klaviyo"

Customer data infrastructure

Go under the hood of Klaviyo's data management





Cart Che

Klaviyo Integrations with Shopify, WooCommerce, Magento, and Others

Event Streams

Relational Databases

Clickhouse

Amazon S

More customer data leads to more powerful insights

That said, as marketing teams collect an increasing amount of data, the technology they use to manage the information and extract insights becomes critical.

In fact, <u>68% of CMOs consider marketing tech stacks</u> as the heart of their marketing strategy, empowering them to better understand and meet customer expectations.

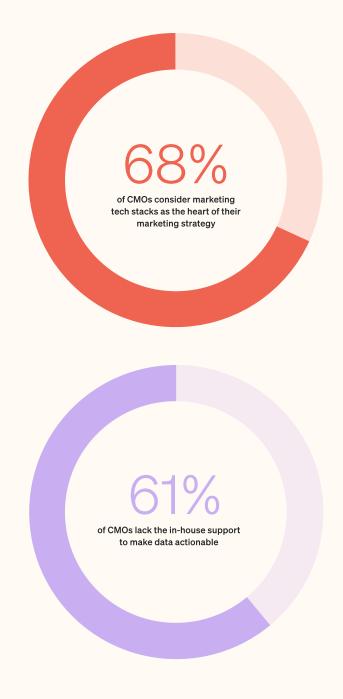
That said, Gartner also found that 61% of CMOs lack the in-house support to make the data actionable—gathering and integrating customer data into their business model still happens in multiple steps. They have to collect the data, transform it into usable formats, and connect it to end applications in a way that makes sense and is actionable for the company.

This is the case when using tools like Salesforce, Adobe, Listrak, HubSpot, Oracle, and many more. They all require data to be defined and pre-configured before they can accept it. For example, you can see in **Salesforce's documentation** their 11-step process to create a data extension to make data usable in their email studio.

More basic platforms, such as Sendlane, only show whether someone has placed an order—not giving any visibility into the actual product(s) purchased.

So, without IT and D&A departments, marketers cannot go anywhere in their personalization goals. And even with that help, they risk losing data consistency.

Klaviyo solves this challenge by going from data to applications, rather than building inflexible data stores tailored for single-end applications. The data platform at Klaviyo was built from the ground up to offer everything a modern marketer needs to store and mine their data without rigid, predefined restrictions.



To understand how Klaviyo's customer data infrastructure works, let's go through modern data challenges, understand the data storage landscape, and then dig deeper into Klaviyo's tech stack.

Contents

The modern-day challenges businesses face when it comes to accessing data

Understanding data stores

Types of data stores and their use cases

The data architecture of Klaviyo

Manage your data the easy way with Klaviyo

Migrate to Klaviyo without slowing your team down

The modern-day challenges businesses face when it comes to accessing data

What's standing between your business and your data

The interaction of modern businesses with data is a lot like the experiences of drivers in the early 1900s, when they had to crank the engine to get their car started—as opposed to pushing a button in a modern vehicle and everything falling into place.

Many businesses are still in the hand-cranking phase of their data management. While they're able to access customer data, they can't use it properly because it isn't clean, refined, or actionable. Instead of finding relevant information at the touch of a button, they have to ping their data departments to define the relevant schema (place) for data, translate data into actionable insights, and transfer data from one system to another.

Common challenges include:







So, even if your business is collecting data, you may not get any value in return because combining data from different silos, refining it by applying relevant schema, and making it actionable for marketers is time-consuming and expensive—i.e., marketers have to go great lengths just to get actionable insights from data, leaving them little time for actual marketing.

Siloed data

Personalization is difficult, especially when data is spread all over the place. Emails, websites, and POS systems aren't connected, so each team has only a fragment of the customer journey.

Or, as Jake Cohen, VP at Klaviyo, puts it: "I have all this data in all these places. I have to get it in one place, ask someone to structure it, and understand how the diverse applications would use it."

