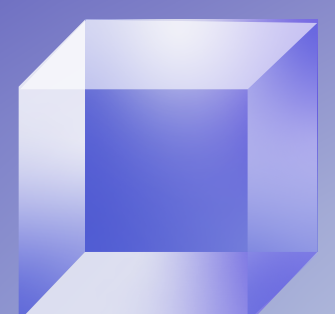
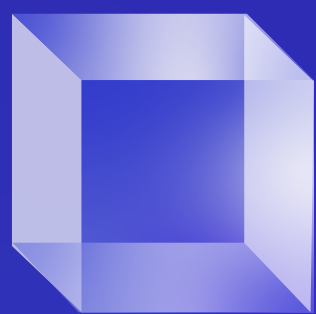




The Composable CDP:

A Future Proof Customer Data Platform

Build your CDP from best-in-breed components
and never look back.



Introduction

In the new era of a privacy-first internet, the Customer Data Platform, or CDP, has become more valuable than ever. Thanks to data privacy changes being implemented across all major mobile platforms and web browsers like Safari, Chrome, and Firefox, organizations must pivot their reliance from anonymous customer identifiers - cookies and device IDs - and implement a first-party data strategy.

A Customer Data Platform is defined by the CDP Institute as “packaged software that creates a persistent, unified customer database that is accessible to other systems.”

CDPs now sit at the top of the [priority list](#) for marketing executives and are an essential piece to any marketing technology stack - ensuring marketers can continue to identify customers and offer them best-in-class omni-channel experiences.

Ordinarily, having all your customer data collected, cleaned, and available to take action on via a single off-the-shelf platform would seem to be the faster and better solution for a business. But the reality is quite the opposite. The CDP Institute's latest [survey](#) found that just “23% of consumer marketers have completed their projects on time and schedule. Only 58% of companies with a deployed CDP say it delivers significant value.”

In contrast, Composable CDPs are faster to implement, easier to maintain, and much more easily adapted by the sales and marketing teams the CDP intends to serve. And by utilizing a cloud data platform, Composable CDPs avoid the constraints of a packaged CDP's “cookie-cutter” data models. Direct control of the data enables companies to properly represent business-specific data models like products, groups, coupons, artists, etc.



Darren Haken

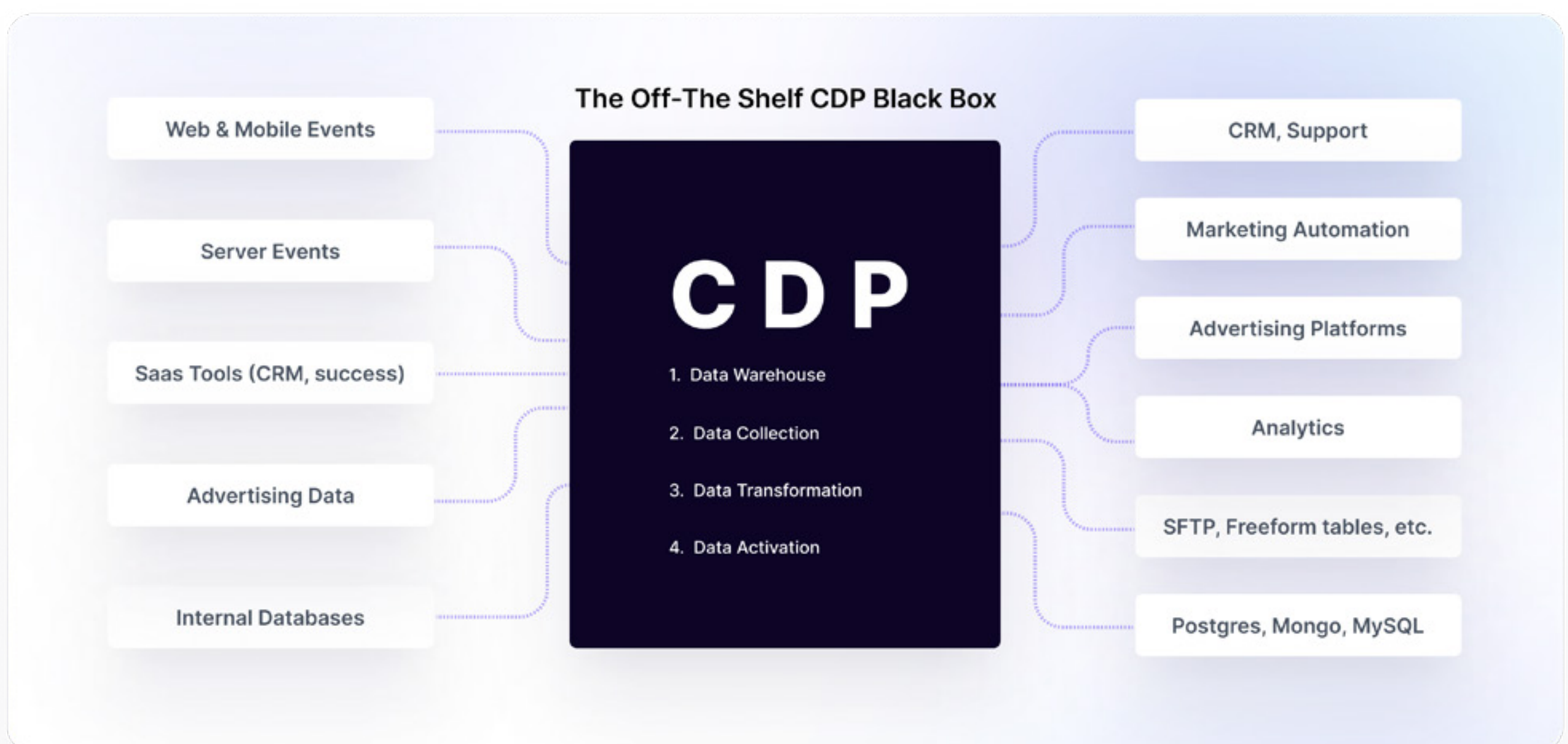
Head of Engineering - Platform
and Data at Autotrader

