

The Composable Customer Data Platform: What It Is and Why You Need One

How modern organizations supercharge marketing efforts with a composable CDP



This is a collaborative whitepaper among Databricks, Hightouch and Snowplow. We thank Nick King (Chief Product and Marketing Officer at Snowplow) and Alec Haase (product evangelist at Hightouch) for their contributions.

In the current era of a privacy-first internet, first-party customer data has never been more valuable to digital organizations. In an effort to effectively capture, manage and activate customer data, the customer data platform, or “CDP,” was born. Gartner defines a CDP as “software that collects and unifies customer data — from multiple sources including first- and third-party — to build a single, coherent and complete view of each customer.”

Here’s the bad news. “CDPs are both one of the most disappointing investments brands made in 2020 and the number one investment they want to make in 2021,” explained a [blog](#) from Databricks and Amazon Web Services citing Gartner numbers.

Simply put, when it comes to rationalizing and putting customer data to work, there is a lot of hype around CDPs, and an alarming rate of failure.



The customer data problem

Marketing teams live and breathe customer data, but they don’t always trust it. “Nearly 80% of marketing leaders say data, analytics and insights are very important to winning and retaining customers,” according to a recent pulse poll of select CMO Council marketing leaders. Every modern marketer wants direct access to reliable customer data, seeking to gain an advantage over their competition, yet in reality, “nearly two-thirds of all marketers are only moderately confident (or worse) in their data, analytics and insights systems,” as reported by the [CMO Council](#).

While most marketers are confused by their data, they are also [baffled](#) by the range of customer data solutions available to them. Many who try to harness customer data ultimately fail, especially as use cases expand. “There is a high risk of failure for customer data projects. Organizations increasingly expect customer data platforms to support customer experience (CX) use cases that fall outside of marketing’s direct control, such as the next best action for customer service or customer journey insights. These use cases depend on, or overlap with, capabilities in IT systems such as master data management or cloud-based data warehouses,” [Gartner](#) stated.

Those that solve the customer data problem have a massive competitive advantage over those who don't.

- **65%** of business leaders say their goal is to build a cohesive data ecosystem and to standardize data collection. ([Treasure Data](#))
- Only **14%** of enterprises have a 360-degree view of customer data. ([Gartner](#))
- **Two-thirds** of marketers lack confidence in their data. ([Gartner](#))

But

- Of the 14% with a 360-degree view, **44%** say they achieved a 360-degree view via a customer data platform. ([Gartner](#))
- **75%** of CDP customers expect 5x or higher ROI, and most see a positive ROI in the first year of adoption. ([Twilio](#))
- Two top-reported CDP benefits include a unified customer view (**88%**) and analytics (**54%**). ([CDP Institute](#))

Embracing first-party customer data

One of the greatest assets of the modern digital organization is first-party customer data. The rapid rise of the privacy-centric consumer has led to a monumental shift away from third-party tracking methods. Organizations are now scrambling to implement a data infrastructure that, leveraging first-party data, can enable the personalized experiences that customers expect with every interaction.

Those that want to engage customers effectively must build actionable intelligence on top of their first-party customer data. Actionable intelligence means improving customer relationships, building consumer trust to capture rich first-party data, and using that data to build intelligence that can optimize the customer experience.

Historically, building a rich, behavioral dataset of customer interest and intent was difficult. Using that data to make meaningful inferences and predictions about individual customers at scale was even more challenging. And activating that intelligence across a myriad of customer touchpoints and marketing channels seemed almost unachievable.

CDPs: Beyond departments to the C-suite

Many organizations have turned to CDPs to help overcome these significant challenges. While CDPs offer a traditional solution to collecting, cleaning and activating customer data, adopting organizations have struggled with their rigid data models, long onboarding times and data redundancy across analytics and marketing tools. The CDP Institute's latest [survey](#) found that only "58% of companies with a deployed CDP say it delivers significant value," leaving much to be desired from these packaged CDP solutions.

Yet CDPs are essential for managing customer data and powering meaningful customer engagement initiatives. They are too vital a solution to be bought as shadow IT or kept under the purview of a single department. While a CDP may be first leveraged by marketing, customer data is so vital that CDPs expand to more and more use cases, becoming both fundamental and strategic. A CDP done right is immensely valuable; when done wrong, it results in frustration and money down the drain. To avoid this, various C-level execs from IT to finance should be invested in implementing a CDP as an element of their broader data management strategy.



C-level execs from IT to finance to data should be invested in implementing a CDP as an element of their broader data management strategy.

The good news is that top execs are taking customer data seriously. According to the [CX Experience Report 2022](#), “65% of business leaders say their goal is to build a cohesive data ecosystem and to standardize data collection.”

With the rise of the modern data stack, platforms like Databricks, Hightouch and Snowplow have emerged to offer future-proof approaches to behavioral data creation, data storage and management, personalization and digital

experiences. These platforms combine to form a best-in-breed alternative to traditional CDPs: the composable CDP, a category-leading solution for modern organizations that want to better manage their first-party data strategies and leverage their existing investment in the cloud data warehouse.

This whitepaper will discuss traditional CDPs, the rapid adoption of composable CDPs, and how the latter can add transformational value to any modern organization.



