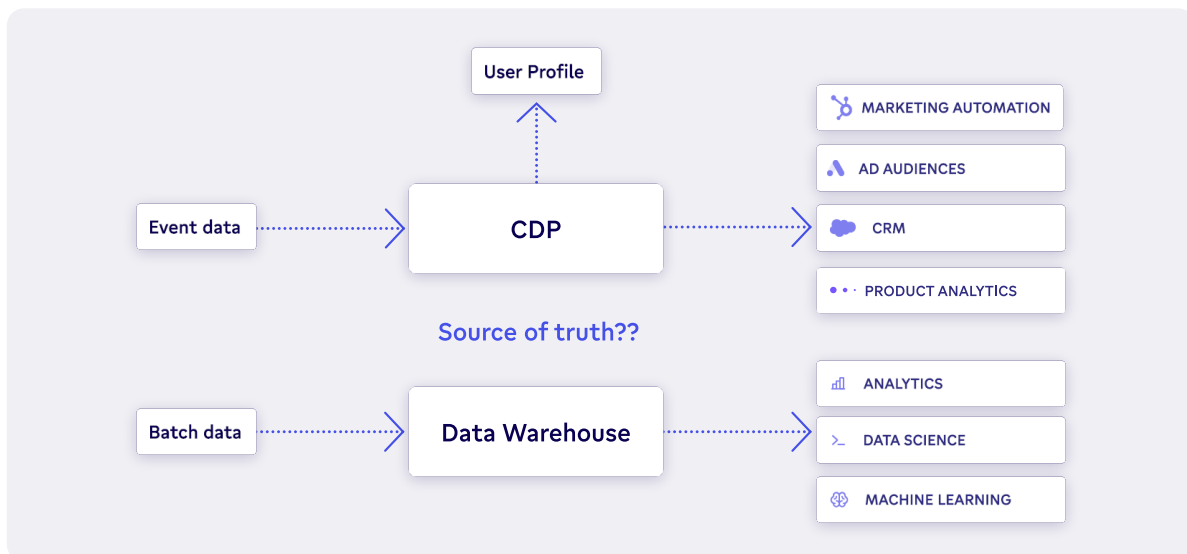
all-in-one CDPs a very expensive event collection solution — especially with other more cost-effective options in the market like [Snowplow](#).

Because their marketing team wants to buy one for data activation purposes and needs them as part of the evaluation and implementation process.

The biggest problem: CDPs are yet another copy of your data

The biggest problem with CDPs is that they claim to be the single source of truth for customer data, but they do not (and cannot) replace your data warehouse. Why? Simply because CDPs don't have all the customer and company-level information to be a complete source of truth for analytics.

Instead, data collected by CDPs is copied to the data warehouse to power trusted analytics, instantly diluting the CDPs' promise of a "single source of truth." While some CDPs are now supporting importing data from the warehouse, doing so results in additional data latency and still doesn't solve the fact that the CDP is still just another silo of data. Due to this fragmentation, data inside the CDP is not as trusted or fresh as in the data warehouse.