“We evaluated CDPs but they are expensive black boxes that are extremely technical and they don’t always play nicely with other tools.”

In this whitepaper, we will discuss what CDP’s are, uncover the pain points associated with traditional CDPs and dive-in to explain how a composable CDP can add transformational value to any modern organization.

What is an Off-The-Shelf CDP?

A CDP is an all-in-one data platform built for marketing teams. CDPs serve as a database for all your customer information with a bundled activation layer to help you leverage your data for marketing automation.



Breaking Apart The Four Components of a CDP

1 Data Warehouse

CDPs provide a proprietary repository of data that aggregates and manages different sources of customer data collected from all of the business's SaaS and internal applications. The unified data cloud is a 360 degree view about each consumer and a central source of truth for the business.

2 Data Collection

Since CDPs are a database of customer data, they need to give you a way to send data to them. To combat this, CDPs exposes an API for developers to track user traits and events that users are taking across applications.

3 Data Transformation

All CDPs usually have an out-of-the-box identity stitching functionality and tools to create custom traits on user profiles.

4 Data Activation

CDPs are useless unless they help you act on your data, so they have integrations to sync enriched profiles and audience segments to marketing channels.

All of these features are bundled together in a tightly integrated all-in-one solution. Off-the-shelf CDPs try to solve for the common use cases across their customer base. Unfortunately this one size fits all approach means those same customers struggle with adhering to the CDP's rigid data models, long onboarding times, and the data redundancy resulting in two sources of truth across analytics and marketing tools.



Weston Rowley

Director, Strategy & Analytics at Lucid

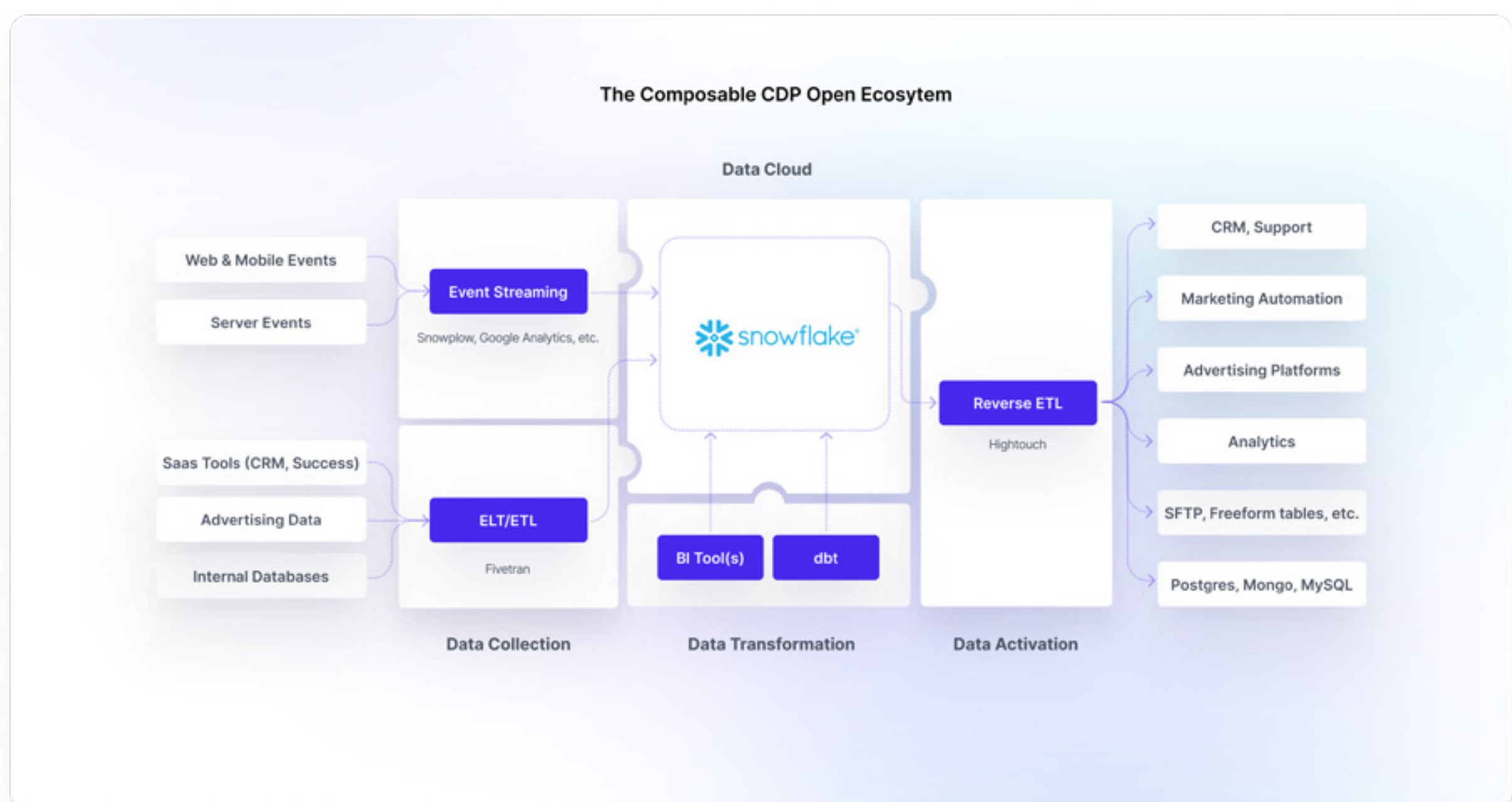
“In many ways, we developed our tech stack with a vision for the future, but we also wanted to solve the problems we have now. The problem with any out-of-the-box analytics suite is that you are restricted to only doing what it supports natively. CDPs are not extensible to what you want them to do.”

