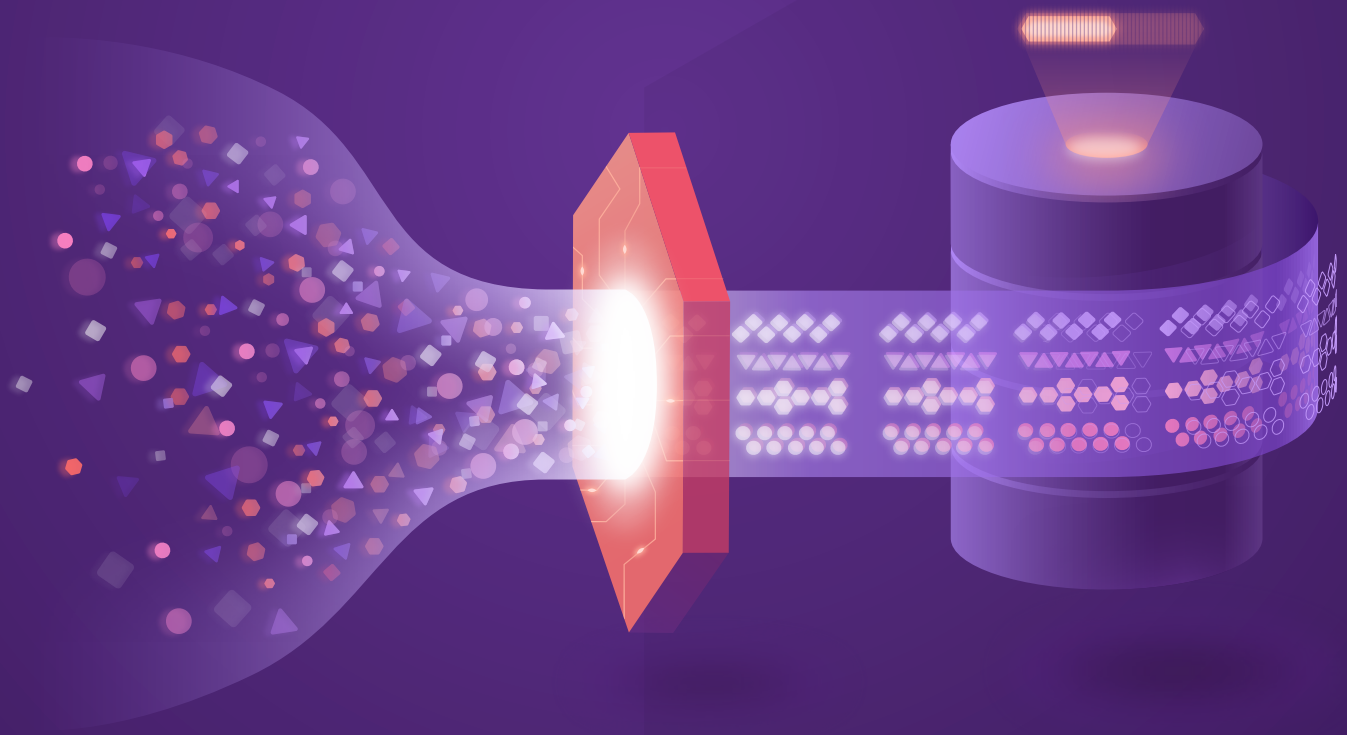
WHITEPAPER

The Death of Web Analytics

and the Rise of the Build-Your-Own Customer Data Platform



HEAP

Contents

Introduction

Web analytics as you knew it is over. The complexity of customer data has evolved beyond a one-size-fits-all solution. And innovative data teams have taken the mandate to build the infrastructure to make customer data a competitive edge.

The past two decades have seen three evolutionary waves in analytics infrastructure. The first wave saw the visit paradigm focused on static websites for marketing teams. The goal was to measure site efficiency by tracking page views and clicks.

The second wave saw the shift to the visitor paradigm, expanding stakeholders to include both marketing and product teams. This optimized acquisition channels, adding mobile and other touchpoints like email, A/B testing, and support.

