

A graphic featuring a large dark blue circle in the center, surrounded by a light blue ring. The background is white with light blue geometric shapes. Inside the dark blue circle, there are faint blue curved lines and a small white star-like icon.

Modern Customer Data Stack

**A better alternative to
customer data platforms.**

Table of contents

01	Introduction	03
	What is the Modern Customer Data Stack?	
02	Data Layer	05
	Data Warehouse	
	Data Lake	
03	Ingestion Layer	06
	Batch Ingestion	
	Buy a Tool	
	Build yourself	
04	Customer Data Model	08
	Identity Resolution & Dedupe	
	Creating Master Models	
	Tools	
05	Visualization & Intelligence Layer	12
06	Activation Layer	13
07	Bottom line: Own your data, own your growth	15
08	Ready to get started?	16

01 Introduction

Did you know that innovative data and analytics leaders are taking control of their first-party data in their warehouses and moving away from relying on third-party platforms like CDPs?

This is becoming increasingly necessary, as user tracking with cookies is going away, privacy regulations are top of mind, and limitations are being implemented by big tech on what's trackable.

Companies need to start owning their own data destiny by taking control of their first-party customer data stack.

This approach provides complete privacy and governance over your own data and unlocks huge potential for innovation and data-driven growth (detailed breakdown of reasons why CDPs are not the future [here](#)).

Companies need to start owning their own data destiny by taking control of their first-party customer data stack.

01 Introduction (cont.)

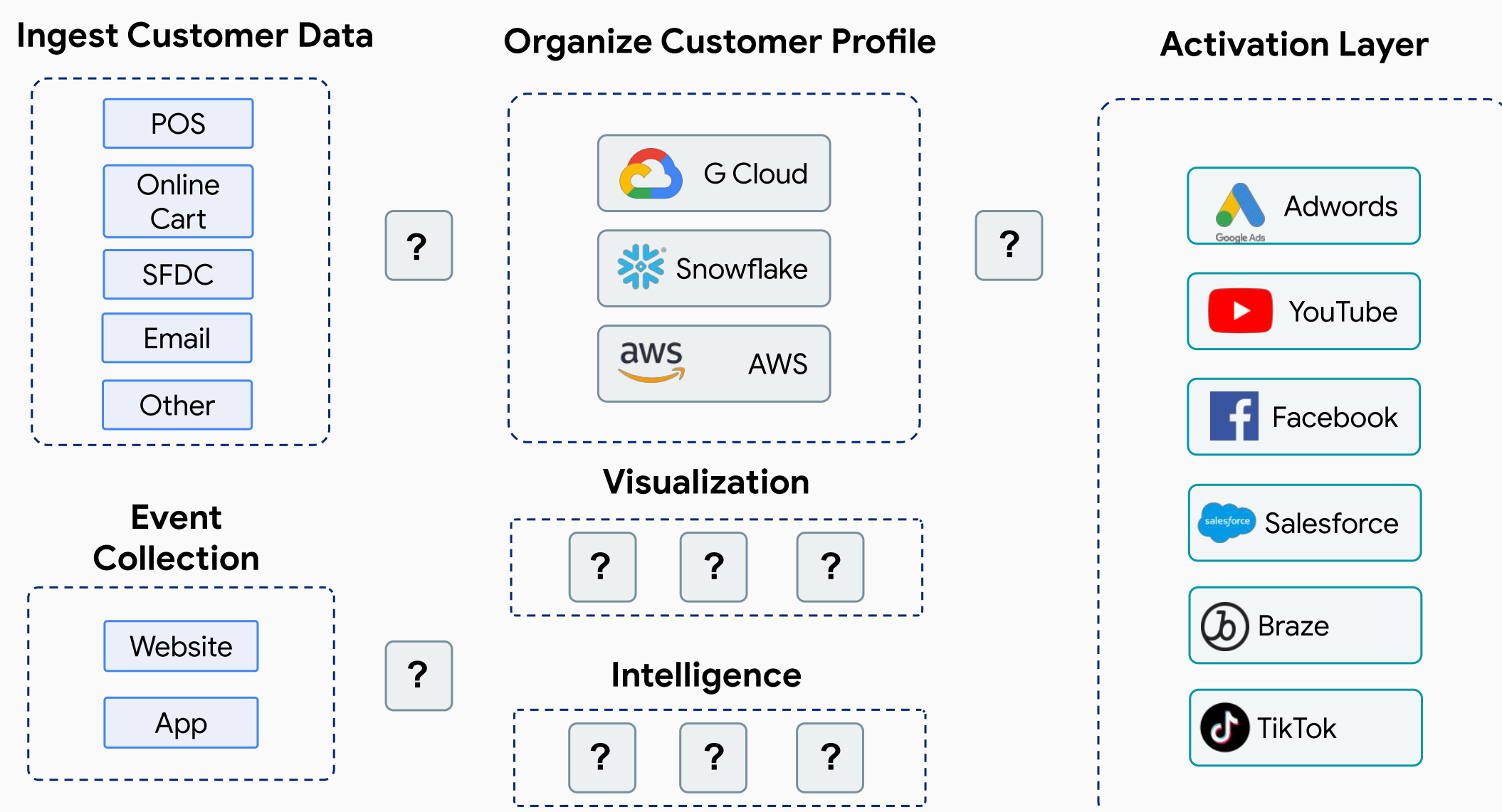
We call this new approach the **Modern Customer Data Stack** and it is fully centered around the data warehouse or a data lake.

The great news is that over the past few years it has become easier to create and own your Modern Customer Data Stack as an analytics leader.

In this article, we're going to take a look at what the Modern Customer Data Stack looks like and what approaches we are seeing ambitious data owners take to bring it to life. We'll also review the best tools you can use to get your Modern Customer Data Stack setup in just a matter of days.

Components of the Modern Customer Data Stack:

- **Data Layer** - The core data storage layer including the data warehouse or data lake
- **Ingestion Layer** - Ingest customer data to your data warehouse from various sources
- **Customer Data Model** - Organize your customer profiles in your data warehouse into a single view
- **Visualization & Intelligence Layer** - Gain insights into customers and marketing efforts to drive growth



02 Data Layer

The Modern Customer Data Stack is centered around utilizing tools that have already paved the way for modern data architectures and building on top of these toolsets. You can reap the benefits of this architecture as long as your data is centralized into a queryable storage layer.