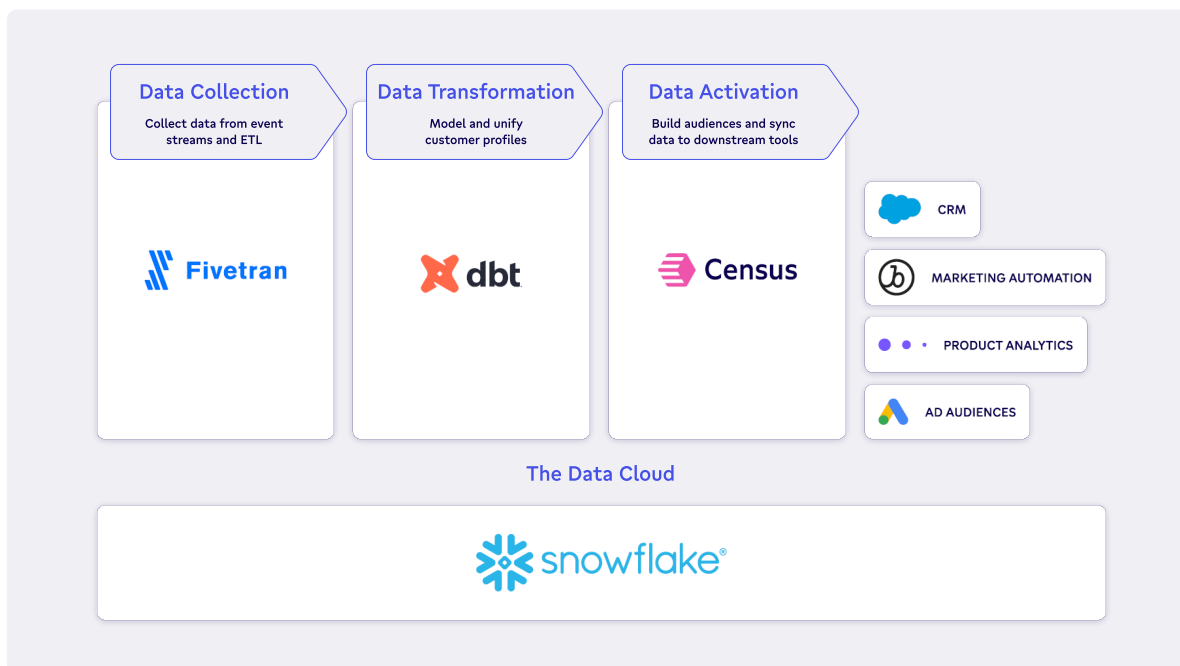


You don't need yet another source of truth

What is the Composable CDP?

The composable CDP is a **best-of-breed** solution that serves as an alternative to buying an expensive all-in-one CDP for data-forward organizations. Data teams can turn their existing data platform into a CDP by adding the data activation layer, since CDPs already duplicate a lot of the functionality from a company's existing data platform.

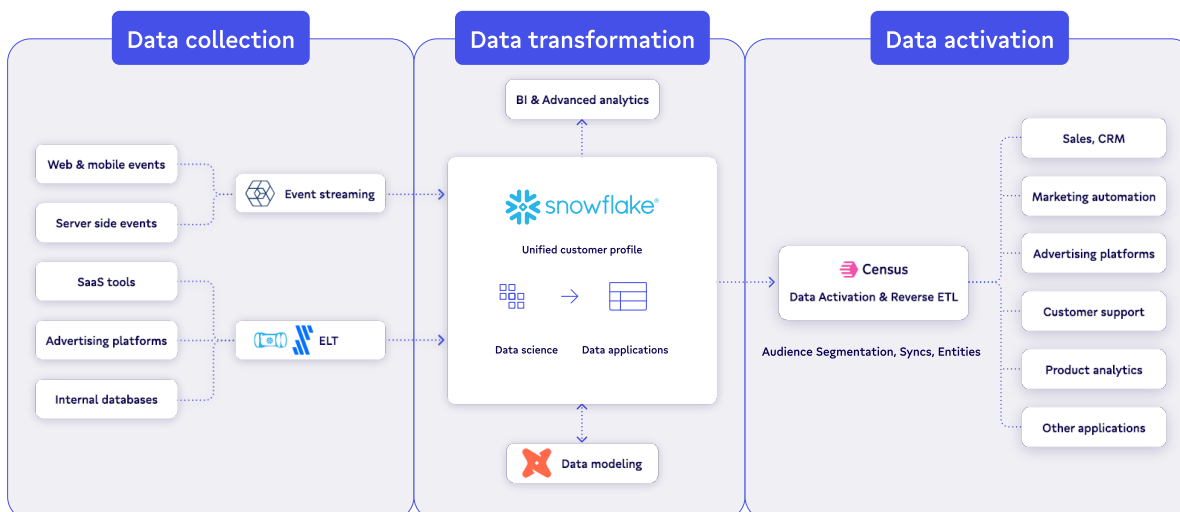
This alternative approach to a CDP is known by a few names: Composable CDP, unbundled CDP, headless CDP, etc. Although the name might differ depending on who you're talking to, they are all ultimately saying the same thing: Use your data warehouse as the source of truth for customer data, and **activate data where it already lives**.