KEY TAKEAWAYS

- The team with the best left-to-right view of their customer data can outmaneuver competition more effectively.
- Changes in privacy rules make an organization’s ability to understand and take action on its first-party data more important than ever. Embrace it for competitive and customer advantage.
- There are many ways the implementation of a CDP can go wrong. Often, the purchase decision is made independently on a departmental level, usually marketing, without C-suite or cross-organizational involvement.
- The C-suite should be active in the scoping and acquisition process for a CDP, demanding a customer data solution that is simple, flexible and extensible.
- An organization’s CDP should be viewed as an element of its broader data management strategy.

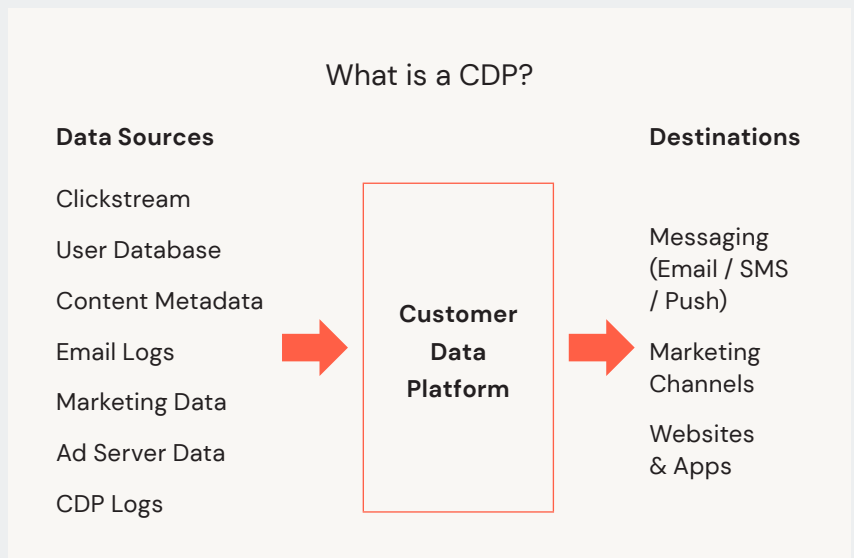


What is a customer data platform?

The definition of a customer data platform has evolved numerous times since its inception in 2013. Gartner currently [defines](#) CDPs as:

A software application that supports marketing and customer experience use cases by unifying a company’s customer data from marketing and other channels. CDPs optimize the timing and targeting of messages, offers and customer engagement activities, and enable the analysis of individual-level customer behavior over time.

Traditionally, digital marketing professionals apply a number of tools to create, implement and measure their multichannel campaigns. However, these tools are too often siloed and



The components of a standard CDP offering:

Data collection

CDPs are designed to collect customer events from a number of different sources (onsite, mobile applications and server-side) and append these activities to the customer profile. These events typically contain metadata to provide detailed context about the customer's specific digital interactions. Event collection is typically designed to support marketing use cases such as marketing automation.

Data storage and modeling

CDPs provide a proprietary repository of data that aggregates and manages different sources of customer data collected from most of the business's SaaS and internal applications. The unified database is a 360-degree view of each customer and a central source of truth for the business. Most CDPs have out-of-the-box identity stitching functionality and tools to create custom traits on user profiles.

Data activation

CDPs offer the ability to build audience segments leveraging the data available in the platform. Thanks to a wide array of prebuilt integrations, these audiences and other customer data points can then be pushed both to and from various marketing channels.

do not offer a holistic approach to customer data. "CDPs promise to solve this by centralizing data collection and unifying customer data from disparate sources into profiles," according to Gartner. "CDPs then create and manage segments and push those segments into priority channels. The CDP is not a substitute for an enterprise's database of record, but it can ensure that customer profile data, transactional events and analytic attributes are available to marketing when needed for real-time interactions."

CDPs offer major benefits. "The promise is extremely valuable – to understand your customers and communicate with them in a consistent way to provide good offers," said Dan Morris, Senior Director of Industry Solutions – Communications, Media & Entertainment at Databricks.

Three routes to CDP – pluses and minuses

There are three main paths to a CDP:

DIY: Building a CDP may be a great solution for an organization if it has a fully resourced development team and its entire data team is highly involved. Note that it's critical to factor in the ongoing support and maintenance required for custom-built CDP solutions over time. Before embarking on a custom solution, consider answers to questions such as: What is our staffing plan when key developers leave the company? What is our CDP product roadmap? Is our CDP extensible, able to add innovative new functionality that may have just come to market?

Traditional CDP: Traditional CDPs, also called fully integrated CDPs, are better for some organizations than others. One benefit is the "one vendor, one solution" scenario. If it can check off most of the organization's boxes, a traditional CDP might be a good fit. However, the cons for data teams include rigid data models, long evaluation and onboarding times, and multiple data silos. In addition, the cons for marketing/operations are vendor lock-in and limited extensibility. If any of these raise concerns, a traditional CDP may not be a good fit.

Composable CDP: Also known as a modular CDP, this approach stitches together best-of-breed components into an extension of the organization's broader data management strategy. With a composable CDP, teams don't need to start from scratch. Any work done acquiring data solutions and harmonizing data can be pulled forward.

Composable CDPs are component-oriented, making them more agile, flexible and future-proof. When done well, a composable CDP can deliver greater functionality by choosing the solutions best suited for their role in the composable stack. As the organization grows and use cases evolve, teams have the flexibility to replace the piece of the stack without affecting the whole system. Specific team members (like the data team) can then work within the context (and tools) they know and are confident using.

Organizational alignment with a data-first composable CDP, in conjunction with a strong data mapping strategy, ultimately helps an organization expand to more use cases faster.