The third wave takes us to the present, as the data and insights layers expand beyond just the web. Specialization leads next-generation data teams to take on managing all customer data. Siloed web data becomes automatically captured in centralized cloud data stores that combine and mine insights across multiple sources. The great casualty is the single-purpose web analytics tool, as companies consolidate funnel and retention reports into business intelligence (BI) tools.

This leads to the new the rise of building your own or it should be the rise of build your own CDP's, which integrates cloud data stores and BI tools with a new layer of automation that makes it possible to efficiently capture, organize, and govern customer data. As customers become more digitally native, data teams must adjust their analytics approach — by leveraging the accelerating volume and complexity of customer data. Teams that do this successfully become the "customer whisperers" in their organization. They help their business respond with agility, bring products and services to market faster, and focus on what matters to stay ahead.

This paper is intended for high-growth enterprises looking to build and own their data infrastructure. The content will help this audience understand the evolution of analytics and how to ride the third wave to a competitive advantage.

Cloud Economics have Changed Tracking Customer Data

Historically, choices in storing customer data were limited to a siloed, single vendor solution or expensive, monolithic infrastructure.

In the beginning, early websites were electronic brochures managed by marketing departments, and efficiency was fundamentally based in page views and clicks. As companies shifted towards optimizing acquisition channels, tracking visits shifted towards tracking visitors. Companies tried to understand how customers engage via mobile devices and email. As visitors started to be grouped in cohorts, A/B testing tools emerged, and product teams joined marketing teams in measuring end-to-end funnels and retention. Multivariate testing tools worked together with purpose-built analytics tools to optimize online conversion rates.

But as user journeys become more complicated, and data sets become very heterogeneous and difficult to integrate, individual vendors can't keep up with the pace of integrating each customer's custom data. As cloud services from Amazon, Google, and Microsoft drive down compute and storage prices, it has become cost-effective to track every customer touchpoint. This includes web click-stream data, offline transactions, and historical transactions. And this data can now be joined, analyzed, and acted upon in real time. **As market leaders make data competencies a differentiator, companies realize they need to own their own destiny and invest in the infrastructure to build and own their Customer Data Platforms.**

Analytics Maturity Curve

Assess your path to leadership.

Companies are only as good as their data infrastructure. To measure and improve your competitive position, benchmark yourself against your peers and competitors with the Analytics Maturity Curve. Identify where your company is in this progression and what you're missing. Heap can help you take the next step to become a data market leader.

1. **Measuring Traffic**

   The first step for companies is to simply understand how many people are accessing various parts of your site. This is a static aggregate view and lacks the information required to drive real change, especially information about visitor cohorts and their behavior.

2. **Optimizing Acquisition Channels**

   As companies gain understanding of what users do when traffic lands on their site, they begin to separate out high-value traffic. They use these groupings to optimize what traffic drivers are best for the business.

3. **Drive User Engagement and Conversion Rate**

   As companies optimize for naturally converting traffic, they start to look at how groups of visitors interact with the site, where they drop off in that funnel, and begin to experiment to achieve higher conversion rates. This is typically the departure point where companies have to abandon a page view model for a more sophisticated user - event model, looking at granular behaviors organized by individual users and groups of users. Each event and user property becomes a building block for segmentation and personalization.

4. **Understand the Customer Digital Journey**

   This stage is where companies realize their user event schema can give them on-page interaction data, but the reality is the customer's journey rarely begins, or ends with just website interactions. This is the stage where companies look to integrate, and combine digital touchpoints from various parts of the customer journey to create a more complete image of where they can focus to optimize.

5. Unsilo Data and Create a Unified Customer View

At this level, most companies find combining disparate datasets with ETL pipelines and cleaning and organizing large amounts of downstream data significantly impacts time and resources left for customer insight and action. This level requires that companies find a durable, automated way to combine customer data in a clean, virtualized schema down stream. This is where companies look to build out their customer data infrastructure in a scalable and powerful way.

Core Data Investment = Competitive Advantage

Leading cloud-first companies invest in a unified customer view.