“Our time-to-value for these projects — to build our CDP customer attributes and push them into Braze to enable marketing campaigns – has really been reduced to minutes instead of having to struggle and build something in-house over a period of months.”

Sameer Vaidya, *Data and Integration Architect, Plume*

How do you build a Composable CDP?





Creating and maintaining a single view of the customer delivers a tremendous amount of value for organizations big and small. Today, the Snowflake Data Cloud helps thousands of organizations across industries centralize all their data, overcome data silos, to generate a holistic view of their customers and unlock value for marketing. With capability that expands beyond modern data warehouses, such as applications, data science, and data sharing/collaboration, the Data Cloud is uniquely positioned as the foundation of the composable CDP, offering organizations a single platform for all their data.

Data Cloud vs CDP

Completeness: The Data Cloud Modern data warehouse covers all of the enterprise/business data (beyond just integrating customer data), which allows for deeper efficiencies and new use-cases to improve customer experience.

Unlimited Storage and Data History: Off-the-shelf CDPs often control their costs by limiting how far back data can be stored within their data warehouse.

Privacy and Security: Off-the-shelf CDPs replicate customer data into their data warehouse, creating privacy and security risks for the business.

SQL Access: Off-the-shelf CDPs don't provide SQL-level access to the data warehouse and instead wrap their solution in proprietary APIs. This limits

internal data teams' ability to work quickly and easily with their existing tools and processes.

No Redundant Costs: Because off-the-shelf CDPs replicate data already in the warehouse, companies pay twice for data warehouse services.

Data Collection

There are two types of data collection services that are part of the Composable CDP:

1. **ELT connectors**, that extract data from SaaS tools and databases
2. **Event connectors**, that capture click events that occur with web and mobile, and other digital applications

Extract , Load , Transform (ELT)

ELT, or Extract, Load, Transform, is the modern approach to replicate data from your SaaS tools and databases across marketing, sales, finance/IT, product, etc. into your data warehouse. The speed and reliability of the warehouse enables transformation to happen after data is stored, and has replaced the traditional ETL approach.

Fivetran



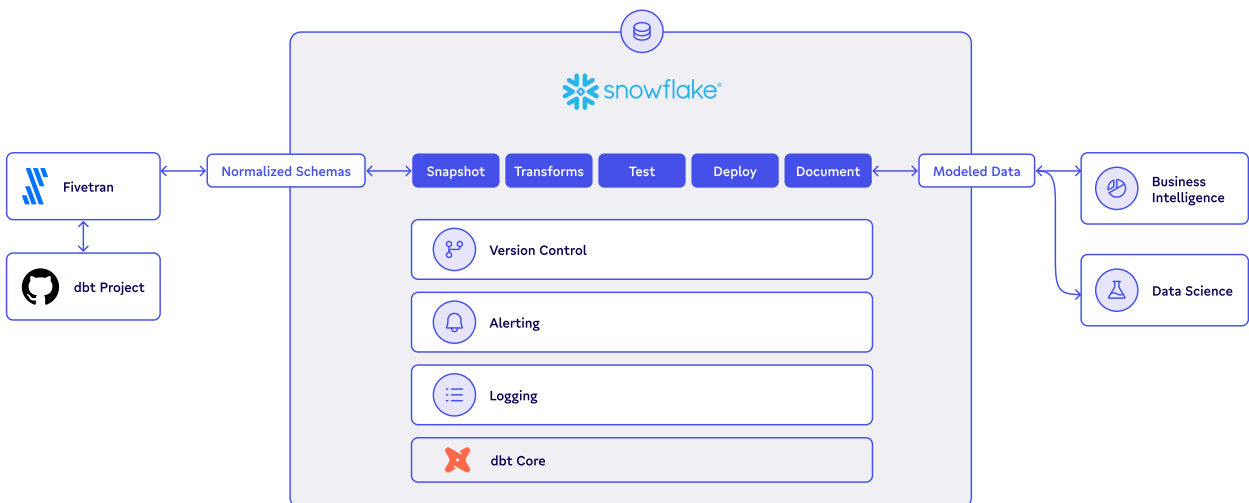
Fivetran, the global industry leader in modern data integration, powers high

volume data movement for businesses. Built for the cloud, Fivetran enables data teams to effortlessly centralize and transform data from hundreds of SaaS and on-prem data sources into high-performance cloud destinations.