Why traditional CDPs may not be the right fit

Most companies struggle with their CDP implementation when they think of the CDP as a “system that’s off to the side,” effectively disconnecting it from the rest of an organization’s data stack. This approach is inevitably not aligned with the existing data management strategy and leads to countless issues and concerns for IT, data engineers, analysts/domain experts and data scientists, all of whom should ultimately drive the CDP’s success.

Beware of shadow IT causing implementation and data group issues

A common challenge with traditional CDPs occurs when marketing makes these purchases in isolation. Often, this results in the marketing team struggling with implementation complexity, reaching out to IT months later with little to show.

“Often a CDP vendor goes straight to the marketing team, convincing them they can stand up the CDP solution, that they don’t need IT support, and bam, they’re good to go,” Databricks’ Morris said. “Turns out that six months later, it’s a little bit more than that. Then IT needs to get involved. And it becomes more challenging. IT ends up living with the sins of marketers past.”

Another challenge that organizations face with their CDP implementations is that traditional solutions often promote the idea of sending data directly to the CDP rather than the organization’s data warehouse. While this may work for the marketing team and their needs, it’s not optimal for the organization and its overall data strategies. Data teams end up challenged with using tools native to the CDP and managing audience segments in multiple places with multiple “sources of truth.” Data extracted from traditional CDPs to the data lake lacks granularity, quality, explainability and completeness, and needs significant wrangling before it can be used for AI-powered data products. Case in point: According to [research](#) published in 2022 by Snowplow, 91% of data practitioners face challenges in delivering

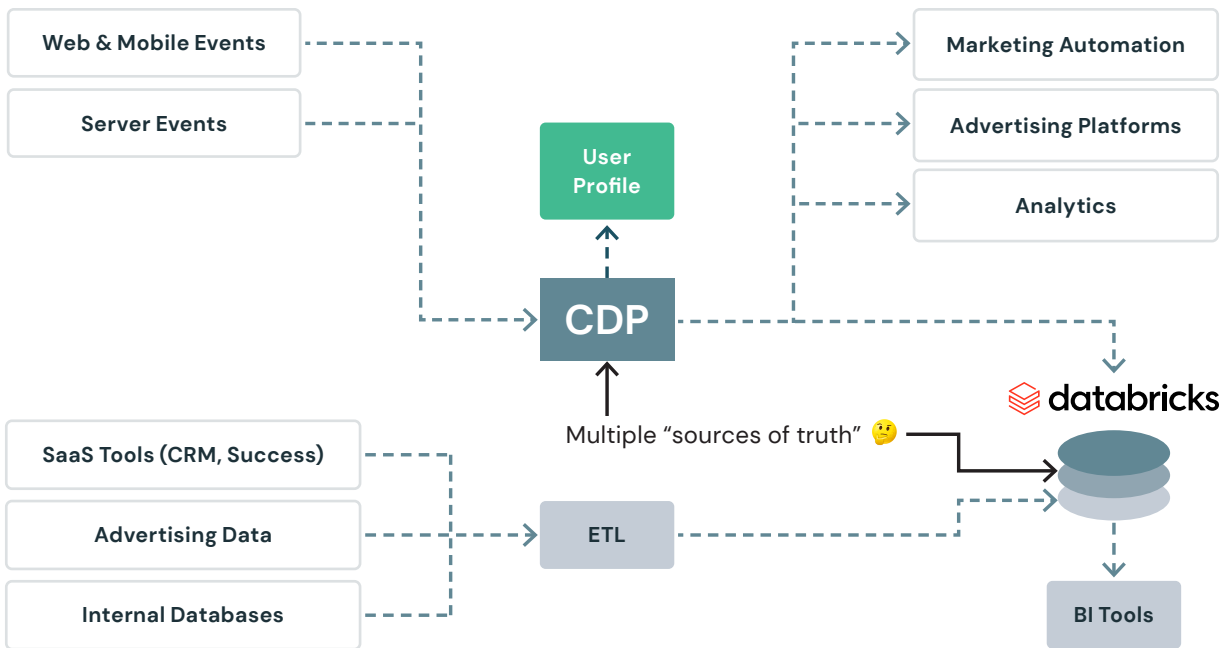
Consider these questions in any CDP assessment process:

- Are you already building a CDP or rolling out a single-vendor CDP solution?
 - Do you already have a lakehouse feeding and supporting your customer data efforts?
 - What is the skillset of your current data team?
 - What solutions do they use and how might these fit into your CDP strategy?
-

The data team challenge: What you can’t see can’t help you

When a monolithic CDP comes with its own data warehouse, how does a team know what’s in it? How do they trust it? In many cases, data professionals don’t.

One marketing professional from a major record label [made this point](#): “I often heard this from the data and BI teams who didn’t trust the CDP’s profile unification model, as traditional CDPs are a black box in terms of modeling. The BI team was tasked with reporting on the total first-party audience using the deduplicated and unified dataset captured within the CDP database. In our case, both the CDP and the BI team used Google BigQuery, so data syncs happened in real time between the two databases. Unfortunately, once the BI team started working with the data, they questioned the unification model so much so that they re-ingested the same data sources like ESP to build their own model.”



actionable insight from customer data. This limits the ability of data science teams to derive data for adjacent use cases, such as content optimization and contribution to machine learning models. Compounding these issues, the CDP can become a siloed copy of an organization’s critical asset: customer data.