Multiple tools

Businesses use multiple tools for different functions across their business. For example, your marketers and tech teams might be using separate—in-house or external—technologies for:

- · Data analytics
- Customer relationship management
- · Website personalization
- · Marketing automation
- Data management
- · Customer reviews
- Customer loyalty

These technologies have hefty license fees and, more often than not, overlapping features. And although tool costs pale in comparison to the cost of severe technical debt in terms of productivity and long-term expenses, together, they make for a poor ROI.

For example, you may be using Salesforce Marketing Cloud for email, Bluecore for automations, Attentive for SMS, MuleSoft for iPaaS, Bazaarvoice for reviews, Annex Cloud for loyalty, Tableau for reporting, Snowflake for storage, and more. All of these providers have their own way they want data structured to make it usable, and they need custom development to share data between them.

While you can still try to integrate all these systems, it often results in partially successful projects, and your marketers will still depend on data engineers to configure even small automation processes.

Take it from the ecommerce executives surveyed in partnership with BWG Strategy—<u>more than half of them couldn't</u> <u>use customer data effectively</u> because it's not easily accessible due to being on multiple tools.

Ineffective analytics

According to Mike Parker, senior director of solution architecture at Klaviyo, marketers want to "focus on what they are good at, such as marketing strategy, creative judgment, and copy creation. Beyond that, they want to be able to move really quickly."

But if the data isn't actionable enough for marketers, they can't do their jobs. The same survey in partnership with BWG Strategy backs that up, revealing that despite feeling like they collected enough data, 71% of the participants found their data unusable because it wasn't clean.

Instead, they need their engineering team to transform the data into predefined schemas so they can make sense of it, leading to significant bottlenecks in their process and making it difficult for marketers to adjust their personalization strategies on the fly.

Before looking into the ways Klaviyo addresses these challenges, let's examine the foundation of data management: data stores.

Understanding data stores

Data storage basics

A data store, or data repository, is a location where you store data for retrieval, analysis, and management.

You can think of it as a digital hard drive. However, it's a more advanced solution than typical hard drives, since it provides you with:



Organization

Usually, data stores have the information stored in a structured manner that makes it easy to find, retrieve, and use it. Examples include databases built on a schema, where data is organized into tables with defined fields.



Accessibility

Data stores make the data accessible through authorized applications, so you can access all of your information from one place instead of dealing with data silos.



Quick retrieval

Data stores are built for quick data retrieval, allowing you to use queries to quickly find and access pieces of relevant information.

Beyond that, compared to a typical hard drive, data stores are not one-size-fits-all solutions. Instead, different data stores serve different functions.

Types of data stores and their use cases

When it comes to data stores, your business has options

There are several different types of data stores you can use, and each one has its advantages and disadvantages. The key to effective data management lies in selecting the storage option that best fits your needs.

Here's a look at nine of the most commonly used data stores and how to best use them.

Relationship database management systems (RDBMS)

An RDBMS refers to a data store that organizes data in tables with rows containing different entries and columns defining their respective attributes. Plus, you have keys to connect the two tables and show their relationship with each other.

RDBMS makes it easy to store structured data and retrieve it whenever required for business use. That's why RDBMS is used in OLTP (online transaction processing), where you need real-time, accurate data to perform different tasks, such as inventory management, order processing, and customer relationship management.

O Data warehouses

While relational databases let you deal with continuous data that involves regular updates, they aren't suitable for batch processing large volumes of data. A common solution for handling large amounts of data is a data warehouse, like Amazon Redshift, Google BigQuery, and Snowflake.

Data warehouses help you store historical data in denormalized schema for faster processing when dealing with large amounts of complex data. They also aggregate data (storing values of sums, averages, and medians) to speed up query performance in OLAP (online analytical processing) operations. For example, you can use data warehouses to aggregate the total number of orders someone has placed over time.

That said, data warehouses are limited in their capability to handle real-time data. That's why many brands store some of their customer data in their marketing platform and use a data warehouse for broader analytics and for keeping a historical record of all data.

Besides that, similar to RDBMS databases, data warehouses can be overwhelmed by excessive queries. As a result, they're not ideal for workflows like personalization that require frequent data retrieval.

NoSQL databases

Depending on the source, data can be unpredictable. For instance, if you rely on third-party APIs to acquire data, you might not get data with a fixed pattern. Instead, each API would send data in a different format, which would also change with time as developers update the APIs.