Data Warehouse



For the data layer of the system, this architecture is powered by and builds upon the performant features of data warehouses.

Companies have already been heavily adopting the use of data warehouses like BigQuery, Snowflake, or Redshift to centralize their data, and the modern data stack builds upon that work instead of replacing it.

By owning your data in the warehouse you open the path for your data analytics and data science teams to innovate in their preferred toolsets while owning your security, privacy, and governance of the data.

Data Lake



If your team is utilizing a data lake architecture using something like AWS Athena, or Azure Synapse, you can still use it as the storage layer for your Modern Customer Data Stack in place of a data warehouse.

The data lake tools also offer a SQL query interface that allows you to use them as the storage layer in this stack.

03 Ingestion Layer

A major step to building a customer data stack is to get all your 1st party data into one place. There are many ways this can be accomplished but we will touch on a few common approaches.

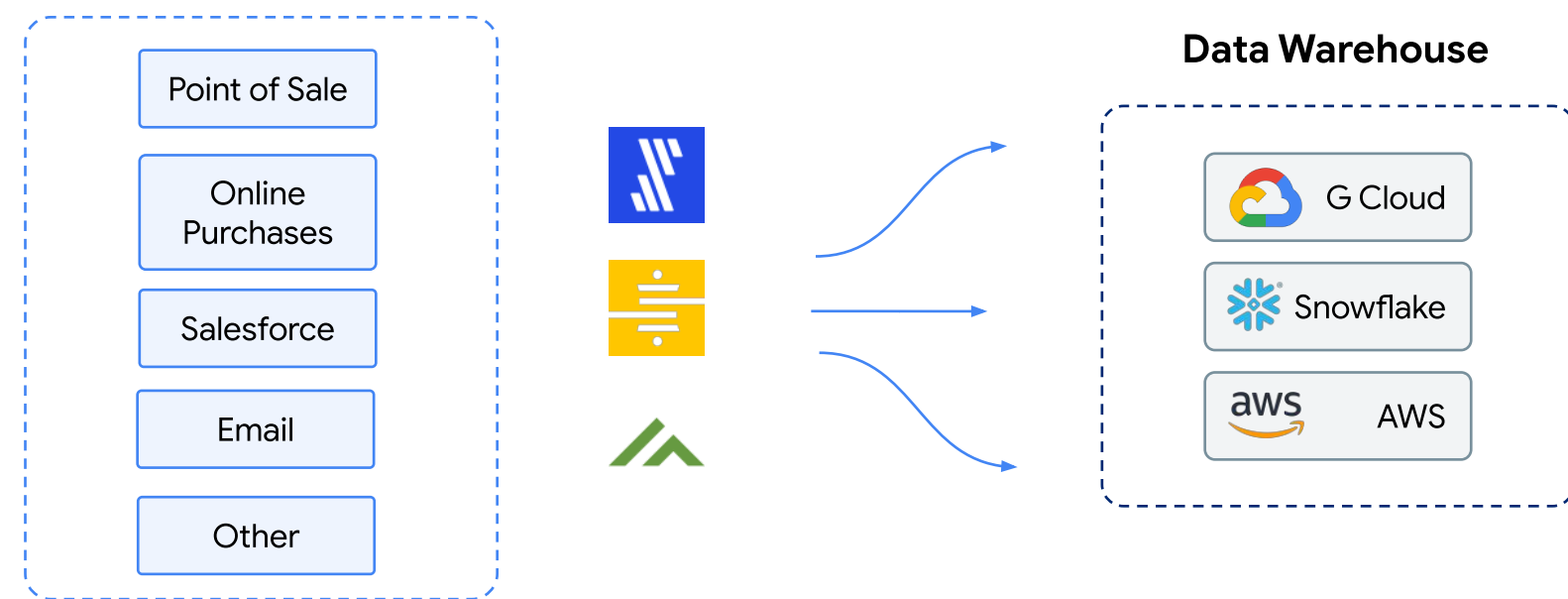
By bringing all the raw data into your data warehouse and modeling it, you unlock a lot of value for your team to use this data for growth. In reality, your data will be fragmented into various internal or external systems.

Some of these systems have APIs, ODBC Connectors, SFTP, or other interfaces that have to be integrated to fetch the data. So what's the approach to wrangling all the data into the warehouse?

Batch Ingestion

The fastest way to get started is to adopt an ETL tool of your choice to load batch data from various APIs and connectors.

Batch Data Ingestion APIs & ODBC Sources



These tools are great for importing your data from different platforms in their raw form on a set cadence.

An example might be ingesting data from your Salesforce instance into Snowflake or BigQuery. You'll find various open-source or third-party tools in this space. The decision on which tool you end up using depends on your available engineering resources, cost, and how much of the process you want your team to own

By using one of these tools, you've got the first step out of the way to get your data landing in your warehouse in a standard schema. Here are a few options we recommend:

- **Open Source** - Singer, RudderStack
- **Third-Party** - Fivetran, Stitch, Matillion

03 Ingestion Layer (cont.)

Product Event Collectors (Web & Mobile)