Now compare your maturity stage to cloud-first companies. They have deeply invested in the core data infrastructure required to create a unified customer view (see chart below for gross salary and headcount in data-related hires). This creates a virtuous cycle in value and growth by leveraging data to deeply understand their customers. It also opens the ability to more rapidly leverage emerging trends in machine learning and artificial intelligence.

This competitive advantage was previously only available to the most sophisticated companies. But recent technology changes in cloud compute and storage economics have made this available to a broader set of companies, even with smaller data teams. It is only recently that inexpensive compute and storage have become ubiquitous enough to drive down the costs of data enough to store basically everything. Upfront costs have plummeted, and the costs of maintaining these data stores have transferred to cloud vendors. Because you pay for what you use and scale up quickly, limits on storing customer data have been removed. This economically opens the door to capturing a complete picture on a customers behavior—now within reach for even small data teams.

Power Business Decisions with Customer Truth

Your business decisions should be guided by understanding how your customers really behave.

Companies leveraging behavioral data are emerging as key industry leaders. According to Gallup, companies that leverage behavioral insights see 85% higher sales growth and earn 25% more in gross margins[1]. Why? Your customers are also more likely than ever to interact with your brand or product online. Customer interactions with brands through a website or mobile have grown 67% and 52% respectively since 2017[2]. Eighty percent of North Americans make purchases online[3] and mobile traffic comprises more than 50% of web traffic[4].

To optimize your bottom line, best-in-class companies develop a full picture of the customer in order to deeply understand these online and in-app behaviors, and how they fit into the broader customer experience. You can optimize your channel spend or identify key cohorts you can target with email campaigns or coupons, as well as more deeply understand attribution or predict customer lifetime value.

Customers expect consistent experiences across digital channels. When they call for customer support, they expect to talk to a representative who has complete context on their customer journey. Point-of-sales systems and call centers can only give an incomplete point-in-time snapshot without the surrounding behavioral data that leads up to a purchase or support call. According to Accenture, 48% of customers will bounce if they have a bad online experience[5].

Finally, technology has emerged that allows you to quickly integrate this data and quickly capture, organize, and analyze a complete view of your customer. You can accelerate your core advantage in data by building your own Customer Data Platform.

[1] *"Behavioral economics," Gallup, gallup.com*
[2] *Kitewheel, State of the Customer Journey 2018*
[3] *Pew Research Center, http://www.pewinternet.org/2016/12/19/online-shopping-and-e-commerce/*
[4] *Statista, https://www.statista.com/statistics/277125/share-of-website-traffic-coming-from-mobile-devices/*
[5] *Accenture Interactive's 2018 Personalization Pulse report*

Build-Your-Own Customer Data Platform

Integrate best-of-breed cloud data warehouse and visualization tools.

Early Customer Data Platforms (CDPs) targeted marketers who didn't have the engineering resources to make a substantial investment in a data platform, but the best companies are building their own CDPs that target their internal data teams' unique requirements.

There are two types of CDPs:

1. **All-in One CDP for Marketers**

   These were the first wave of CDPs, originally targeting small to medium size businesses and non-technical marketing managers. A primary goal is to automate and connect MarTech tools for companies that lack core data analytics teams. These vendors provide end-to-end services, including how data is tagged, stored, and federated across marketing channels. These marketing-focused CDP's are ideal for companies with minimal in-house data resources.

2. **Build-Your-Own CDP for Data Teams**

   These empower larger, fast-growing enterprises with data teams who want to build and own their own data infrastructure to drive digital transformation. A primary goal is to enable these companies to build data as a core competency and in-house competitive differentiator. These CDP's built in-house by data teams leverage best-of-breed visualization and data warehouse components, understanding these industries have deep investment and are innovating rapidly on independent trajectories.

Unlocking this second category has been held back by technical challenges in capturing, organizing, and connecting data sources, but that has changed recently.