In such cases, structured data stores, such as RDBMS and data warehouses, don't work. Instead, you need a NoSQL-based data store, such as MongoDB or Cassandra.

NoSQL databases are a class of database management systems that provide a mechanism for storage and retrieval of data that is modeled in a way other than the tabular relations used in relational databases. They are designed for flexible data handling and real-time analytics of high-cardinality data, which refers to datasets that have a high proportion of unique values.

For example, if you want to offer personalized recommendations to customers, you'll need to store complicated customer data, such as browsing history, device types, and product interactions, which might not fit into a typical structured database.

Queuing systems

Besides long-term data storage solutions, you also have data stores that house data momentarily in a queue. Queuing systems store changes and updates to data and help you perform event-based actions, sync data, and manage asynchronous tasks.

To understand how that works, say a customer adds an item to the cart but doesn't purchase it. That event will generate a message, which is stored in the queuing system until the customer checks out the product or doesn't for a specified duration and abandons the cart.

Examples of queuing systems include RabbitMQ and Apache Pulsar. Many marketing automation platforms use queuing systems to power automations.

Caches (or in-memory databases)

To reduce query load on RDBMS and data warehouses, caches (or in-memory databases) store the frequently accessed data to speed up and optimize data delivery.

For example, if your system frequently executes a database query to get data from an RDBMS, you can cache that. Similarly, you can cache pre-aggregated data stored in data warehouses for efficient retrieval.

Object storage

While data warehouses store structured data, you also need a data store for long-term storage of unstructured data, such as documents, videos, images, and social media posts.

That's where Amazon S3 and Azure Blob come in to help you store such data.

Typically, you can use object storage to store data backups and media files. But you can also use it to run deeper analytics that can scan unstructured data.

Graph databases store information using nodes connected by edges to create a graph-like structure. The edges represent connections between different data elements, making graph databases an excellent choice when it's crucial to understand relationships. For example, these data stores are often used for social media platforms, network analysis, and fraud detection.

That said, graph databases can introduce extra complexity and often require support from developers and analysts who understand graph theory. If your primary data goals are aggregating, querying, and reporting, then a relational database with SQL support would be a better fit.

∨ector databases

In a vector database, information is stored as vectors with a fixed number of dimensions that represent the data. It is often used to store information about media files, such as images, audio, and video.

Vector databases are ideal for running similarity searches. For instance, you can use them to identify products that are similar to one another based on features or to find documents that share similar topics and sentiments. As such, they are often used for recommendations, natural language processing, and Al.

O Data lakes

Data lakes allow organizations to store a vast amount of structured, semi-structured, and unstructured data in its original form without any predefined schemas. As a result, they're commonly used for storing unstructured raw data in its native format and structure.

Since data lakes can store raw data, organizations find them convenient for storing unstructured data before transforming it and moving it to other data stores.

Besides that, data lakes also suit organizations looking to run data analysis, machine learning, and other analytics workloads on large volumes of raw data.

Let's see how all these different data stores come together in Klaviyo's data architecture.

Klaviyo's data architecture

Klaviyo's differentiated approach for intuitive data access

The core of the customer data problem is disconnection—customer data and marketing applications aren't connected seamlessly. As a result, developers have to simplify rich data into simple datasets their applications can understand. So, marketers start at a disadvantage due to their tech stack.

Klaviyo approaches the customer data problem by creating a natural connection—similar to an iPhone and its iOS—by optimizing data infrastructure and building the end applications off that.

So, marketers can, as Cohen says, "access the information they know is intuitively there, but has been locked down due to the friction of the way machines have to operate."

High level: OLAP + OLTP

Klaviyo addresses the complexities of customer data by weaving a robust web of databases that offers flexibility in the backend data infrastructure. Instead of building everything on top of a Snowflake instance, it employs several different technologies appropriate for the use case.

Eric McCall, senior director of engineering at Klaviyo, says, "We're picking the best technologies and implementing on top of those."

So, while many customer platforms compute the data in Snowflake to get data analytics, Klaviyo has dedicated OLAP (online analytical processing) databases to serve analytical workloads.