It’s critical to engage all the major stakeholders, including the data and IT teams, and make sure the direction that’s chosen meets their needs as well, not just the needs of marketing. This will help ensure the CDP implementation aligns with the organization’s broader data management strategy.

Every enterprise has different customer data needs. This is one reason why rigid, all-in-one CDPs aren’t always the answer, and why the rise of a new class of CDP, a composable best-of-breed solution, could be a better option. The next section will discuss the extensive benefits of such a solution, including not having to start from scratch with existing infrastructure while extending the capabilities of a lakehouse to support additional use cases. Most importantly, marketers and data teams will work from the same data warehouse without the loss of CDP capabilities.



KEY TAKEAWAYS

- Traditional CDPs offer many features, but they present critical data choices an organization must be willing to accept or are prepared to overcome for it to be a good ROI.
- Be aware: A traditional CDP can become a siloed copy of an organization’s customer data.
- There are many CDPs to choose from. Get input from IT and data stakeholders to make the right choice.



Solution: The Composable Customer Data Platform

A composable CDP consists of the same components as its traditional counterparts: data creation, collection and integration. By implementing a best-in-class solution at each layer of the composable CDP, organizations can achieve a far more extensible CDP solution that can solve problems well beyond the common use cases of traditional CDPs. Understanding each of these components allows teams to make the most informed architecture decisions when implementing their own composable CDP. IT can carefully and properly implement each selected component, and move faster for less cost without the vendor lock-in of an all-in-one CDP.

There are key advantages to a composable CDP, sometimes called an unbundled CDP. “You don’t need to buy a bunch of tools at once. When you buy an all-in-one CDP, you’re forced to evaluate everything upfront because you’re locking yourself into a single platform. As the name indicates, the unbundled (or composable) CDP is the exact opposite,” [argues](#) Tejas Manohar, founder of Hightouch. “Rather than make a large upfront (and risky) commitment to an all-in-one vendor, the unbundled CDP allows you to solve the most important problem in front of you, incrementally. This allows you to choose the best solution and components for your business. You’re able to educate yourself throughout the process, and also future-proof and swap out certain components down the line when your needs change, or when a specific tool just isn’t cutting it.”

“Rather than make a large upfront (and risky) commitment to an all-in-one vendor, the unbundled (composable) CDP allows you to solve the most important problem in front of you, incrementally. This allows you to choose the best solution and components for your business. You’re able to educate yourself throughout the process, and also future-proof and swap out certain components down the line when your needs change, or when a specific tool just isn’t cutting it.”

Tejas Manohar, Founder, Hightouch

With composable, you can build on what you have

Most organizations likely already have a data management strategy and a data platform. With a composable CDP, an organization can align its teams to its data management strategy and build upon its data platform. IT and data teams can also move more quickly; instead of starting from scratch, they simply extend what they already have.

Data professionals like to work with solutions they know and trust. When a new CDP is implemented off to the side of an

A composable CDP provides many benefits

CTO / CIO	Data scientists & analysts	Marketing
<ul style="list-style-type: none"> • Reduce vendor lock-in • Right tool for the right job • Additive and future-proof • Data ownership • Reduce TCO 	<ul style="list-style-type: none"> • Support for use cases adjacent to CDP • Contribute model output • Avoid duplication of business rules 	<ul style="list-style-type: none"> • Consistent segment definitions • Organizational support and alignment • Better data quality and completeness

organization's existing data platform, issues arise. New and redundant workloads are added to already taxed resources. Data engineers have to work within tools native to the CDP, not the ones they know and love. Analysts and other domain experts must manage business rules and audience segments with the CDP, when they are accustomed to using web analytics tools like Adobe Analytics or Google Analytics. Finally, data scientists love having a single place to get all the customer data. How exactly do they contribute their models to the CDP?

When it comes to a CDP, don't reinvent the data wheel. "Using an internal data warehouse benefits marketers by tapping into the growing data and people resources within other parts of the organization. In the last few years, companies have moved their legacy systems onto a modern data stack. In the process, they have already integrated the same systems as a CDP. In addition, the BI team has staffed up with analysts and data scientists to mine the data for better business insights," argued Lee Hammond, owner of KLH Consulting, on [CMSWire](#). "These internal resources can deliver far more than reports. They can enable marketing teams with richer, directly actionable data about their customers directly to all activation channels—email, advertising, ecommerce and B2B CRM systems."

The CDP should be part of an aligned marketing, data and IT ecosystem.

When it comes to TCO and ROI, composable wins big

A traditional, monolithic CDP, when done right, indeed offers a positive ROI. Composable CDPs, when it comes to the total cost of ownership and ROI, stand to do even better. Some 75% of CDP customers expect 5x or higher ROI and most see a positive ROI in the first year of adoption. And, according to the [2021 CDP Institute User Survey](#), top CDP benefits include a unified customer view (88%) and analytics (54%).

A traditional or monolithic CDP can deliver serious payback. Composable offers even more. A composable CDP is cheaper to operate since data teams can get up to speed more or less immediately. By using best-of-breed components, organizations get a better return by going composable. A composable CDP is an effective way to align marketing, data and IT.

Consider a marketing team that bought a monolithic CDP, but did not get what was promised. Now think about the effort required by the engineering team to make all the required connections to make it work. A CDP should be about making teams as efficient as possible; otherwise, an opportunity cost is incurred if they're not working on their assignments. And if team members don't support the CDP, that creates a significant cost, especially when their organization is trying to move fast and innovate.

And, indeed, the marketing team needs to move fast if critical KPIs are to be met. The truth is, organizations move faster when they don't start from scratch, and are more efficient when they have support and buy-in from all their different teams.