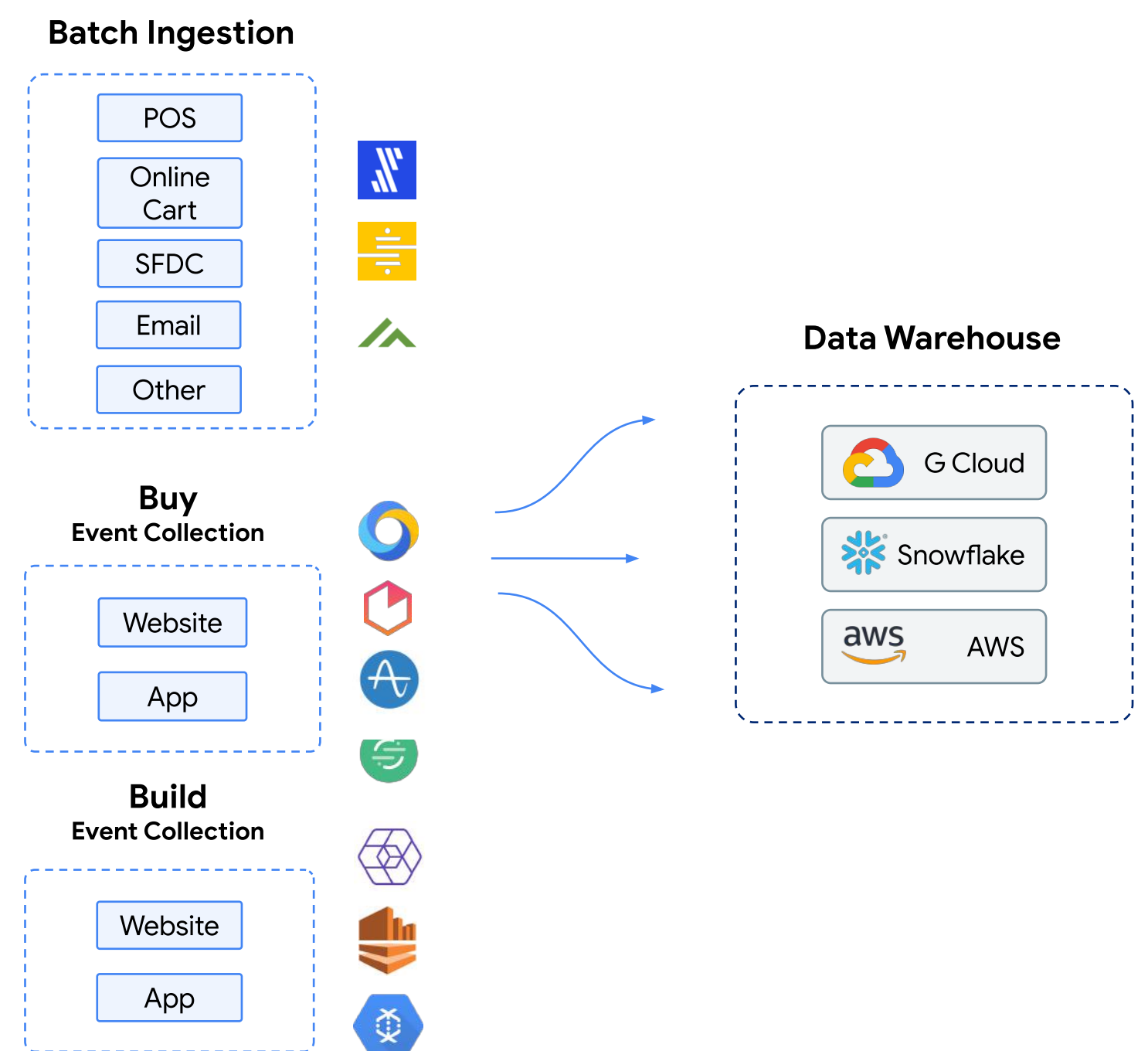
In addition to importing batch data, a lot of companies looking to get product analytics implemented will opt for tools that directly integrate with their web or mobile apps for event collection.

This approach requires more work upfront by your engineering team to integrate the SDKs in your mobile apps and snippets on your web apps. They allow you to track product events like logins, searches, or clicks. For the product event collectors, you have the option to either **Buy** or **Build** yourself.

Buy a Tool

To get your data stack setup faster, buying a tool is the way to go. If you choose to go the buy route, you have options like Segment, Mixpanel, Amplitude, or Heap available in the market. As a starting point, review which analytics tools your team is already using today.

Usually, your engineering team is already collecting events somewhere via GA 360, Segment, Amplitude, or Heap. In terms of the tools, [Amplitude](#) and [Heap](#) can sync to your data warehouse every 10 minutes. [Segment syncs](#) at least twice per day to your data warehouse depending on your plan. [GA 360](#) offers daily syncs to Bigquery.



Build yourself

If you expect to scale your collection while saving on costs, you might want to consider the build option. Most teams are using Amazon Kinesis, Kafka, or PubSub to roll out their own systems, or they use an open-source framework like Snowplow.

If you have a strong engineering culture and are going with the building approach, we recommend Snowplow as we see companies getting the farthest, fastest. You can deploy the Snowplow stack on GCP or AWS via [Terraform modules](#) to get it up and running quickly in your public cloud to start collecting events in your Data Warehouse.

04 Customer Data Model

Now that you have your customer data in raw form in your data warehouse, it is time to organize it to make it useful to your teams. This is the part of the problem where good architecture design and thought will go a long way.

The goal is to organize the raw data into actionable models or entities that work for your business use cases. This stage of the stack can involve a few key components:

- **Identify Resolution** - Identifying the same users in different data sources
- **Master Data Models** - Creating a final/clean view of your customers and associated facts and dimensions.

A word of caution here is to start simple and not try to boil the ocean. Building models incrementally and iterating quickly will get you to real value sooner. Remember, our goal is delivering value in days, not months.

One approach that we've seen works well is to start by creating business objects needed for activating growth within marketing and sales — customers, transactions, and events.

Identity Resolution & Dedupe

The goal of Identity Resolution is to deduplicate individuals both within data sources and across data sources, leaving you with a single identifier (id) for each customer you have. This is an important step to resolving a single view of your customer for communications purposes.

Identity resolution can seem like a daunting undertaking but even for this problem, there is a simpler way to get started. You might feel the itch to go down the rabbit hole and look at black-box/AI-driven identity resolution solutions but forget about them for now.

We recommend starting off with rules-based and precedence-based identity resolution where you match on specific keys between different tables.