What is a Composable CDP?

A Composable CDP has the same architecture as an off-the-shelf CDP but sits on top of your organization's cloud data platform and combines it with best-in-breed components for ingestion, transformation, and activation.

Control of the data cloud provides significant advantages over the black box database provided by off-the-shelf CDPs.

Composable CDPs enable a **single source of truth** across the entire organization, **greater flexibility**, **improved data security**, and **greatly reduced launch times**.



A Composable CDP is Assembled from Best-in-Breed Components

Comparing Composable to off-the-shelf CDPs often leads to misleading comparisons to a buy vs. build approach typical of software investments. In reality, components that make up a Composable CDP are ‘purchased’ software assembled in lego-like building blocks.

Separate best-in-breed components for Data Collection, Data Transformation, and Data Activation are superior to the all-in-one counterparts baked into an off-the-shelf CDP. More importantly, separate components are more extensible and can be combined to solve problems well beyond a off-the-shelf CDP’s common use cases.



William Tsu

Customer Success Operations at Blend

“We’ve found to have more success with a purpose-built tool, versus something that does everything ‘okay’. There’s certain things you just want to do well and not have to worry about it, on the ELT side with Fivetran and then on the Reverse ETL with Hightouch.”

The Evolution of CDP

So, given all of the above: Why haven't companies simply built their CDPs on top of the data warehouse? It's easy to discard CDPs by saying they're all-in-one solutions and not "best of breed," but there's more to the story if you pull back the curtain.

The CDP category pre-dates the growth and adoption of modern cloud data platforms. First coined in 2013, the CDP grew out of tag management companies (Segment, Tealium) which captured massive amounts of click-stream events from digital touchpoints: email, e-commerce, website, and mobile experiences. The traditional relational database systems were simply not suited for managing this velocity and volume of data.

CDPs became the first to adopt modern data warehouse technologies but did this "under the hood" because their customers didn't have their warehouse ready to activate. Simply put, CDPs built their own customer data "source of truth" because there was no other option available.

Today, however, modern data platforms have firmly taken root within the business environment. In 2020, Snowflake, the company behind the Data Cloud, had the largest software IPO. Now, any business not leveraging the Data Cloud to build their single source of truth for all their business data is rapidly falling behind.

As Tejas Manohar, Founder of Hightouch and early engineer at Segment, said: "It became clear that the way customers viewed data warehouses and BI had totally changed. The warehouse was no longer just an advanced analytics tool you reached for when you couldn't get an answer in Amplitude or Mixpanel—it was the source of truth across your business. Data-mature companies no longer need the all-in-one ingestion and unification components of a CDP."



Composable CDP Components

1. Data Cloud

2. Data Transformation

- a. ELT
- b. Event

3. Data Transformation

4. Data Activation

Data Cloud

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 a best of breed component of the composable CDP, offering organizations a single platform for all their data, instant access to governed data, and the ability to seamlessly act on their data to extract maximum value.