Composable CDPs can also offer lower total cost of ownership (TCO) and higher ROI. With composable CDPs, organizations do not have to start from scratch, but can instead leverage the data solutions they already have. They can also bring in

support for new capabilities as they are needed, instead of buying everything up front. In terms of ROI, the boost comes from being able to move faster because they don't have to wait for a big implementation to enjoy CDP benefits.

Creating a customer data platform that works for a specific business is one of the smartest and most strategic things a leadership organization can do. In the next three to five years, this is something organizations absolutely must figure out – and do right.



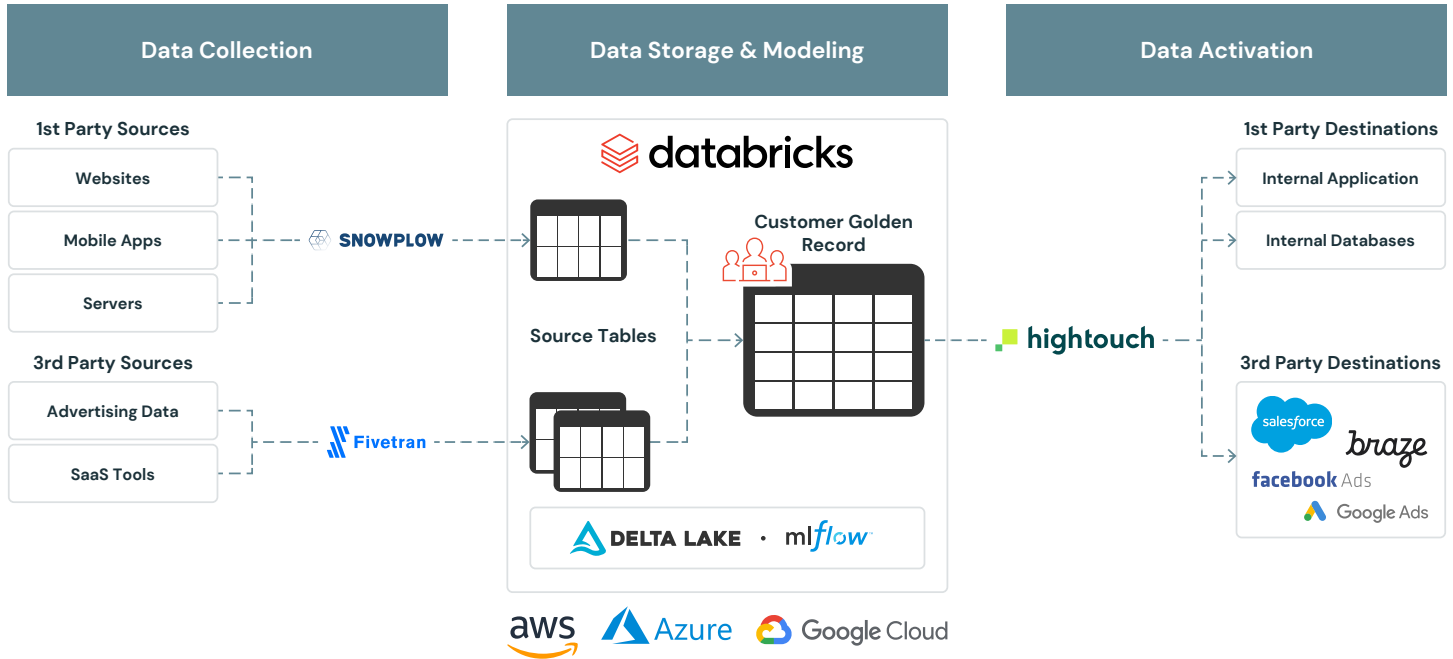
KEY TAKEAWAYS

- Composable means the CDP doesn't just start as a blank slate but remains malleable, ready to respond to new technologies and changing business needs.
- While marketing is the primary CDP consumer (at least initially), a composable CDP brings value to other parts of the organization through improved processes, supporting adjacent workloads and avoiding opportunity costs.
- Think of the CDP as an extension of the organization's broader data management strategy, not simply something that lives off to the side.



A deep dive into the composable CDP: The art of the possible

In order to realize the full potential of the composable CDP, it's time to revisit the three main components of the CDP and the composable, best-of-breed solutions that enable possibilities beyond traditional options.



Behavioral data creation (Snowplow)

Data creation is the foundation of a composable CDP. With Snowplow's [Behavioral Data Platform](#), data teams can manage the end-to-end [creation of behavioral data](#), collecting first-party behavioral data from every customer touchpoint (mobile app, website, server-side, etc.).

Snowplow offers the ability to define an unlimited number of events that can be enriched with any number of entities and properties, creating data bespoke to the business and providing unlimited opportunities for activation and personalization. The approach to events and entities helps organizations better understand customer behavior, activate data and support other evolving use cases within the business.

Accurate and compliant identification of users is the cornerstone of any CDP. Snowplow captures out-of-the-box and [custom user identifiers with every event](#). With in-stream privacy tooling, including personally identifiable information

(PII) pseudonymization, organizations have a complete and compliant view of every customer interaction.

Once data has been created, Snowplow's private deployment model and native connector to Databricks means an organization's unified event stream lands in the Delta Lake in real-time, ensuring it maintains complete ownership of its data. Data is modeled out-of-the-box at an interaction, session and user level, creating a single view of customer behavior without data wrangling. Data is ready to stitch with other data sources or can be used to develop future BI or AI data products right away.