Generally, a given data store **can't be optimized for everything**.

Parker states that, "Some data stores prioritize data flexibility—like making it easy to store and retrieve complex data structures (e.g., JSON) and high-cardinality data—which turns out to be really important for some use cases, such as event-driven automations and viewing detailed behavioral timelines."



Different use cases typically lead to using different storage technologies. This produces a classic design challenge for software applications: How can a given app be designed to be good at each of these?

MIKE PARKER, SR. DIRECTOR OF SOLUTION ARCHITECTURE, KLAVIYO

"Other data stores are optimized for querying in different ways across datasets, which is really important for other use cases, such as reporting, analytics, and machine learning. Different use cases typically lead to using different storage technologies. This produces a classic design challenge for software applications: How can a given app be designed to be good at each of these?"

"This is where Klaviyo excels by using a collection of integrated data stores, creative techniques to process and index data, and caching technologies to ensure that our application can handle transactional and analytics use cases. And even though this is incredibly complex and elegant, it is abstracted from end-users of the application—so in addition to eliminating typical tradeoffs between flexibility and performance, the platform is really easy to use. This triad of flexibility, performance, and ease of use is what makes Klaviyo unique."



This triad of flexibility, performance, and ease of use is what makes Klaviyo unique.

MIKE PARKER, SR. DIRECTOR OF SOLUTION ARCHITECTURE, KLAVIYO

In fact, Klaviyo uses a mix of the best RDBMS, data warehouses, NoSQL-inspired databases, queuing systems, caches, and object storage databases to serve both OLAP and OLTP (online transaction processing) use cases.

Every event Klaviyo ingests goes to three primary databases—RDBMS database MySQL for OLTP, ClickHouse for OLAP, and object storage via Amazon S3. And every one of those databases stores different details of the event for the desired use case.

McCall shares that Klaviyo "picks out the pieces of the event that are useful in a context and saves it to a more specialized database."

For example, say a visitor is browsing products at your Shopify store. Klaviyo will ingest that data via its integrations and let you develop real-time trigger workflows to personalize the customer experience due to its OLTP databases.

In particular, Klaviyo's relational databases, like RDS (MySQL), will provide low-latency access to granular data—customer profiles, browsing history, and production interactions—that marketers require to move their customers toward a purchase.

On the other hand, Klaviyo's OLAP databases help group customers who've browsed a specific product category, purchased a particular item, or interacted with marketing material—i.e., powering Klaviyo's **segmentation engine**.

Plus, since Klaviyo doesn't archive relevant customer data, unlike other marketing tools, its OLAP databases are also instrumental in analyzing customer behavior over time.

Director of Product Alexandra Edelstein explains, "What makes Klaviyo unique is that the platform stores all data, over all time. Instead of preaggregating information, Klaviyo stores raw data in ways that allow for richer analysis and processing." As a result, you get a more nuanced understanding of your customer base over time, which isn't available with other marketing tools.

In contrast, Braze only stores the product purchased from an order for the last 30 days.



What makes Klaviyo unique is that the platform stores all data, over all time. Instead of pre-aggregating information, Klaviyo stores raw data in ways that allow for richer analysis and processing.