04 Customer Data Model (cont.)

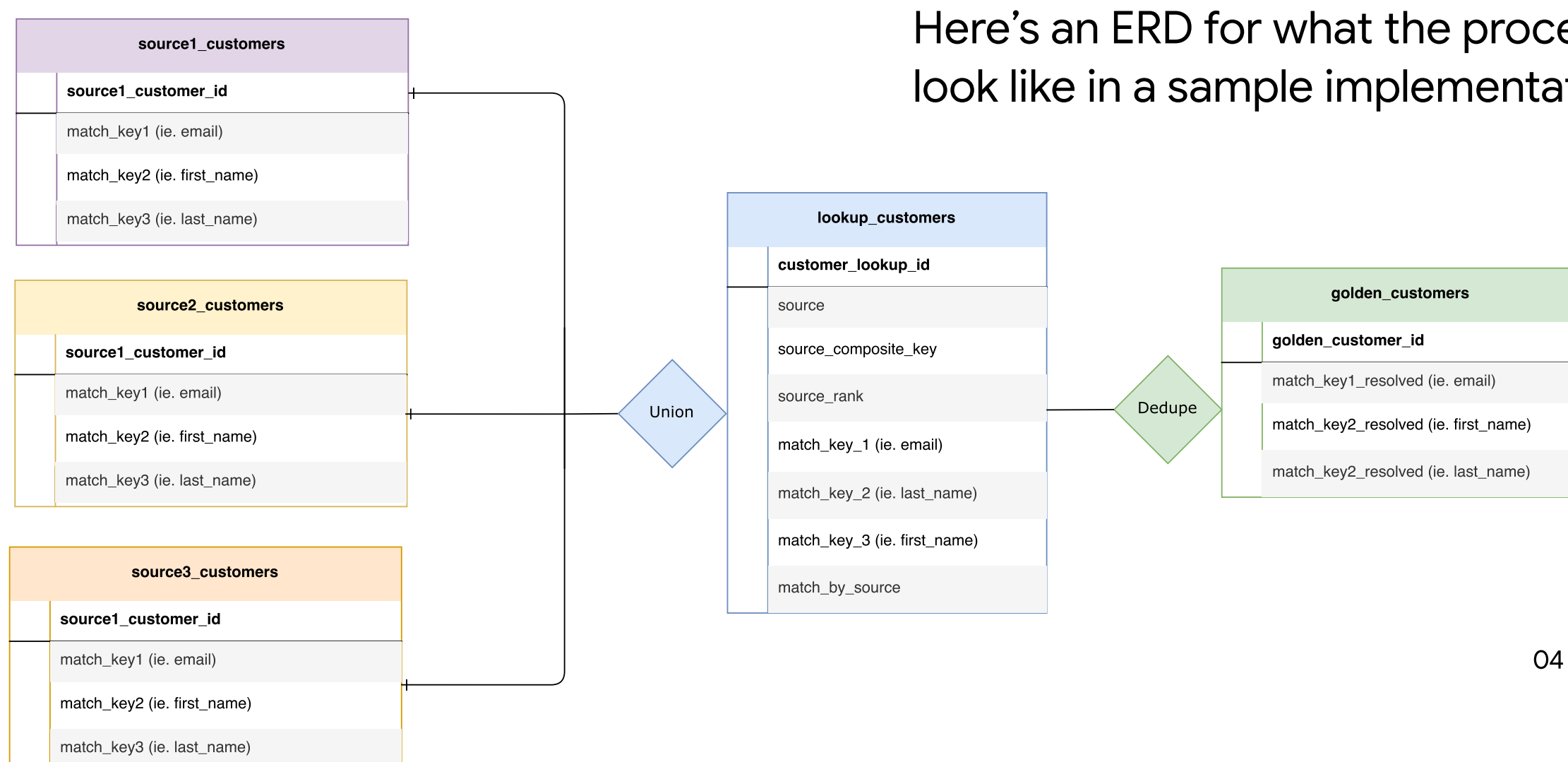
The key steps for SQL based simple identity resolution in your Data Warehouse:

- **Identify match keys** - The first step is to determine which fields or columns you'll be using to determine which individuals are the same individual within and across sources. A typical example of match keys might be an email address and last name.
- **Aggregate Customer Records** - The second step is to create a source lookup table that has all the customer records from your source tables. We call this the lookup table.

- **Match & Assign a Customer ID** - The final step is to simply take records that have the same (or in some cases, similar) match keys and generate a unique customer identifier for that matching group of customer records. Every customer id that is generated can be used to link the customer sources together going forward.
- **Roll in more sources** - As you get more sources you can start rolling them into the same process by setting the correct rules and precedence for the source.

Flywheel has built models to make simple identity resolution in the data warehouse approachable and practical. If you are interested in learning more about our approach, you can get in touch with a Flywheel Engineer [here](#).

Here's an ERD for what the process can look like in a sample implementation.



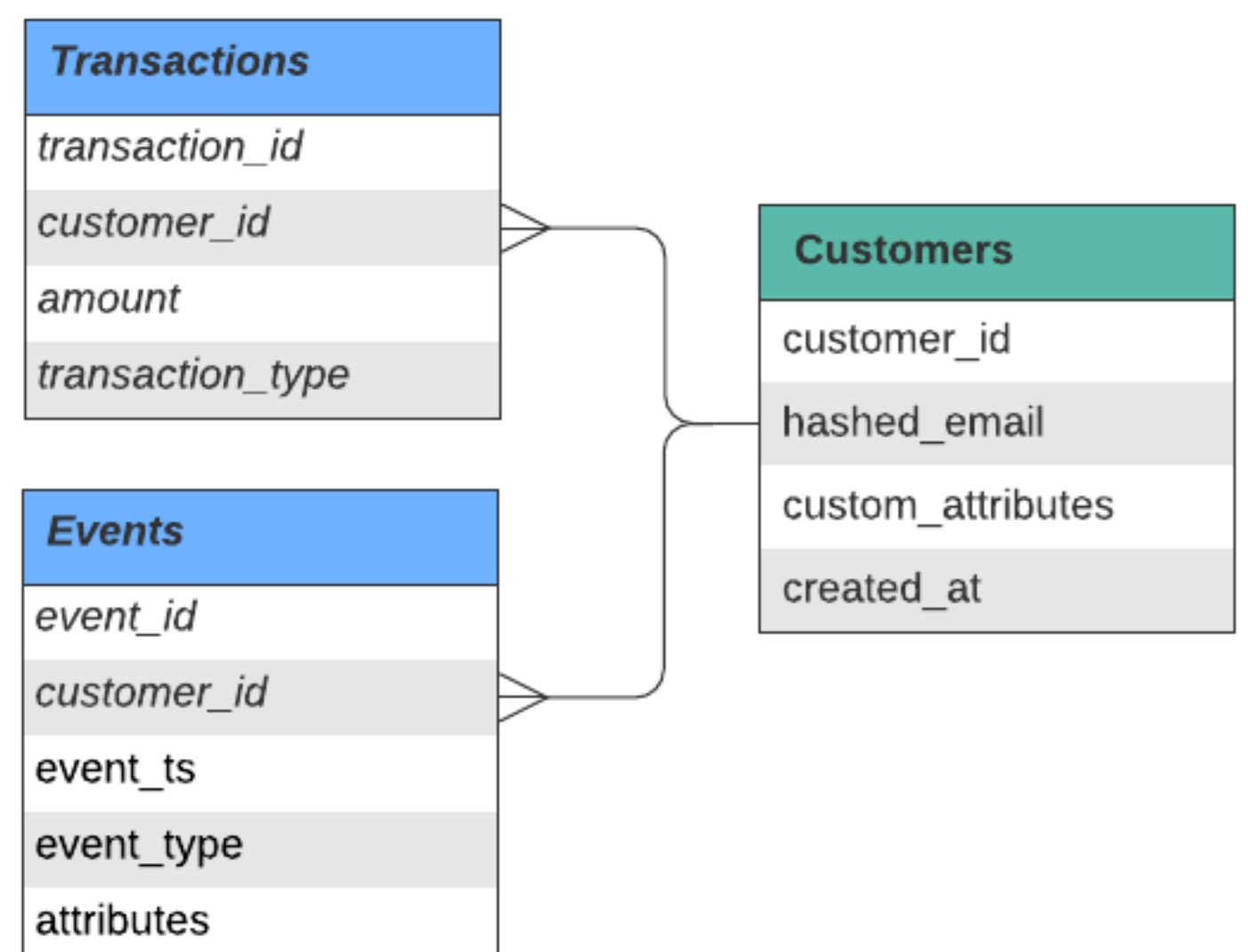
04 Customer Data Model (cont.)