The advantages of Snowflake include:

1 One Platform for All Data

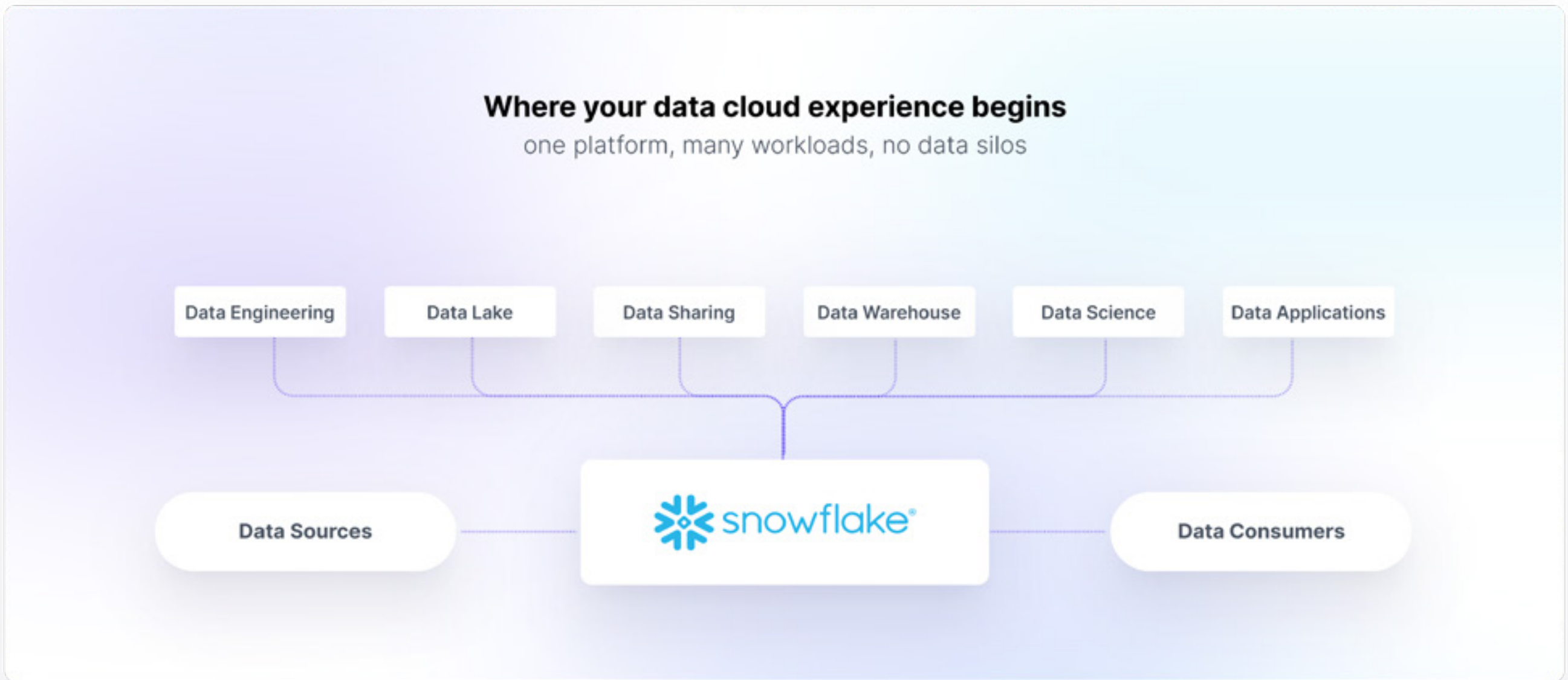
The Data Cloud is a single platform for all data from any source, in any format and for any method. It integrates both structured and semi-structured data, whether the sources are internal or external, and does so for any type of data method (batch, stream, or shared). Having one platform that integrates all then organization's marketing data holistically is a prerequisite for laying the foundation for data-driven marketing.

2 Instant Governed Access

The Data Cloud mitigates data privacy risks by limiting the creation and movement of disparate data copies as well as through built-in security and access controls. Moreover, the Data Cloud ensures organizations are able to govern their data with reduced latency with data sharing and offers near real-time access to updated datasets. With its purpose-built-for-cloud architecture, Snowflake ensures that the Data Cloud is able to support all data users and applications simultaneously, significantly streamlining data flows and increasing efficiency.

3 Seamless Activation

No data-driven marketing is possible without activation of the data itself, and the Data Cloud is uniquely positioned to help organizations extract value by seamlessly enabling access of data for activation and analytics. By opting for the Data Cloud, organizations can integrate with a broad ecosystem of best-of-breed partners across CDPs, engagement engines, marketing clouds, and more. Furthermore, organizations can scale up, or down, and only pay for what they use, providing the needed flexibility and scalability to cope with peak intervals.