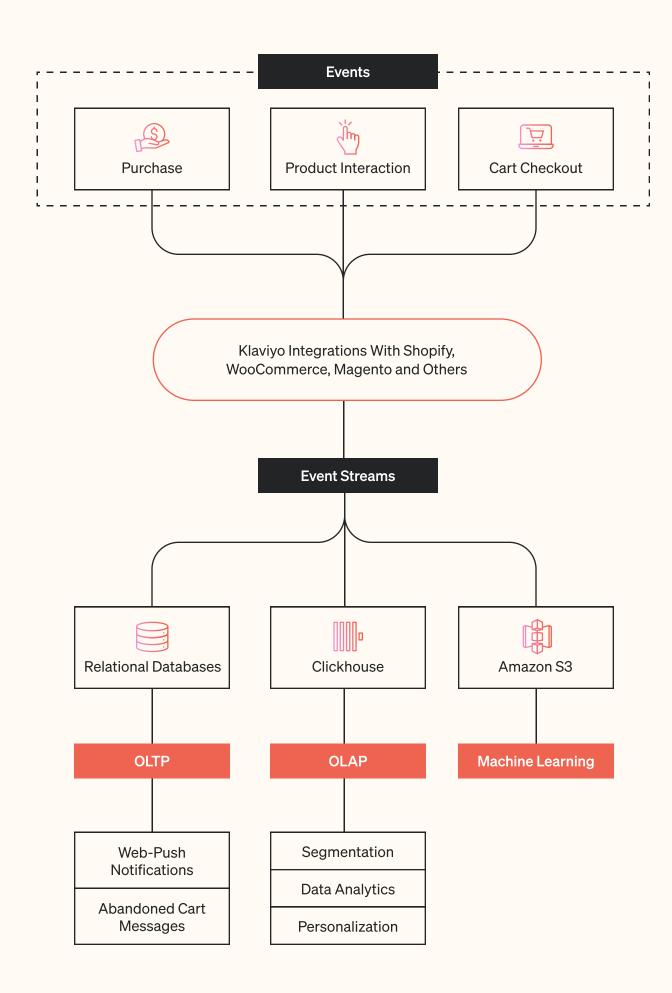
ALEXANDRA EDELSTEIN, DIRECTOR OF PRODUCT, KLAVIYO

Beyond that, you need to purchase a data extension to be able to look back up to 2 years. Bloomreach has similar challenges, billing customers based on the amount of data brought into their system and the total amount of data stored. For customers to control their costs, Bloomreach has extensive documentation about deleting their customer data.

The problem is that marketers and data teams don't have a crystal ball to know what data they will need in the future. So, while other providers have limitations on data due to their data infrastructure, Klaviyo is built to scale for the near-infinite data generated by today's brands.

Beyond that, Klaviyo simplifies the customer data equation by leveraging its roots as a database company. Instead of making marketers structure their data, Klaviyo lets marketers import data that best fits their business.

The reason? According to Cohen, Klaviyo's proprietary event pipeline "will go look at all that data and restore it in structured ways because the platform already knows how it's going to be used."



Deep dive:

Klaviyo's customer data infrastructure and the underlying event pipeline rely on multiple technologies communicating with each other. Let's explore those parts of Klaviyo's tech stack one by one.

Event streams

Event streams are Klaviyo's data source. Every digital customer interaction with your business generates an event. Examples include a customer checking out a product, clicking on a link in the email, and inquiring customer support for help.

Whatever the event, once it enters Klaviyo's pipeline, Klaviyo ingests it in real time by transforming it into different formats for machine learning, aggregation, and point queries and stores it in its relevant databases.

Every data processed by the Klaviyo engine is outputted into separate branches to serve the end-use cases of customer data.

OLTP

Events are generated throughout the customer journey—whether a customer visits your websites or leaves a review via your Android app—and you must record such instances in your data infrastructure.

Klaviyo handles this granular data via MySQL, as it helps quickly retrieve information. So, whenever a marketer wants a list of everything a customer purchased from your brand last month, MySQL can source that information with very low latency.

OLAP

When the data is of high cardinality, developers typically turn to NoSQL databases for flexible data storage. That said, Klaviyo addresses the flexibility factor without depending on NoSQL databases.

While Klaviyo used to have a NoSQL database, the platform has been switched to a combination of MySQL and ClickHouse (an OLAP database) to do everything a NoSQL database does but better.

In particular, since ClickHouse is more suited for analytical tasks as an OLAP database, it powers Klaviyo's segmentation engine without the high computational footprint of a NoSQL database.

Caches

Klaviyo uses key-value storage solutions via Redis and Memcached for frequently accessed data to reduce latency.

For instance, Allen Chaves, CTO at Klaviyo, shares that users typically access the "last 10 orders, not the last 1,500." So, Klaviyo put similar frequently accessed data in Redis and Memcached to speed up data operations.

And it's not just limited to relational databases. Instead, Klaviyo places caching mechanisms wherever there is data that finds frequent use—whether it is ClickHouse or even APIs—to lower the latency to the data.

Queuing systems

Customer-generated events don't directly go from an event stream to a database. Instead, they pass through temporary data stores or queuing systems before getting processed.