Its 200+ prebuilt connectors that automatically adapt as schemas and APIs change ensures consistent, reliable access to data. Fivetran makes data always accessible to analytics organizations that rely heavily on using a centralized cloud data warehouse or data lake.

“Fivetran helps us recognize what’s working and influences how we optimize the opportunities we uncover. It makes us agile and more responsive to customer needs. Most importantly, though, is that we’re able to deliver this agility without requiring engineering resources.”

Greg Roodt, *Head of Data Platforms, Canva*



Fivetran vs CDPs

There are several benefits to using Fivetran’s connectors instead of a CDP:

Data completeness: Fivetran’s 200+ prebuilt connectors are optimized

for both SaaS apps and data lakes and support a broad variety of databases. Fivetran supports data sources spanning all departments and use cases, whereas CDP prebuilt integrations tend to focus solely on marketing and sales.

Query ready: Fivetran's integrations are purpose-built to normalize the data so that it's query-ready for your data team to work with in the warehouse. Fivetran connectors typically come with a prebuilt entity relationship diagram for easier immediate report building.

Privacy and Security: Fivetran replicates but does not own data during ingestion to your data warehouse, alleviating any privacy and security concerns.

Data freshness: Fivetran syncs data as frequently as every five minutes, allowing you to act on the latest data.

Event collection

Capture rich, quality behavioral data across all platforms and channels in a common format and stream it into your data warehouse.

With recent changes in browsers and operating systems, the implementation of event collection has become more challenging in today's privacy-conscious environment. Due to the blocking of third-party cookies, it is now difficult to link anonymous users across different domains. Additionally ad blockers and cookie consent pop-ups can prevent client-side collection services from functioning, resulting in a loss of data.

Some CDPs such as Twilio Segment primarily gained popularity due to their

event streaming capabilities that overcame the limitations of reporting tools like Google Analytics which, at the time, were only capable of tracking page views rather than individual click events.

With the growing emphasis on privacy and the emergence of the data warehouse, the features of event collection platforms have undergone changes. In 2020, Google released a new version of Google Analytics, known as GA4, which is designed to support user event tracking and it is recommended to be deployed server-side, to avoid being blocked by ad-blockers. Similarly, other popular event collection tools such as Mixpanel, Amplitude, and Heap have also upgraded their services to support server-side deployments as well as allowing for seamless data synchronization with data warehouses.

Why choose your own event collection vs CDP

Snowplow and Snowcat are a new category of event collection tools designed specifically for the modern data stack. Unlike CDPs these tools only send data to your data warehouse and do not store a copy of the data themselves. Notably, we don't recommend Segment or Rudderstack in this category purely from a cost perspective, as both of these companies position themselves as all-in-one CDP solution providers and end up being more expensive down the line.

The two biggest benefits of using a best-of-breed event collection solution over a CDP are:

- **Data Ownership:** Data is sent directly to your data warehouse, thereby avoiding privacy and data residency issues.
- **Data Quality:** Specialization in event collection translates to richer and better quality data than event collectors designed for a more specific purpose like CDP or Analytics. For example Snowplow collects 120+ user

properties, whereas Twilio Segment's CDP only captures 23 user properties.

Data Transformation

Data transformation is the process of joining different data sets into useful groupings. CDPs provide a specific form of data transformation known as identity resolution. Identity resolution stitches together a unified customer id from data across multiple systems. Identity resolution is one of the key services off-the-shelf CDPs provide as part of their packaged offering, however, many companies outgrow the basic capabilities offered by these CDPs, and data teams often do not trust them.