Creating Master Models

For creating your first customer view, you've got the first problem of identity solved. This is followed by getting your data pipelines or ETL processes in place to build your master models that will drive analytics.

To drive quick value we recommend starting with a “**Customer** → **Transaction** → **Event**” framework. This framework entails creating the three key objects from your source data. Here is a bit more detail on this type of [schema](#).

- **Customers** - Create a table of your customers with the ability to quickly add new fields
- **Transactions** - Join key from customers table to their transaction history
- **Events** - Any events you track for each customer



If your company is a double-sided marketplace or has different business entities, you can change these master data models to follow what makes sense for you.

For example, for a double-sided marketplace, you might have tables for both sellers and buyers as different entities. For a B2B business, you might have separate accounts and contacts entities.

04 Customer Data Model (cont.)

Tools

While there are many ways to take source data and transform it in your warehouse, we are seeing innovative analytics teams move faster with an open-source stack.

Over the past few years, open-source tools for creating the Modern Customer Data Stack have come a long way, making managing and maintaining your data easier.

A lot of teams have been switching over to Data Build Tool (DBT) for building, maintaining, and quickly iterating on models with production-grade data pipelines. Our team uses and is a big fan of DBT. The same thing can be accomplished with other ETL tools or even Lambda serverless architectures.

When it comes to workflow orchestration, Airflow has been widely used in the space for running data pipelines or machine learning models.

It can also be deployed as a managed service on both GCP and AWS, so your team doesn't have to manage the infrastructure. Other alternatives to Airflow include Prefect, Dagster, KubeFlow e.t.c.

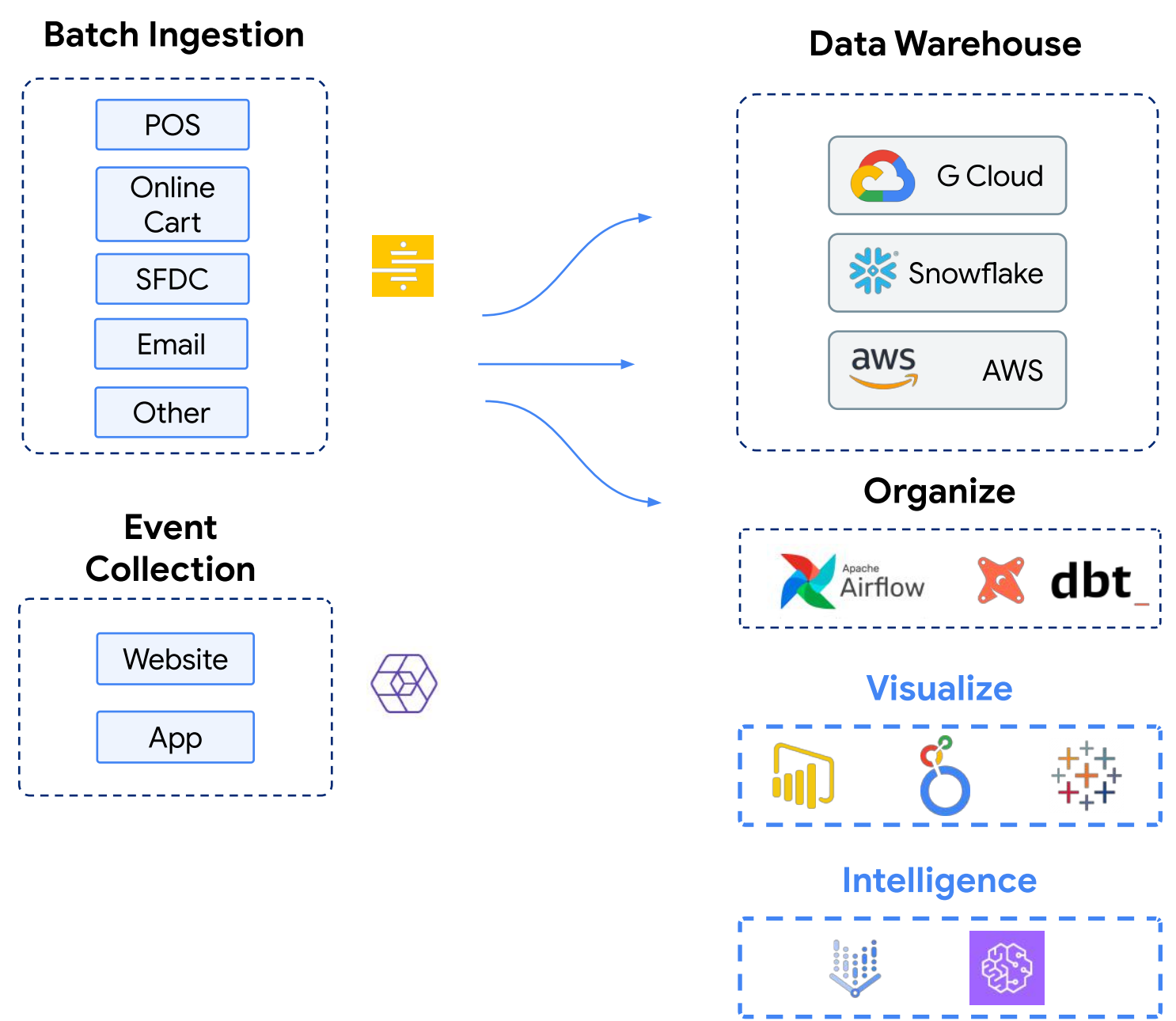


05 Visualization & Intelligence Layer

One of the main drivers behind setting up a Modern Customer Data Stack is to have the flexibility to drive insights, grow your business and improve products. As soon as you centralize your data into a data warehouse you instantly unlock these insights through visualizations and the ability to run predictive analytics. For visualization, you can hook up the usual suspects like Looker, PowerBI, DataStudio, or any of the other tools.