Previously, Klaviyo used RabbitMQ as its queue, which was a robust messenger broker but couldn't keep up with Klaviyo's 20+ systems behind the scenes. It was having reliability issues with high-volume publishing.

That's why Klaviyo adopted Klaviyo Messaging Service (KMS), which can handle more volume than RabbitMQ and is centralized—so internal teams no longer need to worry about operational aspects of queuing, as was the case with RabbitMQ.

KMS primarily uses Pulsar, which is better suited for event-driven workloads a customer data solution deals with. Still, it's not entirely dependent on Pulsar; it's also using Kafka to supplement it.

Manage your data the easy way with Klaviyo

Spend less time on development and more time on marketing

But what does Klaviyo's tech stack offer a business trying to stay ahead of the competition? Primarily, it's ease of use and time to value. Instead of depending on data engineers, working extra time to clean the data, or investing in expensive third-party integrations, your marketers can turn to Klaviyo to go from data to insights within milliseconds.



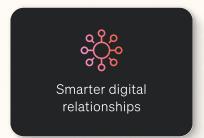
We engineered things the hard way, so marketers can do things the easy way.

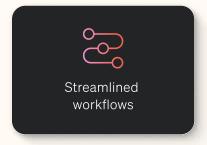
MIKE PARKER, SR. DIRECTOR OF SOLUTION ARCHITECTURE
KLAVIYO

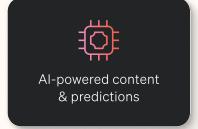
In particular, Klaviyo offers the following sweeping benefits:

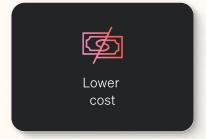














Easier data centralization

If you have to clean up the data during data ingestion, your customer data infrastructure isn't working for you. Instead, you're working for the data platform.

With Klaviyo, that's not the case. It comes with layers of abstraction, so you don't need to predefine all schemas to store data. Klaviyo will take the data from different sources, normalize it, organize it, and make it available for use in marketing, analysis, or later use.

So, whether you have customer data coming from integrations, direct APIs, or bulk files, Klaviyo will ingest that data and keep it in hot storage—available for fast, frequent access—indefinitely so you can put it to use whenever you need it.

In fact, that's how Klaviyo was born—a place to store data. According to Chaves, it'll remain that way because Klaviyo never "archives or deletes relevant customer data. Instead, you have a centralized place to put all your data, with no concerns of volume or speed to retrieve."

So, whether you're a growing ecommerce store or a large enterprise, you can find Klaviyo convenient for your data needs since it can scale up according to your requirements without harsh limitations.

On the other hand, some marketing platforms built through acquisitions—e.g., Salesforce Marketing Cloud—don't accomplish data centralization in a similar way. Instead, they require data to be moved around in the platform because all the acquisitions aren't operating off the same unified store.



Deeper customer insights

With centralized data, your distributed teams get access to the entire customer journey—from how customers engage with your emails and respond to flash sales to what makes them convert and become loyal advocates of your brand. Klaviyo includes numerous ways to analyze your data through flexible reports and rich visualizations, with pre-built and custom report builders.

That said, Klaviyo isn't limited by what your customers are telling you. Klaviyo also lets you read the market via personalized benchmarks. With the anonymized data of more than 130,000 brands in its database, Klaviyo can tell a brand how it compares to its peers.

So, instead of comparing apples to oranges, you get to compare apples to apples. If you're a beauty brand with \$1 million annual revenue, you can look at how your revenue per recipient compares with that of similar brands to see where you stand rather than checking vague benchmarks.

In contrast, brands such as Iterable, Cordial, and Braze only have a few thousand customers spread across multiple industry segments, so there isn't the same wealth of data to provide these insights.



Smarter digital relationships

McKinsey & Company shares that brands that excel at personalization get <u>40% more revenue</u> than their average counterparts. And since omnichannel marketing lets you approach customers via multiple touchpoints, you can depend on it to create a personalized customer journey.

McKinsey & Company shares that brands that excel at personalization get <u>40% more revenue</u> than their average counterparts. And since omnichannel marketing lets you approach customers via multiple touchpoints, you can depend on it to create a personalized customer journey.