One of the workloads in the Snowflake Data Cloud is the data warehouse, which is a prerequisite to a Composable CDP.

Working with the data warehouse capability within the Data Cloud provides significant advantages over using a data warehouse bundled in a CDP, including:

- 1. 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.

2. All Business Data – 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.
- 3. Privacy and Security** — Off-the-shelf CDPs replicate customer data into their data warehouse, creating privacy and security risks for the business. As a result, most CDPs are not HIPAA compliant.

4. 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.

5. Redundant Costs — Because off-the-shelf CDPs replicate data already in the warehouse, companies pay twice for data warehouse services.



Ryan Newsome
Head of Data at Compare Club

“We didn’t choose an off-the-shelf CDP because a substantial amount of engineering is required to get it up and running, and at the end of the day, it is just a second source of truth. Why would we buy a CDP when all of our data modeling is already being done directly in Snowflake?”

Data Collection

Packaged CDPs provided two types of data collection services:

1 ELT connectors, which extract data from SaaS tools and databases into the Data Cloud.

2 Event connectors, which capture click events that occur within digital merchandise stores, and mobile applications.

In the era of the modern data stack, data collection is handled by companies that specialize in each of these areas.

Extract, Load, Transform [ELT]

ELT, or Extract, Load, Transform, is the modern approach to ingesting data from any application database, or data source within your organization. The speed and reliability of the Data Cloud enables transformation to happen within the data warehouse and so have replaced the traditional ETL approach.

Benefits for Dedicated ELT Connectors

- CDP vendors don't have the same dedicated resources to maintain this volume of ELT connectors.
- In many cases, your data engineering team has already licensed and implemented ELT connectors like Fivetran's to power their business reporting needs.



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.



Adam Smith

Analytics Manager at Imperfect Foods

“Fivetran’s automated data connectors are prebuilt and preconfigured. After the initial setup, you can quickly and reliably transfer data from the source to Snowflake in minutes”

Fivetran vs. CDPs

Fivetran and CDPs were created to address different issues, leading to several key differences, including:

- Fivetran’s 200+ prebuilt connectors are optimized for both SaaS apps and data warehouses and support a broad variety of databases. These integrations are purpose-built to normalize the data so that it’s query-ready for your analyst team to work with.
- Fivetran centralizes data from hundreds of SaaS and on-prem data sources into cloud destinations. All Fivetran connectors (where data exists) will ingest all historical data available in the data source.
- Fivetran connectors typically come with a prebuilt entity relationship diagram for easier immediate report building.
- Fivetran supports data sources spanning all departments and use cases, whereas CDP prebuilt integrations tend to focus solely on marketing and sales.
- Fivetran replicates but does not own data during ingestion to your data warehouse, alleviating the aforementioned privacy and security concerns.
- Fivetran syncs data as frequently as every five minutes, allowing you to act on the latest data.

Snowflake + Fivetran

- Fivetran is embedded within the Snowflake product and you can kickstart a trial of Fivetran through Partner Connect.
- The stack offers:
 - Data accessibility with scale - regardless of data source
 - Automated, zero maintenance data pipelines
 - Secure data replication
- Over 2,000 joint customers leverage the stack to make every decision data-driven with reliable, high-volume data integration that scales with the Data Cloud.

Event Collectors

As the name suggests, event collectors track your users and their associated click events when browsing your website, e-commerce store, app, or any other digital experience.

CDPs, Analytics applications, and Ad platforms all use event collection to power their respective products. Until recently, implementation was easy: just insert JavaScript into your digital experience and start collecting data.

Browser and OS changes are upending how event collectors work in today's privacy-sensitive landscape, with significant implications for both analytics and CDP tools:

- Blocking third-party cookies makes it impossible to join anonymous users across domains.
- Ad blockers and cookie consent pop-ups block client-side collection services from even firing, resulting in data loss.

CDPs developed their own event collection tools to:

- Stream data directly to their application database.
- Overcome data limitations of reporting tools like Google Analytics which was designed for a time when tracking page views, instead of individual click events within a page, was the norm.
- Structure user identifiers (IP, Device) to join user identifiers downstream in the data warehouse.

The New Generation of Analytics as Event Collectors

Privacy constraints and the rise of the modern data warehouse are re-shaping the offerings from existing event collection applications.

Google Analytics 4 [GA4], introduced in 2020, is a wholly rewritten architecture to support user event tracking. GA4 provides tight integration with BigQuery, Google's data warehouse offering, so all data captured can be sent to your data warehouse for enrichment and data modeling. Google recommends deploying GA4 server-side so tracking fires independent of ad blockers.

In a similar vein, Mixpanel, Amplitude, and Heap have all upgraded their offerings to address these same challenges: All three services now support server-side deployments with the ability to sync data to your data warehouse.