The intelligence layer is where things start to get exciting with the ability to create ML models or extend innovation with any tools of preference for your data science or analytics teams - python, spark, SQL. Your teams can now be wizards with all the data available at their fingertips. As we've written about before, this is a key differentiator of the Modern Customer Data Stack vs using a Customer Data Platform ([link](#)).

A simple yet often used example is the creation of a customer churn score model. This can be easily accomplished by leveraging tools like Amazon SageMaker, BigQuery ML, or Vertex AI for training, tuning, and predictions. The predictions from these models can be easily rolled into your models and served to your Marketing or Sales teams to prioritize their outreach to customers.



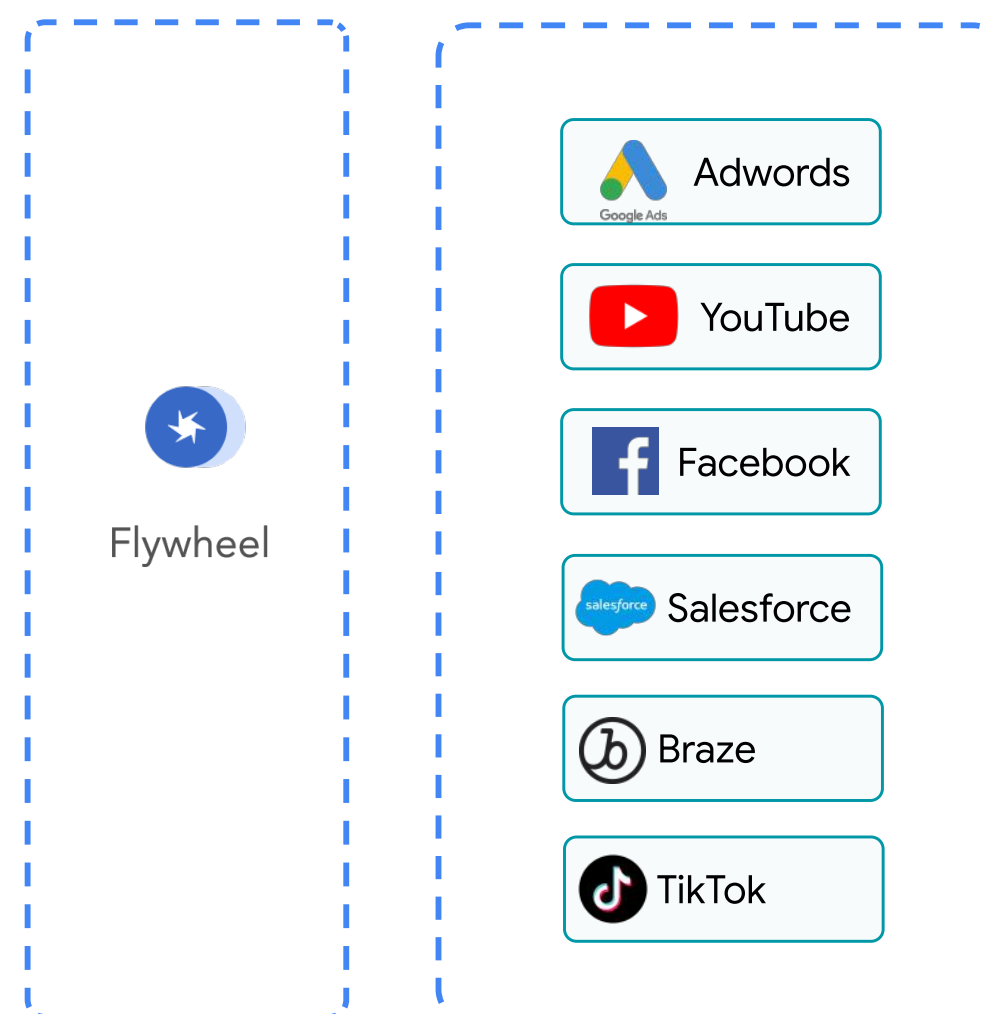
Once this Intelligence Layer takes shape at an organization it becomes a virtuous cycle driving growth. Your data team understands your business and customer data best, so they can create the most effective insights and models. These models are used by your business team to drive marketing and sales campaigns to influence customer behavior and growth. The business team then begins asking for more insights and predictions driving further customer data innovation.

And at the end of this cycle, your company owns the entire data and insights asset, giving you the flexibility to build on it and leverage it across new systems while never being locked into a specific 3rd party. That is the magic of the Modern Customer Data Stack.

06 Activation Layer

After your data has been organized and ready for driving insights on your customers, the next key question is how do you activate those 1st Party Customer Data insights with your sales, marketing, customer service, and product teams? After all, the key reason to own your Modern Customer Data Stack is to be able to activate it to drive changes in customer behavior and growth.

Typically, we see companies using a wide variety of tools across teams to engage customers including Salesforce, Marketo, Braze, Facebook Ads, or Google Ads. So, building on our previous example, how would you take a list of customers that are likely to churn and engage them in these tools?



Today, if you wanted to run a Facebook Ad campaign for these at-risk customers you'd have to create a SQL query, pull a list from the data warehouse, download it to CSV, and upload it into Facebook Ads or build an integration.

This process is highly manual, time-consuming, and creates too much friction to realize the full potential of the Modern Customer Data Stack.

This process is highly manual, time-consuming, and creates too much friction to realize the full potential of the Modern Customer Data Stack.