Yali Sassoon, co-founder and chief strategy officer of Snowplow, summarized the immense benefit of such a solution:

Our ability to engage with customers and prospects effectively depends on our ability to understand those individuals: who they are, what they are interested in and

what they are trying to accomplish right now. Behavioral data is the fuel that powers that insight — it reveals not just what decisions those individuals make, second-by-second and minute-by-minute, but how they make those decisions, understanding customer intent and improving predictions. There is no type of data that is more predictive or explanatory, which is why using this data to segment customers, directly in your lakehouse, is so effective.

Storage and modeling (Databricks)

Creating and maintaining a single view of the customer delivers a tremendous amount of value for organizations big and small. Whether it's used by a marketing team to facilitate cross-sell/upsell opportunities or a product team to personalize the user experience, the value of this asset can and should be realized across all organizational boundaries. This can be achieved by using the [Databricks Lakehouse Platform](#) as the storage and modeling layer of an organization's composable CDP.

The benefit of this approach is three-fold. First, the Databricks Lakehouse Platform natively supports any type of data, whether it's batch or streaming, structured or unstructured. This means having a common way to work with all of an organization's data, whether it's clickstream data streaming in from Snowplow, marketing data that is updated in batches from [Fivetran](#), unstructured text data from a customer service tool such as Zendesk, and more.

Second, all this data being managed in an organization's lakehouse can be used directly for BI and ML/AI. For

The screenshot shows the Snowplow Data Catalog interface. The main content area displays the 'add_to_basket' event details. It includes a table of event properties:

PROPERTY	DESCRIPTION	TYPE	REQUIRED	EXAMPLES	VALIDATION CRITERIA
location	The location within the store that the product was added from.	string	✓	-	Must be one of: <ul style="list-style-type: none"> home product_listing product_page basket
mechanism	The mechanism used where the customer added the product to their basket.	string	✓	-	Must be one of: <ul style="list-style-type: none"> standard_button similar_items recommended_items featured_items last-in items
quantity	The quantity of the item added in this specific event.	number	✓	-	Minimum value = 1

Below the table, there are two informational messages:

- Additional properties are not permitted to be sent with this event.
- All events record additional properties. Dependant on the tracker type and implementation each event may also record over 100 additional properties. (View additional properties)

The 'Event map' section shows a diagram of the event structure. The 'EVENT add_to_basket' is linked to four entities: 'user', 'product', 'screen', and 'basket'. Each entity has a list of properties:

- user:** user_id, title, first_name, last_name, (+5 more properties)
- product:** product_id, category, brand, size, (+5 more properties)
- screen:** screen_name, app_section
- basket:** basket_id, basket_subtotal, promo_code, basket_total, (+5 more properties)

Add an unlimited number of entities and properties to enrich event data.

The screenshot shows a Jupyter Notebook titled '00_Composable_CDP_Demo'. The code defines a function for model evaluation:

```

1 def evaluate_model(params):
2     #instantiate model
3     model = XGBClassifier(use_label_encoder=False, learning_rate=params["learning_rate"],
4                           gamma=int(params["gamma"]),
5                           reg_alpha=int(params["reg_alpha"]),
6                           reg_lambda=int(params["reg_lambda"]),
7                           max_depth=int(params["max_depth"]),
8                           n_estimators=int(params["n_estimators"]),
9                           min_child_weight = params["min_child_weight"], objective='reg:linear', early_stopping_rounds=50)
10
11     #train
12     model.fit(X_train, y_train)
13
14     #predict
15     y_prob = model.predict_proba(X_test)
16     y_pred = model.predict(X_test)
17
18     #score
19     precision = average_precision_score(y_test, y_prob[:,1])
20     f1 = f1_score(y_test, y_pred)
21
22     mlflow.log_metric('avg_precision', precision) # record actual metric with mlflow run
23     mlflow.log_metric('avg_f1', f1) # record actual metric with mlflow run
24
25     # return results (negative precision as we minimize the function)
26     return {'loss': -f1, 'status': STATUS_OK, 'model': model}
    
```

Define model evaluation for hyperopt using MLflow.

example, using [Databricks SQL](#), the return on ad spend for marketing campaigns can be easily analyzed using an organization’s BI tool of choice. Likewise, with MLflow and AutoML natively integrated into Databricks, data scientists can easily train and productionize models, such as propensity to churn, from behavioral data leveraged from Snowplow, and then use the output of those models for activation via Hightouch.

Lastly, because the Databricks Lakehouse Platform is simple, open and collaborative, the single view of the customer can be managed and governed in a consistent and scalable way, for present and future use cases alike.

Data activation (Hightouch)

[Data Activation](#) is the final piece of the composable CDP. All business teams, from sales and marketing to support and customer success, need relevant, accurate and near real-time customer data to add critical context to the software they already use.