	All-in-One Marketing CDP	Build-Your-Own Infrastructure CDP
Data Available	• Data limited to digital journey (web, app, email, CRM)	• Contain all relevant sources of data, customized to your business
Company Size	• Target small to mid-sized companies with limited developer/data resources	• Large, fast-growing companies and data centric organizations
Team Ownership	• Marketing	• Data
Benefits	• Automated • Non-technical users	• Customizable • Contain complete data set
Cons	• Black box • Limited data set • One size fits all	• Requires developer work
Primary Goal	• Automate and connect MarTech tools for companies that lack core data analytics teams	• Build data as a core competency and in-house competitive differentiator

Challenges in Build-Your-Own Customer Data Platforms

Key challenges that have made data infrastructure more inefficient and unreliable include missing data, poor instrumentation, lack of standardization, complex identity resolution, and fragile pipelines.

Missing Data: As business needs evolve—especially in experimentation-driven cultures—it is impossible to brainstorm every nuanced question your team or company might face in the future. Spending sprint cycles brainstorming a list of "what-ifs" can detract from analysis, action, and driving towards KPIs. However, when these gaps in the data arise, teams become less reliant on data and increase reliance on gut to make hypotheses.

Poor Instrumentation: As experimentation/iteration accelerates, your in-app, email, website, and experimentation data is constantly changing. Keeping up with these changes and maintaining data collection pipelines has required extensive engineering resources.

Lack of Standards: **Without clear standards, joining disparate data sets with different schema is inefficient and costly.** Data from email is delivered in a different schema from data from the web and experimentation tools; and all the data is arriving in your warehouse at different times, making it hard to reconcile how up-to-date your insights and actionability are.

Complex Identity Resolution: Tying cookie-based data to credit bureau information for example can be hard, especially when you're trying to reconcile anonymous, pre-conversion behavior with an authenticated behavior or browsing behavior with email behavior. Many data teams spend significant internal resources implementing advanced identity resolution.

Fragile ETL Pipelines: Up to 80% of a data team's time can be spent cleaning, integrating, and maintaining data pipelines to create a holistic user view. The cost of incorporating behavioral data can consume a significant majority of a company's data resources, decreasing the amount of time spent on delivering insights, taking action and developing predictive models.

The Modern Customer Data Architecture addresses these pain-points by leveraging recent advances in cloud data warehouse and BI technologies.

The Modern Customer Data Architecture

Building your own Customer Data Platform requires three distinct components: **Your Best-of-Breed Cloud Data Store**, **BI Visualization tools**, and **Customer Data as a Service**.

1. ## Cloud Data Store

   Behavioral customer data should be collected in your choice of cloud data store. Upstream changes should be seamlessly synchronized with your source of truth as a first class citizen. ETL/pipeline management is automated: as event definitions change, data changes are retroactively adjusted in the data store.

2. ## BI Visualization Tools

   Visualize data insights across multiple sources by leveraging your multi-data BI tool of choice. Because data has been cleansed, organized, and verified before being modeled, BI tools can easily incorporate these data without worrying about integrity.

3. ## Customer Data as a Service

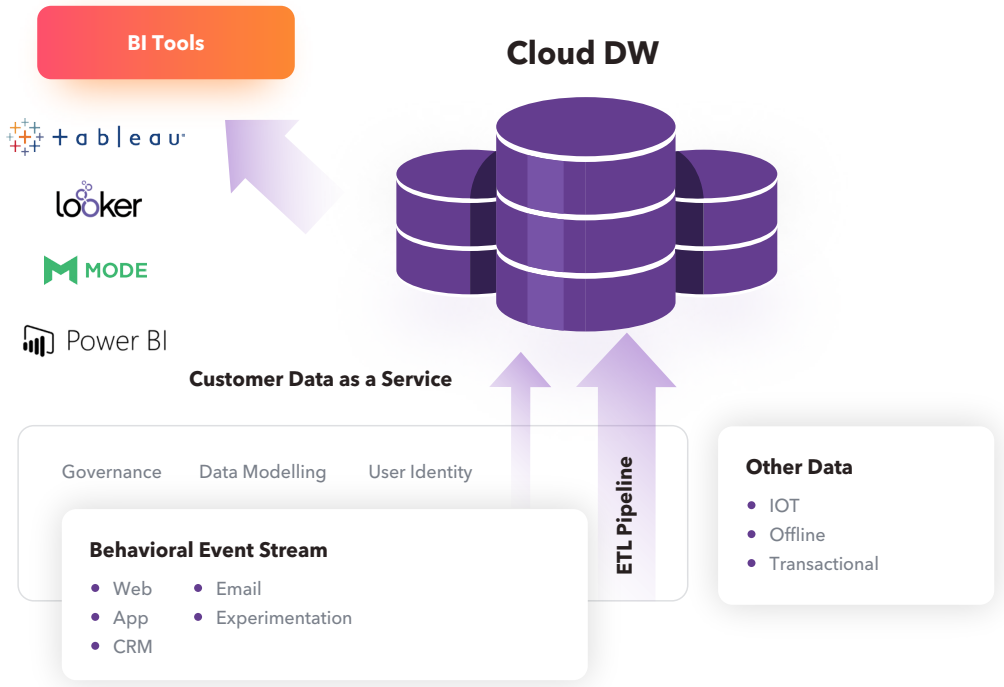
   Automates how behavioral data is organized, verified, and connected with your cloud data store. This abstracts away the annoying parts of analytics, including cleansing data, maintaining manual pipelines, integrating data, and verifying data integrity.