CDPs force customer data to conform to a rigid structure. This might appear efficient at first, but businesses come in all shapes and sizes, meaning they require much more flexibility than what these CDPs can offer. The composable CDP is built on the data warehouse, which means it's built on top of trusted models and transformation maintained by the data team using a tool like dbt. This flexibility allows you to represent relationships between users and entities like accounts, teams, workspaces, or devices. Your business can have complete control over how you unify customer identity, rather than identity resolution capabilities provided by CDPs that you have no control over.

dbt



dbt™ is a SQL-first transformation workflow that lets teams quickly and collaboratively deploy analytics code following software engineering best prac-

tices like modularity, portability, CI/CD, and documentation. Now anyone on the data team can safely contribute to production-grade data pipelines.

Once all your raw data has landed in your data warehouse, you can use SQL to clean up and transform the data into clean tables/views and build unified customer profiles to productionize downstream. Critically this enables organizations to use the same data models for both analytics and activation, ensuring a true source of truth for customer identity.

Benefits of using your data transformation vs CDP's identity resolution

Flexibility — With the power of SQL and relational data, you have the full flexibility to represent anything you want within the warehouse. You can build the most accurate and robust user definition and represent any other entity — accounts, workspaces, products — specific to your business and its relation to your customers.

Completeness — As the single source of truth, the data warehouse is the most complete source of data on your customer from all sources, including POS systems, data science models, and even offline data.

Ownership — You own the models and the data. You can change these models whenever you want and distribute them at will without being limited to the opaque tooling within a CDP.

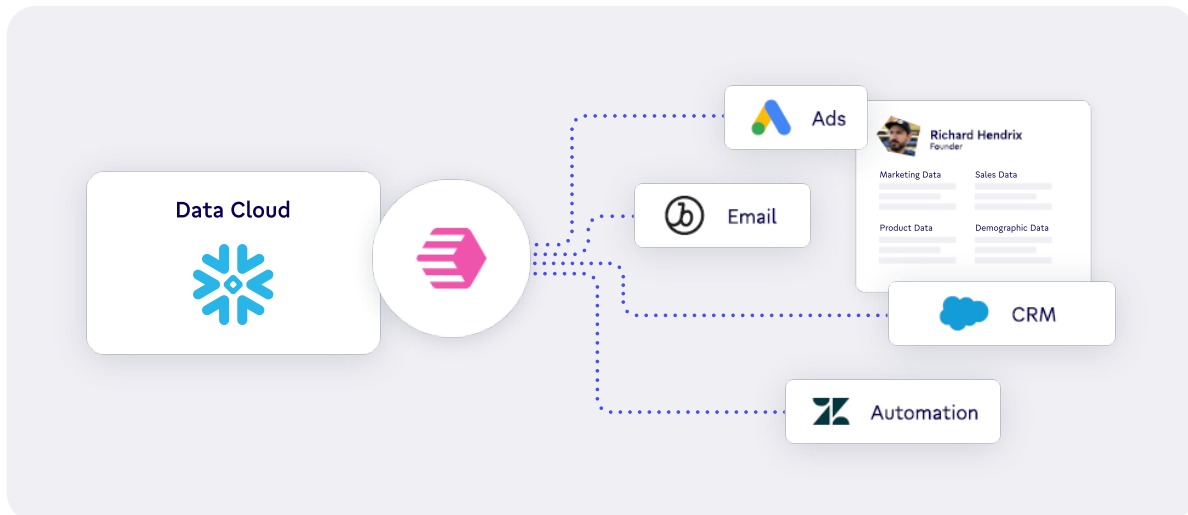
Open source identity resolution solutions

If your organization doesn't have the resources to do all the required mod-

eling for identity resolution in-house, or if you have more advanced needs, there are now third-party identity resolution services popping up that focus on merging identities in a warehouse leveraging ML technology. One such service is [Zingg](#) (here's a post about using Zingg within [Snowflake](#)). This is very much an emerging space and we expect to see more options in the coming years.

Data Activation

Data activation is the missing link needed to connect the data collection and transformation efforts of the data team and make that data accessible to marketing teams to activate across all their channels. Data activation tools like Census enable marketers to unlock data directly from the warehouse – all without needing to know SQL – while still enabling data teams to maintain governance and control that comes from being built on the data warehouse.