Businesses that have invested in these analytics products would do well to leverage their data warehouse integrations rather than start with a standalone CDP's event collection tools.

Event Tracking for the Data Cloud

Snowplow and Snowcat are a new category of event collectors designed for the modern data warehouse. Unlike CDP and analytics-based event collectors, these new tools only send data to your data warehouse; there is no upstream data warehouse with a copy of customer data outside your business's control.

The architecture of these solutions has numerous benefits in data ownership and quality over their analytics and CDP counterparts, including and especially:

1. Data Ownership—Data is sent directly to your data warehouse, thereby avoiding privacy and data residency issues.

2. Data Quality—Specialization in event collection translates to more and better quality data than event collectors designed for a more specific purpose like CDP or Analytics. A simple example is that Snowplow collects 120+ user properties, whereas the Segment CDP only captures 23 user properties.

For businesses who have outgrown their CDP and want more control of their data, [Snowplow](#) and [Snowcat](#) are excellent options.

Data Transformation and Identity Resolution

Data transformation is the process of joining different data sets into meaningful and useful groupings (tables). CDPs provide a specific form of data transformation known as identity resolution. Identity resolution stitches together a unified customer id from all the different systems ingested into the data cloud.

Identity resolution is a key service — one could say the key service — off-the-shelf CDPs provide as part of their packaged offering. To provide identity resolution as a service, CDPs must make assumptions about the user definition — user identifiers and attributes — that power their identity resolution.

Off-The-Shelf CDP Identity Resolution Limitations

In practice, most businesses struggle with the constraints imposed by a CDP's one-size-fits-all customer definition. For example, in a B2C context, a customer is an individual purchaser, whereas, in a B2B context, customers need to be associated with their company.

Combine prescriptive user definitions with variations of a business's model (ie, Is it a physical good, digital subscription, consumption-based, or one-time purchase?) and you will begin to understand the complexity involved in identity resolution.

Composable CDP Data Transformation Benefits

As companies begin to leverage the data cloud as the central hub for all customer data, we see a rapid shift towards companies **preferring** to do identity resolution directly in Snowflake. Compared to legacy CDP solutions, the tooling available to transform data within the Data Cloud is more flexible and robust. It empowers data teams to align their analytics with operationalizing and distributing data to the rest of the business. **This data modeling task is one of the few things that data teams should own in-house**, as clean modeling will inevitably become specific to each company. Accurate business-specific data models are the engine that powers all analytics, decision making, and orchestration that sits on top.

- **Completely Flexible** — With the power of SQL and relational data, you have the full flexibility to represent anything you want within Data Cloud. 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.
- **Complete** — As the single source of truth, the Data Cloud will contain all the other data (outside of only clickstream events) from other data pipelines, including from internal databases and your other SaaS tools. When you build a customer profile from the Data Cloud, it will contain all the critical insights from these data sources.
- **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 black box tooling within a CDP.

Recommendations and Vendors

Taking on Identity Resolution in-house can sound intimidating, and it's worth noting that for more advanced needs with **large and varied data sets**, there are third-party tools that focus on merging identities in a warehouse that leverage ML tech. If you have complex needs and don't have the resources to take this on in-house, it may be worth checking out the two options below (that we're very curious about):

- **Zingg** - Open Source solution for entity resolution that works natively on the warehouse (on the data type and model of your choice), without much preprocessing or data cleansing. Here are two recent posts on using Zingg within [Snowflake](#) and [Data-bricks](#), respectively.
- **Truelty** - Snowflake-native solution, primarily for larger B2C organizations. Truelty automatically generates the code to natively process deduplication within your Snowflake instance. Typical use cases involve cross channel, cross app, and anonymous to known identification.

In most cases, however, an in-house method using just SQL can get you very far, and our recommendation is to start simple and focus first on your highest-value data sets. We walked through a basic example of this SQL logic in our earlier blog post titled [Identity Resolution in SQL](#).

dbt, one of the hottest modeling tools in data today, simplifies this type of merging within Snowflake, as it allows analysts to use just SQL but do so in a way that empowers them to collaborate, test, and continually abstract the complexity away from the otherwise hairy dependencies that would arise.

For more on identity resolution and how to get started with dbt see our two-part guide [“Identity Resolution - Why CDPs Fall Short”](#)