	Legacy Web Analytics	Build-Your-Own CDP	
Visualisation / Insights	One-off Visualization Tools (i.e. Web only, Product guides only)	Best-of-Breed BI Visualization Tools	looker, tableau
Cloud Data Store	Proprietary Data Stores	Best-of-Breed DW or Data Lake	snowflake, Google Big Query, amazon REDSHIFT, Azure
Customer Data as a Service	Limited One-way DW synch, Low Perf Fragile Data Pipelines Fractured User ID's Missing Data	Retroactive Synch Virtual Events Unified User Identity Verified/Audited Automated Data Capture at Scale	HEAP Customer Data as a Service

How Customer Data as a Service Works

Customer Data as a Service is the best way to bring clean, complete customer data into your company's source of truth. It autocaptures imports critical behavioral event streams, including web, app, email, CRM, and experimentation data. It fully cleanses and normalizes the data against a user-centric schema, and allows you to model and verify the data by separating event definitions from raw data. Finally, it abstracts away the annoying and error-prone parts of managing ETL/data pipeline processes, freeing up resources to focus on higher order, more valuable data science.

	Without CDaaS	With CDaaS
Data Collection	• Manual integration throughout app, website • APIs to extract data from third party tools	• Automated collection across web, app, experimentation suite, ESP, and CRM • Optional APIs for data enrichment
Identity Resolution	• Manual data standardization and identity resolution where possible	• Advanced APIs automatically resolve identities across devices, channels and platforms
Data Modelling	• Requires extensive joins and transformations downstream • Frequently results in missing data	• Allows you to combine, and update events and property mappings • Updates warehouse with retroactive data
Data Governance	• Characterised by a lack of standard governance features across platforms	• Governance features incorporated into platform • Advanced permissions, verification, inactivity notification and more
ETL	• Requires engineering resources to build and maintain data pipelines and data transformations	• Managed ETL process • Schema automatically standardized • Data retroactively updates to reflect any upstream changes



Introducing Heap Customer Data as a Service

Initially, Heap was founded to automate how web and mobile customer data sets are captured and analyzed. The company has since evolved to help data teams automate how they capture, organize, and connect customer data into their source of truth. Built by data teams for data teams, Heap innovates around collection, identity resolution, virtualization, governance, and simplifying data pipelines.

Autocapture a Rich Set of Event Stream Data

Heap pioneered Autocapture to provide a complete and retroactive data set. With Heap, you can automatically capture event stream data from a variety of sources such as web, app, and email, and standardizes them into a user centric model. Capture every click, tap, swipe, pageview, and more without code.

Consolidate Omnichannel Journeys with Best-in-Class Identity Resolution

Consolidate a customer's entire digital history by merging anonymous behavior with authenticated behavior. Seamlessly join data across channels, devices, and platforms with Heap's advanced Identity APIs. Create multi-tiered identities and update identities as user progress along their digital journey.