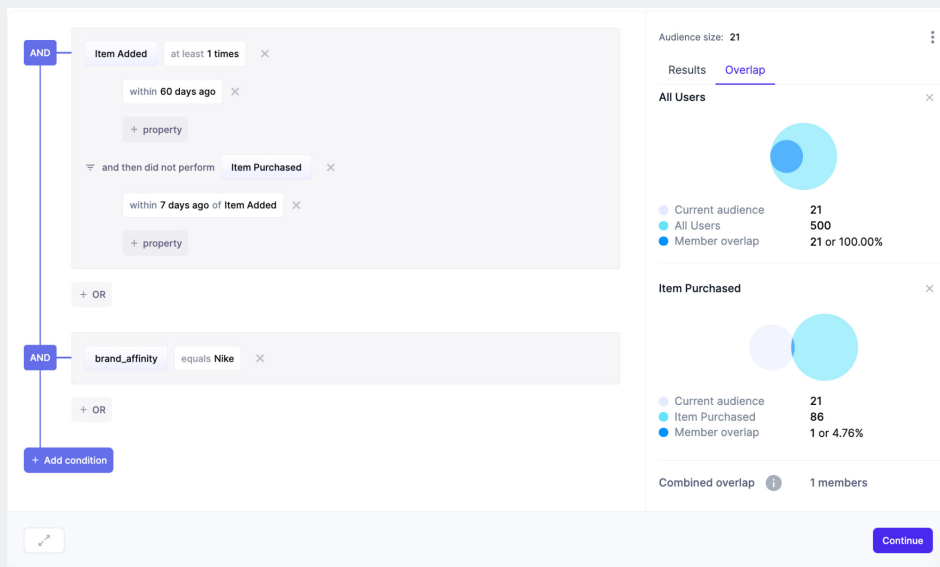
Leveraging a technology coined “[Reverse ETL](#),” Data Activation platforms stream data out from the data lakehouse to any business application. Whether an organization is [enhancing communications](#) with customers via CRM, [optimizing ad spend](#) with audience targeting, or [personalizing email/SMS campaigns](#), Hightouch makes data actionable, with no scripts or APIs required.

Thanks to Hightouch’s no-code [audience builder](#), marketing teams are able to centrally manage all their cross-channel personalization and targeting efforts within the platform. Teams can [connect Databricks as a Hightouch source](#) and, in minutes, take action on their lakehouse data in any of their various downstream marketing tools.

According to Hightouch’s Manohar, “Data Activation is the method of unlocking the knowledge stored within your lakehouse, and making it actionable by your business users in the end tools that they use every day. In doing so, Data Activation helps bring data people toward the center of the business, directly tying their work to business outcomes.”



With every element in a composable CDP, organizations can choose the best-in-class data collection, storage, modeling and activation tools that fit the requirements of each of their teams.



Define the data to sync with SQL or a no-code audience builder.



The broad benefits of a composable CDP

By harnessing the power of best-in-class tooling to create a composable CDP, there are four key benefits over a traditional CDP:

Better data governance

In today's privacy-conscious world and with ever-evolving data legislation, taking ownership and having full control of customer data is paramount. Rather than a traditional CDP managing all of an organization's customer data, a composable CDP provides full transparency, assurance and auditability at each step of a customer's data architecture.

The composable CDP is also the central repository where customer data is managed via a single profile that changes as the customer requirements shift. Controlling what PII is collected, how data is stored and modeled, and what data is shared with marketing partners ensures that an organization can comply with GDPR, CCPA and future legislation. With a single source of truth confirming who the customer is, their profile and their privacy and consent restrictions, controls can be more easily managed. Once customer data has been created, whether through a mobile app, website or other means, the customer consent is stored and applied whenever that customer is engaged. A composable CDP enables this kind of data governance at scale.

Better results with better data quality

Advanced personalization and segmentation of an organization's campaigns rely on a consistent source of well-structured, reliable, accurate, explainable and compliant behavioral data describing what customers are doing minute-by-minute. With a composable CDP, an organization can determine the events and entities that match its business and decide how its data is modeled for activation.

Although behavioral data can be exported from a traditional CDP, in reality, their data models were never intended to be used outside their platform. CDP data exports from irregular table structures and requires complex joins and transformations before it can be activated.

With a composable CDP, data science teams can directly leverage the behavioral data in their organization's lakehouse, along with Databricks' enormous data processing capability, to build AI models specific to the organization's data, product or business goal instead of relying on the black box models offered by traditional CDPs. With greater model accuracy, businesses can create additional opportunities and revenue from their campaigns.

As mentioned at the beginning of this whitepaper, marketers have a problem with data quality. "Nearly two-thirds of all marketers are only moderately confident (or worse) in their data, analytics and insights systems," explained the CMO Council.

A composable CDP is not only the place where accurate, up-to-date customer data resides. It is also where this data is cared for — where best practices for data quality are applied.

Future-proof and modular by design

Composable CDPs are future-proof by design, allowing organizations to avoid the vendor lock-in and one-size-fits-all approach associated with any all-in-one platforms. With every element in a composable CDP, organizations can choose the best-in-class data collection, storage, modeling and activation tools that fit the requirements of each of their



Rather than a traditional CDP managing all of an organization's customer data, a composable CDP provides full transparency, assurance and auditability at each step of a customer's data architecture.

teams. As the requirements of the business evolve, organizations can continue to invest on top of their composable CDP, rather than implement a new stack from scratch and rebuild the business logic, which would entail high risk and cost. The modularity also opens up opportunities to engage with either internal data science teams or with third-party data science partners to develop custom AI/ML solutions.

And last, but not least, organizations have the flexibility to determine their approach to [identity resolution](#) to ensure their team is able to deliver accurate and compliant marketing campaigns. The business has complete control over how and when to stitch together user identities, leveraging every customer data point available.

A single source of truth across marketing and other teams

Instead of adding another data silo to the tech stack, teams can do more with the single source of truth they already have — their lakehouse. With the lakehouse as the single source of truth for the composable CDP, all teams have access to the most comprehensive customer profiles and insights from across the business and can activate it through Hightouch with an easy-to-use UI and workflow.