Erik Edelmann

Senior Analytics Engineer at Vendr

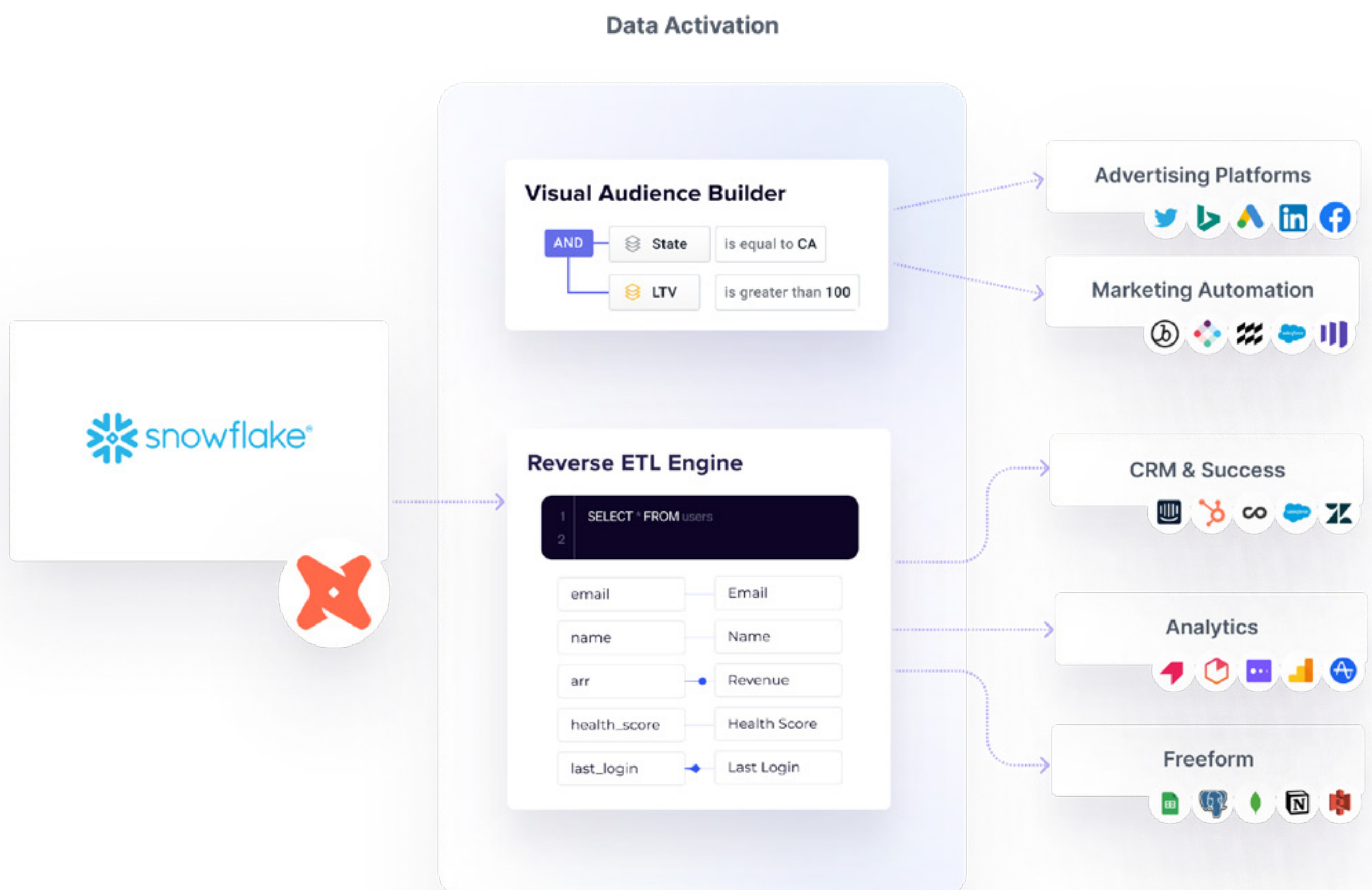
“Data is core to our business. Taking that data and putting it in the right place at the right time is key and that is exactly where Snowflake, Hightouch, and dbt help.”

Data Activation

Data Activation is a recent arrival within the modern data stack ecosystem. Prior to this, the only way to activate data from the warehouse was through internal development to build and maintain connectors for business and marketing systems. Companies adopting a Composable CDP approach is a direct outgrowth of the recent availability of Data Activation as a standalone service.

Like a mirror image of ELT providers such as Fivetran, Data Activation companies provide an extensive library of pre-built connectors **from** any data warehouse, such as the Snowflake Data Cloud, **to** an ever-growing library of marketing and business applications.

Reverse ETL, a popular way to describe this architecture, is the process of copying data from a central data warehouse to operational systems of record, including but not limited to SaaS tools used for growth, marketing, sales, and support.



Segmentation, Audience Building, and Syncing

During Data Activation, marketing or data teams define what data they want to sync and how often.

In the CDP context, the data definition is often an audience segment derived from profile attributes of the customer or actions they have taken: Location, Recent Sign-ups, Purchases, or Lifetime Value.

Outside the CDP context, businesses use Data Activation to sync other data types, such as Products, Content, and Reports.

The Data Activation platform then syncs this data with SaaS applications at whatever intervals meet the business needs, which is typically how often the underlying data changes —every minute, hour, or day.

Benefits of a Data Activation Platform

Data Activation Platforms are a first-class technology within the modern data stack. Unlike CDPs, which belong only to marketing teams, Data Activation platforms work across the entire business — including marketing — activating data anywhere needed. As a result, Data Activation platforms must support many more destinations and user interfaces than a CDP provides.