Retroactively Model Events with Virtualized Data

Virtual Events retroactively name and manage any customer interaction without touching your codebase. This enables advanced data governance with an audit trail: always have access to what an event means, who defined it, and when it was last updated. This eliminates the need to dig through your site's HTML or reference an outdated tracking plan.

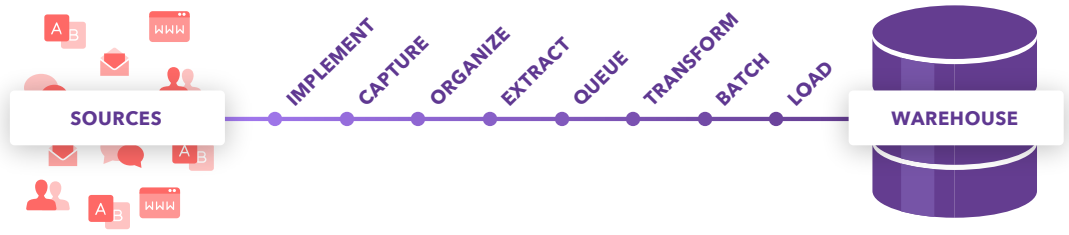
Build Trust with Data Governance

Data integrity is at the core of adopting and utilizing data. Heap offers a suite of features to ensure the data is accurate so you can trust the decisions you make. Monitor event changes with an audit trail. Identify site changes and broken data with Inactivity Notifications. Align your team with living event definitions and annotations. Control your environment with advanced team and role based permissions. Eliminate the opaqueness of Excel-sheet driven tracking plans and unclear ownership.

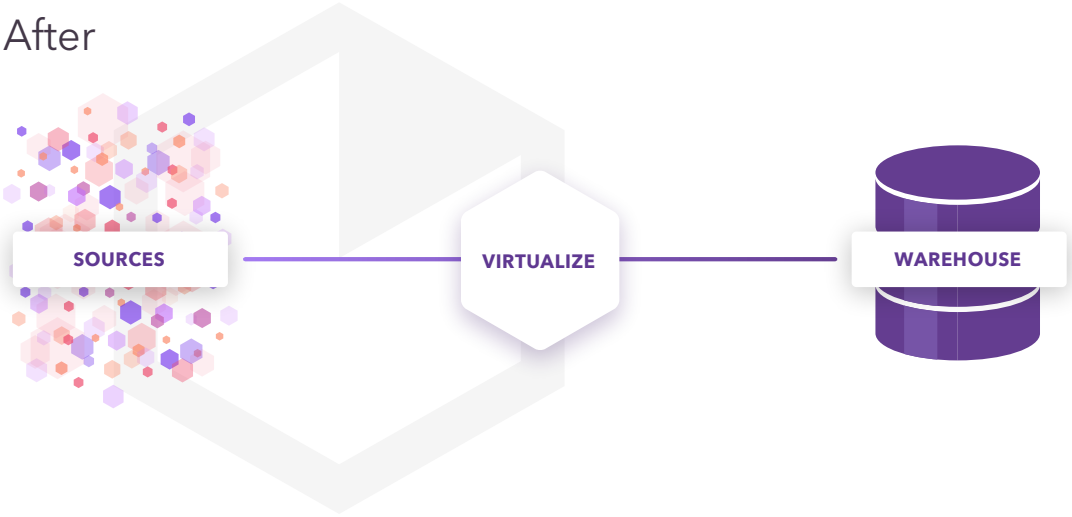
Spend more time on insights, not ETL plumbing

Heap Connect controls the flow of data from your site, app, and more to your data warehouse or data lake, so you can focus on developing insights not maintaining pipelines. Heap syncs retroactive data and updates event and user tables with newly added and modified information. The standardized schema makes it easy to join clean behavioral data with all your other sources.

Before



After



Learn more about Customer Data as a Service

Learn more about how Heap automates bringing behavioral data into your cloud data store with Customer Data as a Service at www.heapanalytics.com/cdaas.

WWW.HEAPANALYTICS.COM