Census — the leading data activation platform



Census enables marketers to:

- **Sync customer attributes and lists** from their data warehouse into “systems of action” such as Salesforce Marketing Cloud, Marketo, and Hubspot, and update them in real-time as the data is transformed. This pattern is known as [reverse ETL](#).
- **Build dynamic customer (and other entity) segments** by filtering across trusted data models from the data warehouse and [syncing them to all their advertising and marketing automation platforms](#) from a single place.

Just take it from [Canva](#). By using Census to activate their data, they sync millions of customer records every day from Snowflake to their marketing tools like Braze, Facebook Ads and Google Ads. Now, they can leverage the power of their Data Cloud to send personalized marketing messages, and target the right audiences.

In conjunction, data teams can:

- **Automate [reverse ETL pipelines](#)** to sync data to downstream tools used by business teams.
- **Define approved entities (or business objects)** for marketing teams to activate. For example, they can define the underlying approved “users” and “events” tables that marketing teams can use for audience building.
- **Maintain [full governance and control](#)** of all customer data leaving the warehouse with robust logging, [observability](#), and access control features.

“A data analyst looks at data, looks at dashboards, finds insights, and is then able to activate it using Census all in one step. It’s really, really powerful and I think we’re going to get a lot of use out of it.”

Cuong Duong, *Marketing Analytics*, Canva

The screenshot shows the 'Definition' tab of a segment named 'Active High CLV Black Friday Target List'. The segment is based on 'Rewards Program Users'. The estimated size is 102,445 (18% of the base). The segmentation logic is as follows:

- Records which meet **All** of the following conditions...
 - and** Have **Properties** that match **All** of the following conditions...
 - and** email is not blank
 - and** Performed **App Events**
 - with Event named **product_view**
 - occurring **at least** 5 times **between** 30 days ago and 90 days ago
 - and** Have **Orders** that match **All** of the following conditions...
 - order_total greater than 1000

Buttons at the top: View SQL, Preview Results, Save Segment, Cancel.

Census Segments enables marketers to self-serve data with powerful visual segmentation on top of the data warehouse

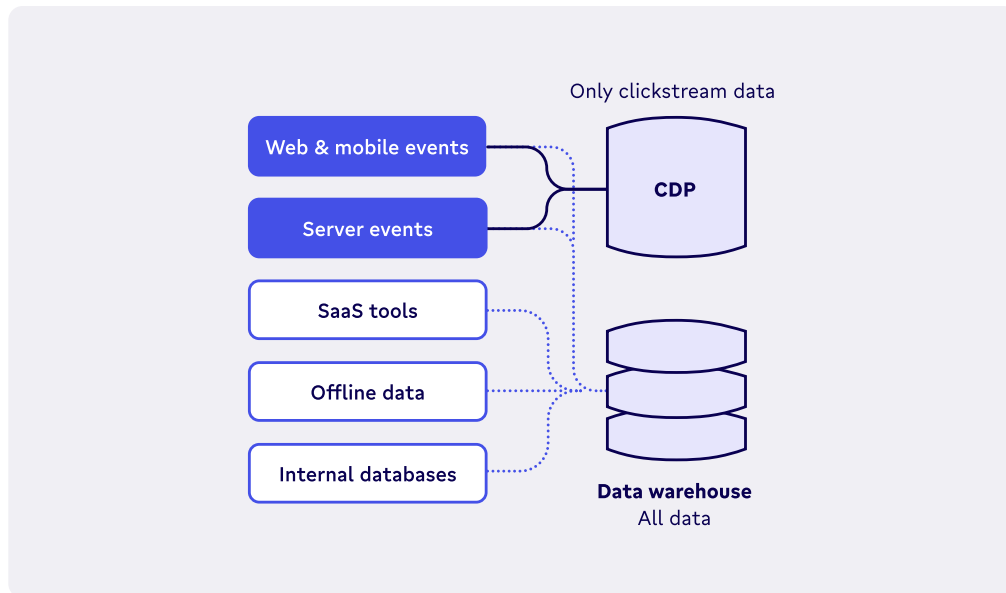
TL;DR: Data activation tools bridge the gap between marketing teams and data teams and fundamentally change the way data and marketing teams collaborate.