More Integrations — Data Activation platforms support marketing, sales, reporting, ERP, and collaboration applications. Hightouch, a leading Data Activation platform, now has over 100 integrated destinations across all the business categories.

Data Friendly — BI and engineering teams can use SQL and dbt to provision and sync data from warehouse to business application.

Marketer Friendly – No-code interfaces allow marketers to easily build and sync audiences from the data warehouse to their business applications.

hightouch

Hightouch is the world’s leading Data Activation platform, syncing data from warehouses directly into your SaaS tools. All business teams, from sales and marketing to support and customer success, need relevant, accurate, and real-time customer data to add critical context inside the software they already use. Whether you’re enhancing communications with customers via CRM, optimizing ad copy, or personalizing email, Hightouch makes your data actionable —all with SQL, no scripts or APIs required.



Adam Smith

Analytics Manager at Imperfect Foods

“Hightouch has made it really easy to sync our data in Snowflake to our various marketing platforms so we can optimize around our most successful customers. With Hightouch we can duplicate our audiences and send data to any destination of our choosing immediately.”

The Six Data Activation Frameworks at Hightouch

Hightouch provides six activation frameworks to enable the broad range of use cases their customers require.

1. Objects / Reverse ETL:

- **Example Use Case:** Sync product engagement data from the warehouse into Salesforce. Sales needs to see how much a customer has consumed and which features they've interacted with so that they can sell the right features and limits.
- **Common Destinations:** Sales CRMs (**Salesforce, Hubspot**), marketing CRMs (Braze, Iterable), Intercom, ERPs and finance tools (**Netsuite and Anaplan**), and 70+ more. This also includes database destinations like Postgres and MongoDB (for in-app personalization).

2. Event Syncing:

- **Example Use Case:** Sync purchase event data to Braze so that marketers can automatically enroll buyers of specific products in relevant email campaigns, based on the color and category of the purchased product.
- **Common Destinations:** **Amplitude, Mixpanel, Braze events**, and **Iterable events**.

3. Hightouch Notify:

- **Example Use Case:** Any time a customer workspace surpasses the limit of the free tier, send a message in Slack tagging the relevant salesperson and notifying them that they should reach out and ask the workspace to pay (PQLs).
- **Common Destinations:** **Slack, Mattermost**, and **Microsoft Teams**.

4. Hightouch Audiences:

- **Example Use Case:** When someone fills out a coupon form and enters their email address, upload their email to a Facebook, Google, and Tik Tok audience so that they receive ads about products they're interested in. When they purchase a product, remove them from all three audiences.
- **Common Destinations:** **Facebook Audiences, Google Ads, Bing, Tik Tok, Snapchat, Mailchimp, Sendgrid, SFMC**.

5. File Uploads:

- **Example:** Clients run ads on an advertising platform. The identities of clicking users are stored in the advertising platform's data warehouse and shared with the clients via SFTP.
- **Common Destinations:** **SFTP, S3**.

6. Tasks:

- **Example Use Case:** Query workspaces that just graduated from four users to five users (PQL). Automatically create a task in the helpdesk tool (Jira or Zendesk) so that a support rep can reach out to the workspace and ask if they need help setting up any premium features.
- **Common Destinations:** **Asana, Jira, Zendesk, Salesforce**.

Customer Case Studies

Each customer of Hightouch uses the product in a different way, and the majority eventually graduate to doing more than Reverse ETL:

- **Vendr** automates customer touchpoints
- **Compare Club** reduced cost per lead by 10%
- **Retool** increased email response rates by 32%
- **Circle CI** activates dbt models to measure ROI
- **Blend** reduced reporting time by 50%
- **Nando's** improves customer loyalty & reduces data integration time by 75%
- **Lucid** increased ROAS by 52%, new users by 37%

S U M M A R Y

Getting started with a Composable CDP

As we’ve seen, implementing a packaged CDP can be pretty challenging, especially at an enterprise level. Because CDPs revolve around their customer database, your product engineering team must implement data collection by tracking user traits and events across various websites, backend services, and apps via your CDP’s APIs and SDKs. Implementation can often take 3-6 months before the marketing and data activation can begin.

The Composable CDP allows you to solve the most important problem in front of you incrementally. This will enable you to choose the best solution and components for your business. You can educate yourself throughout the process and future-proof and swap out specific components down the line when your needs change or when a particular tool isn’t “cutting it.”

Since all companies are already investing in their data warehouse for analytics, you can start activating your data in hours or days with a Composable CDP instead of months or quarters with a typical off-the-shelf CDP.

Authors



Lee Hammond
Product Evangelist
Hightouch



Lourenco Mello
Product Marketing
Snowflake



Amit Kapadia
Product Marketing Manager
Fivetran