The single source of truth also has applications outside marketing use cases. It can also power other use cases ranging from internal reporting to product analytics.

Darren Haken, head of engineering (Data & Platform) at Auto Trader, gives this example: “I’ve never seen engineers and marketing people talk to each other as much as they are right now. All our objectives are now aligned on shared outcomes,” he said. “Access to first-party data has made our data engineers have more empathy for our marketers, and our marketers are now more willing to provide insight into the campaigns they are running.”



KEY TAKEAWAYS

- The composable CDP delivers a single source of truth and better data governance at scale.
- Data science teams don't need to rely on the black box models offered by traditional CDPs.
- Composable means no vendor lock-in — the organization's stakeholders call the shots on which component works for them, not a single software provider.

Spotlight: Composable CDPs deliver optimal identity stitching

When it comes to a person's identity, there are many different identifiers. There are known identifiers like PII, email addresses and so forth, and anonymous identifiers such as device IDs that come off a laptop or smartphone.

Identity stitching is the process of tying together these identifiers to build a more complete view of user behavior. The aggregation of these identifiers builds a profile that enables organizations to understand their customers, enhance customer experiences, and engage them with relevant and accurate messaging.

One of the great benefits of a composable CDP is the added flexibility and transparency to identify data and how that data is stitched together. Rather than the rigid parameters of a traditional CDP, the composable CDP opens up the possibilities with out-of-the-box stitching (four or more IDs in a single atomic table) and the ability to define customer IDs (by enhancing the tracking JS) while the stitching logic can be defined at a later date.

The result: Organizations have the freedom to define their identity logic based on their needs and, therefore, more accurately identify users of their website or application across multiple devices, before and after they log in.



The composable future: The expanding role of CDP and CCDPs

Traditional CDPs have made their impact in marketing, and are aimed at attracting and keeping customers. Currently, the primary use cases for CDPs include acquisition (57%), customer value (65%) and retention (61%), according to the [CDP Institute](#). But that's only the beginning. The composable CDP stands to open up a range of use cases beyond the marketing department.

“Interest in CDPs has not been confined to marketing: CDP search activity on gartner.com among CIOs increased by 91% between 2020 and 2021. Executive leaders expect that customer profile data will support the rapid digital transformation pursued by IT, marketing and cross-functional CX teams,” Gartner found. “Data to power customer

experiences is beyond the sole purview of marketing teams and their technology solutions. As CIOs develop new digital products (not just digital storefronts for physical goods), CDPs face a range of new demands from digital commerce, sales, operations and customer service to unify and disseminate data to power relevant customer engagements. Where the CDP may initially have served marketing teams seeking to make use of data from elsewhere in the organization, the demand is now bidirectional. Expansion of use cases beyond those orchestrated by marketing will challenge the CDP in terms of data governance, replication and workflows.”

Composable CDPs are potentially the enabler for organizations to deliver on these expectations.



Conclusion

Implementing a traditional CDP can be challenging, especially at an enterprise level. Because traditional CDPs revolve around their native customer database, product engineering teams must implement data collection by tracking user traits and events across various websites, backend services and apps via CDP APIs and SDKs. Implementation can often take three to six months before marketing efforts can even begin.

The composable CDP allows organizations to solve the most important problems incrementally, enabling them to choose the best solution and components for their business. Teams can educate themselves throughout the process and future-proof and swap out specific components down the line when their needs change or when a particular tool isn't cutting it.

Thanks to the partnerships of Snowplow, Databricks and Hightouch, getting started with a composable CDP has never been easier. Snowplow's new Databricks Loader allows businesses to load directly into both Delta Lake and Databricks. To get started with Snowplow and Databricks,

contact us or deploy Snowplow Open Source. And with the launch of Hightouch on Databricks Partner Connect, Databricks customers can establish a secure Hightouch integration in just a few clicks.

Further instructions on how to get started with each of the market-leading composable CDP tools can be found below:

- Snowplow Getting Started snowplow.io/get-started/
- Databricks Getting Started databricks.com/try-databricks
- Hightouch Getting Started app.hightouch.com/signup



KEY TAKEAWAYS

■ Vet the solution

When looking for a CDP, make sure the solution can successfully harness customer data.

■ The C-suite

The C-suite should be highly active in the CDP acquisition process and choose a customer data solution that is agile, future-proof, simple, extensible and part of the overall data management strategy.

■ Harnessing customer data = competitive advantage

If your competition has superior customer data, even your most loyal customers are easy pickings.

■ Optimize first-party data

Composable CDPs help enterprises collect and optimize first-party data, thus building competitive and customer advantage.

■ CDPs are giving way to composable

While traditional CDPs can be great, composable is changing the playing field – and the very definition of CDP.

■ Why start from scratch?

Don't replace a data lake that's trusted (and the data folks already love) with one that isn't.

■ Overcome CDP limits with composable

Composable CDPs are overcoming the limitations of traditional all-in-one approaches.

■ Adapt to change, drive into the future

Composable CDPs are extensible, and easily respond to new technologies and changing business needs.

■ Escape single vendor handcuffs

With composable, there is no vendor lock-in. The business calls all the shots.

■ Extend, don't replace

A composable CDP is a logical extension of an organization's broader data management strategy, not a new solution living off to the side.

[Get Started Today](#)

Discover more about the
Databricks Lakehouse for Media & Entertainment