06 Activation Layer (cont.)

The big missing piece to unlock the potential of your Modern Business Data Stack is the activation layer– the ability to take the customer insights you are generating from your Modern Customer Data Stack and activate them across your sales, marketing, and customer service tools. [This is the gap that Flywheel fills.](#)

[Flywheel Software](#) is an activation layer on your Data Warehouse that takes just minutes to set up. It empowers your teams to create customer segments in a visual builder without using SQL and activate them to your marketing, sales, and customer service [destinations](#) for a variety of [use cases](#) in minutes.

You might have seen the idea of Reverse ETL taking shape in this space to send specific attributes about customers to software tools.

Flywheel has taken this concept even further by creating the first Smart Reverse ETL purpose-built for enterprise sales and marketing use cases. This process is highly manual, time-consuming, and creates too much friction to realize the full potential of the Modern Customer Data Stack.

07 Bottom line: Own your data, own your growth

The Modern Customer Data Stack is giving enterprise companies a way to own their growth in an uncertain environment.

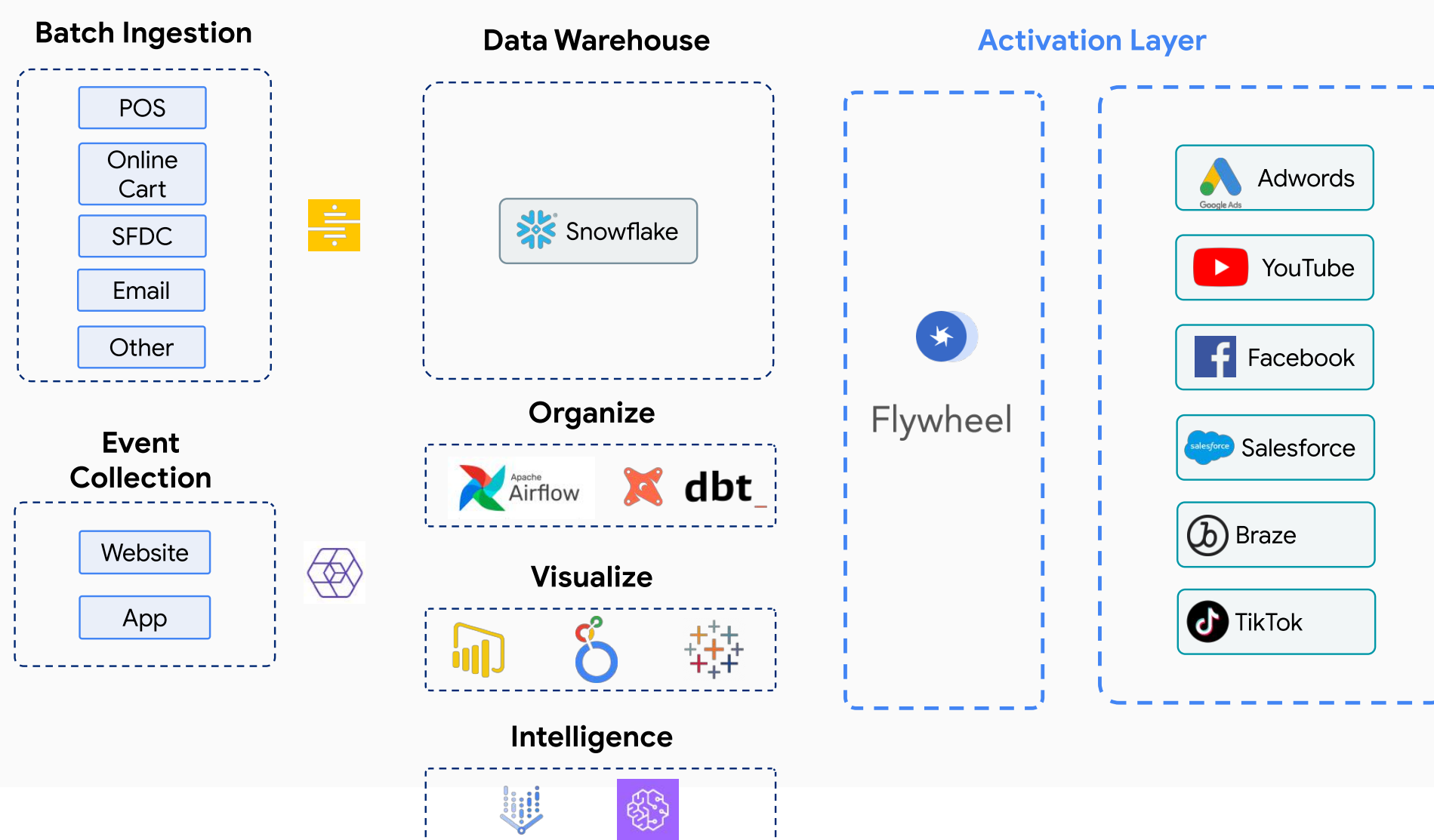
Over the past few years, innovative cloud-based tools have made it easier for companies to build their own Modern Customer Data Stack, lowering the cost of ownership.

At the same time, the need to bring first-party customer data in-house has increased dramatically for enterprise organizations with changes at the government and big tech companies.

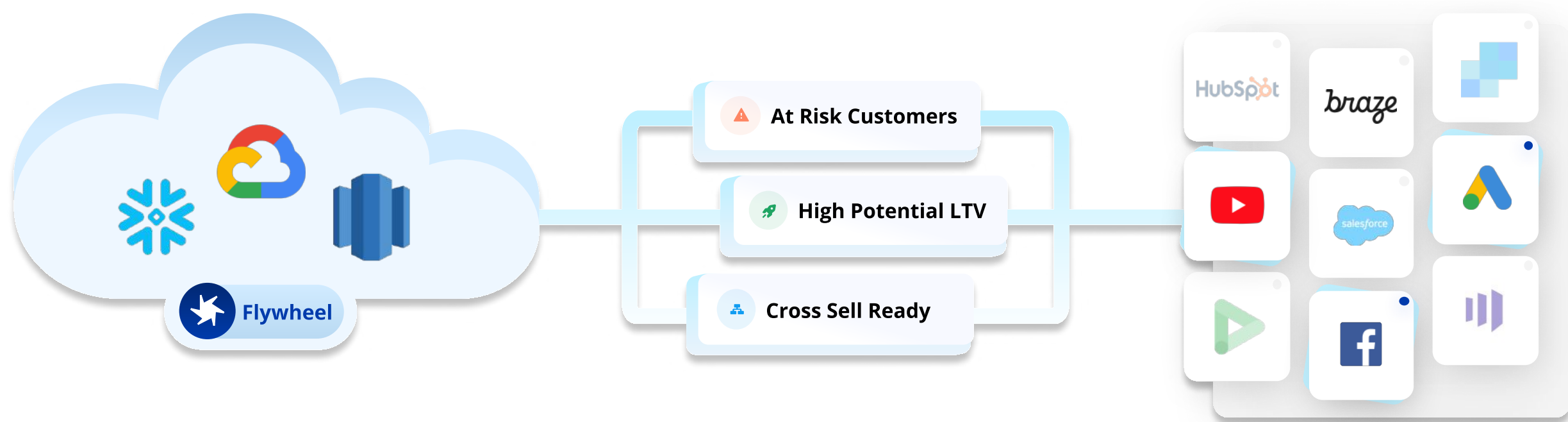
These key factors are making the Modern Customer Data Stack a top priority for enterprises in 2022 and we've been hearing the concept resonate across industries.