Old way	New way
 Marketing Mark 5:08 PM Can I have a list of all users who purchased X last year with lifetime spend > \$500?	 Data Daisy 5:46 PM Our orders data for the X SKU is ready - you can start segmenting in Census now!
 Data Daisy 5:12 PM Sorry, you'll have to add your request to the backlog. Likely we'll pick it up next quarter 😞	 Marketing Mark 5:48 PM Awesome! Our team was able to build those new audiences for google and fb during the meeting! 🥳

Benefits of the Composable CDP

A true single source of truth for customer data ★

While legacy CDPs on the market only give you a partial copy of your customer data, the composable CDP is built on top of your data warehouse to provide you with a complete view of ALL your customer data. With this 360-degree view, organizations can easily leverage data from all sources, including POS systems, data science models, and even offline data to get a comprehensive understanding of their customers.



Faster Time-to-Value

CDPs are expensive. Like really, really expensive. This cost might be justified since it's an all-in-one solution, but if you're dropping a big chunk of change on a CDP, you need to be able to justify the value. And when you buy a legacy CDP, you have to think about everything before making your purchase because once you're locked into the platform, it's game over.

On top of a long and difficult onboarding process, you won't be able to run any campaigns for months. However, with Census, organizations can connect to their warehouse and activate existing customer data so they can launch campaigns in minutes (literally) instead of waiting months.

More flexibility

CDPs force customer data to conform to a rigid structure. This might appear efficient at first, but businesses come in all shapes and sizes, meaning they require much more flexibility than what these CDPs can offer.

The composable CDP is built on the data warehouse, which means it's built on top of trusted models and transformation maintained by the data team using a tool like dbt. This flexibility allows you to represent relationships between users and entities, so your business can have complete control over how you unify customer identity.

Better data governance

Rather than an off-the-shelf CDP managing all of your customer data, the composable CDP is built on top of your data warehouse so you get to leverage all your existing data governance, security, and observability protocols for managing your customer data. This is particularly important with growing privacy legislation like GDPR and CCPA mandating the customer's right to be forgotten. It's never been more important to know exactly where all your customer data is going, and have control over all of it.

Future-proof by design

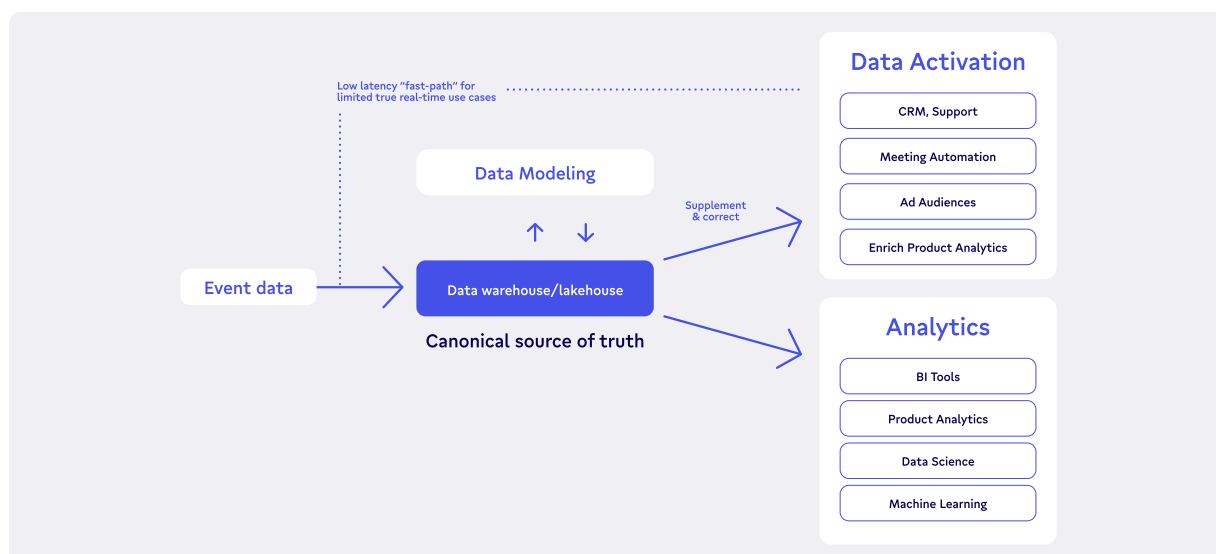
Composable CDPs are future-proof by design and allow you to avoid vendor lock-in. Since every element in a composable CDP is modular, you can choose the best-in-class tools that fit the requirements at that time. As the requirements of your business evolve, you can continue to invest on top of your composable CDP as opposed to implementing a new stack from scratch.