That said, omnichannel marketing remains a tough concept to implement, as you must analyze several identifiers—phone numbers, emails, and app IDs—to develop profiles.

And that remains a challenge.

"One of the things that a lot of marketing tools get wrong is that they might treat each of those identifiers as a distinct person," says Parker.

So, instead of omnichannel marketing powering your personalization, it may work against you. You might be sending duplicate messaging across different channels without even knowing that you're doing that.

However, Klaviyo enriches a single profile with all the different identifiers to ensure nothing gets missed using **identity resolution**.

To get into the weeds, Anthony DelPizzo, lead product marketing manager at Klaviyo, phrases it as: "Klaviyo will take data from any source across a brand's tech stack, using a pre-built integration, a custom-built integration with our APIs, or a flat file. Without needing to predefine your metric schemas, Klaviyo will automatically normalize and unify that data into single customer profiles, using identity resolution—specifically, deterministic merging."



Streamlined workflows

If you're struggling with below-par marketing results, look into how much time your marketers actually get to spend on marketing operations.

For instance, say you rely on enterprise marketing software like Salesforce Marketing Cloud. In that case, your marketers will likely need to work with a cloud administrator who writes SQL queries whenever they want to build an automation.

They might also depend on a templating expert to write email templating code to send bulk emails.

And these are just two examples of your marketers spending time explaining their requirements to two specialists instead of finalizing the upcoming Black Friday email campaign.

As a result of Klaviyo's smart customer data infrastructure, marketers can avoid relying on other teams for core marketing operations. They can use one-click integrations, pre-built automations, and a library of templates to get their work done in seconds.

"A marketer can have an idea in the shower and execute on it that same day," says Parker.

Titan Fitness, a brand that <u>saved 75 developer hours</u> monthly by switching to Klaviyo, said that "Klaviyo's ease of use and extensibility with apps empowers team members to do their jobs."

Titan Fitness saved 75 developer hours (about \$15k monthly) by switching to Klaviyo.





Al-powered content & predictions

As a customer data platform, Klaviyo has the data it needs to have the upper hand in all things artificial intelligence. In particular, since Klaviyo never archives relevant customer data, you can use intelligent algorithms with years of data in their backend to make strategic decisions.

While you can see various intelligent tools throughout the UI, major <u>AI-powered processes inside</u> <u>Klaviyo</u> include:



Faster content creation

Klaviyo makes the creative process easier with generative Al. Users can create winning content faster, like SMS campaigns or email subject lines, by typing a short prompt.



More personalized marketing

With machine learning algorithms analyzing historical customer data, Klaviyo suggests relevant products and the best time to send messages to get buyers to convert.



Purchase prediction

Klaviyo's predictive analytics use abundant customer data to determine who will buy when, the lifetime value of any customer, and more.

In addition, Klaviyo has Segments AI, which uses natural language processing (NLP) to enhance the segmentation engine's ease of use. Customers simply need to describe the people they want to reach, and Klaviyo will define the segment instantly.



Lower cost

Klaviyo removes the need to operate multiple data systems. You don't have to set up instances of Snowflake or Google BigQuery, define complex databases, or configure data synchronization.

Plus, you also optimize your costs since the technical labor won't need to spend all their time on helping marketing teams execute their day-to-day tasks. Instead, they'll be able to focus on more needlemoving projects.

According to Forrester's TEI Report, Klaviyo's API functionality and ease of use enables the reallocation of both graphic designer and developer labor. This allows the composite organization to realize \$392K in savings.

Migrate to Klaviyo without slowing your team down

When it comes to Klaviyo, robust doesn't mean slow.

You can get started with Klaviyo in a matter of days, if not hours, without parting with your historical data.

All you need is to import data, whether structured or unstructured, into Klaviyo, and it'll ingest it in a few hours to a day. Even millions of historical orders can get imported in less than 24 hours.

And in less than 30 days*, you'll have your intelligent marketing platform sending messages, with live flows, IP warming, and end-to-end integrations.

Store and manage your data more effectively to power smarter digital relationships.

Get a demo