We are excited to see the Modern Customer Data Stack take shape and help high-growth companies enhance their growth engines.

The systems that companies put in place now will determine their ability to own their data and their growth for years to come.



08 Ready to get started?



If you are thinking that you want to own your own Modern Customer Data Stack and are wondering if it's realistic for your organization, the answer is yes, and it's simpler to set up than you might think.

Flywheel was created to help companies build and own their own Modern Customer Data Stack. No matter what stage you are in, we can help accelerate your journey. If you need help with ingestion or data modeling, our professional services team of cloud and data experts can help.

If you are looking to activate your existing data, the Flywheel Customer Segmentation platform is deployable in minutes to start driving revenue on your Modern Customer Data Stack. Reach out to us at solutions@flywheelsoftware.com or schedule a consultation [here](#).

CONTACT US

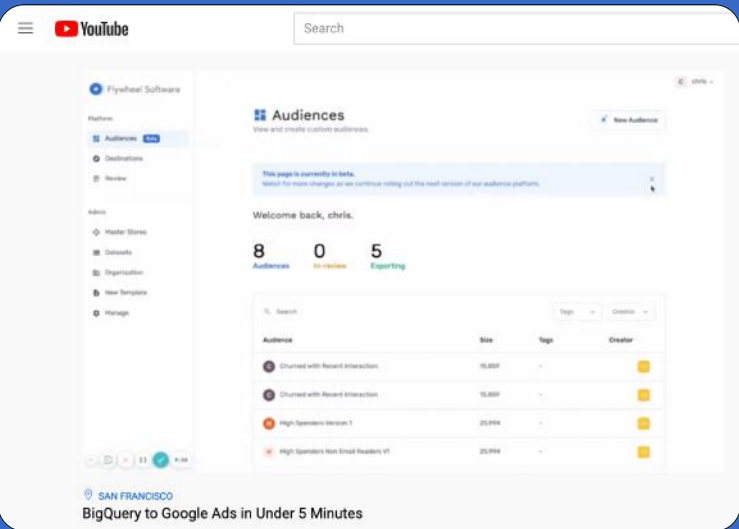


Flywheel Software

Email: solutions@flywheelsoftware.com

Schedule demo: flywheelsoftware.com/contact-us

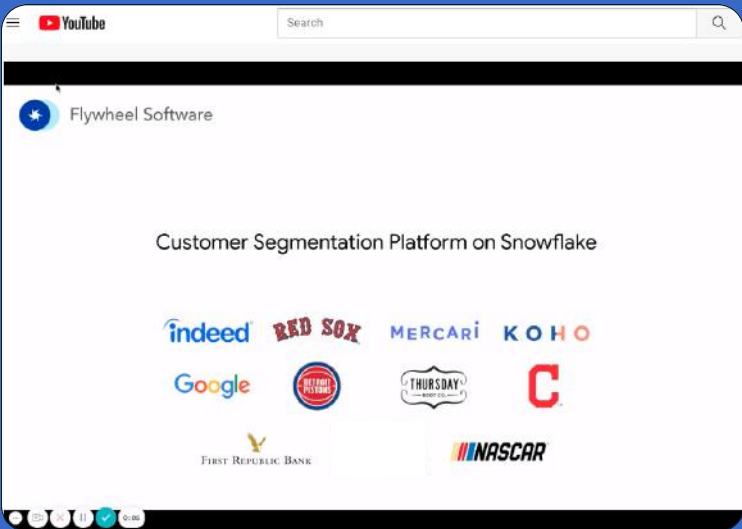
Linkedin: [@flywheel-technologies](https://www.linkedin.com/company/flywheel-technologies)



On-Demand BigQuery Demo

Learn how you can activate customer data on Google BigQuery to any destination in less than 5 minutes.

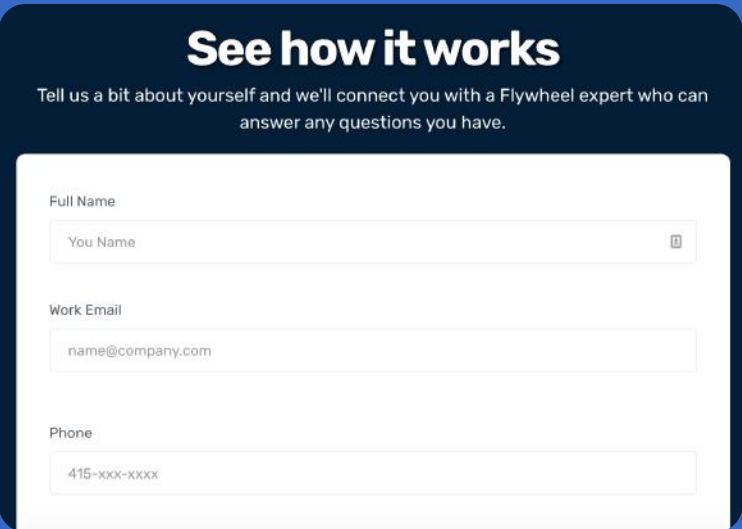
[Play Now](#) →



On-Demand Snowflake Demo

Learn how you can activate customer data on Snowflake to any destination in less than 5 minutes.

[Play Now](#) →



Personalized 1:1 Demo

Get a tailored walkthrough and learn how Flywheel can activate the customer data in your Data Warehouse.

[Request a Demo](#) →