Composable CDP FAQs

What about real-time use cases?

The first question to answer is how real-time do we mean when we say “real-time”? Data warehouses continue to get more and more sophisticated, especially in relation to streaming use cases. In that sense they are near real-time. The end-to-end latency from an event being generated to reaching a destination tool can be as fast as 5 minutes. Specialized reverse ETL tools like Census support “continuous” syncs which means they ping the data warehouse every 1–2 minutes to check if there’s new data to sync over.

There are honestly very few true use cases where you need data sub 5 minutes, but for those extremely low latency use cases you can use your event collection tool to stream events directly into your destination. Here we’ve seen most companies follow a hybrid approach, where only specific events get streamed directly into marketing tools to activate in true real-time, but those same events are also streamed into the data warehouse. You should keep a close inventory of those low latency use cases, but ultimately the canonical record for those events will still be the copy in the data warehouse. This is the record that will be used for analytical purposes and also the record that can be amended down the line if necessary.



Does my data need to be modeled before I activate it?

Generally data does need some level of modeling before it's activated to ensure data quality as well as data governance. Some data teams may be concerned that this step could slow teams down as new event streams will need to be modeled before they are accessible for analytics. Although this may have been the case before, advancements in data technology have made it so new events can be productionized faster than ever (we're talking a few hours after you start collecting them):

- Event collection tools like [Snowplow](#) handle a lot of the structuring and flattening of events into warehouse readable schemas, which means the data already lands in a readable format in the data warehouse.
- Data teams have adopted software engineering practices like testing and version control to streamline productionizing data
- Data warehouses natively support JSON and other complex data types

From a data governance perspective, it's actually beneficial for the data team to have data producers model their data before it's published, since they understand the structure of their data the best. This follows the decentralized data ownership pattern outlined by Zhamak Dehghani's [data mesh](#) approach where decentralized teams adhere to a set of standards around format, metadata etc when publishing data and enables the organization to more easily comply with centralized governance. Remember that time spent modeling data now before publishing saves 10x time in technical debt later on.

Do I need an event collection tool before I buy data activation?

If you have no event collection set up at all, we do recommend that you first choose the event collection solution that is best for your needs before looking at data activation.

That being said, many organizations already have some level of event collection in place, and may be looking at replacing or upgrading their existing event collection solution. At this point it's important to ask yourself, what is the biggest challenge my organization is facing today? If you already have some level of event collection in place today, don't fall into the trap of "data perfectionism" — believing that you need to collect more data or model it to perfection before you start activating it.

The reality is that [73% of company data isn't being used today](#) and until you start activating data in your business, you want to have the necessary learnings on what more you actually need to do. Data activation unlocks more value from your existing data and creates an important feedback loop for the data team. This new value and learnings can then be reinvested to make more informed decisions on upgrading your data stack.

Get started with the Composable CDP today

Every company eventually needs to invest in some sort of "customer data platform." But rather than buying an expensive all-in-one solution, why not leverage the building blocks for a composable CDP built on your existing data warehouse? That way, you can start delivering value today and avoid yet another copy of your data.

There's no denying that when it comes to the future of the CDP, the tailwinds are heading in the direction of the data warehouse. CDP vendors like Twilio Segment and mParticle have recognized this by recently announcing their own versions of "reverse ETL connectors," but their rushed attempts to "slap on" reverse ETL are clearly an afterthought.

The rise of the data warehouse as the single source of truth, paired with best-of-breed tools like Fivetran, dbt and Census takes away the need for organizations to invest in expensive all-in-one CDP solutions that don't provide time-to-value. With data activation and reverse ETL democratizing access, the data warehouse is best positioned to become the system of customer record that powers not just your marketing technology, but the entire business operations of the company.

Learn more at getcensus.com.



getcensus.com