How to Build your Customer Data Technology Stack

HEAP

Contents

Why Your Customer Data Technology Stack Matters

WHY YOUR CUSTOMER DATA TECHNOLOGY STACK MATTERS

Customers are everything these days. Better understanding your customers is a high impact way to improve your business – it allows you to increase conversions on your website, improve the customer experience of your product, measure the impact of your new initiatives, and uncover new sources of revenue. In today's world, relevant customer data is distributed across many different tools and silos.

Let's say a customer of yours sees an ad for your product on TV, and then a few days later notices your Facebook ad. Intrigued, they Google your product name and stumble across a blog post of yours. They decide to sign up, and start getting drip emails delivered by your email service provider. After a while, they run into an issue and hop on the phone with a support rep. Each of these interactions with this customer and your business lives in a different silo. To deliver a holistic customer experience across these tools, you must have a customer data technology stack that is intentionally designed to achieve this goal. When you build it well, you can break down data silos and emerge with a clear and unified view of your customers.

Generally, there are four activities that your stack should enable you to do:

① **Gather relevant customer data.** This includes behavior data on your site or app, interactions across your touchpoints, transactional data, and anything else that contributes to creating an actionable 360 degree view of your customers.

② **Organize your data.** All the data from step one is captured and stored in different formats and schemas, as well as in different tools. To make it usable, it must be unified and organized into consistent formats.

③ **Analyze your data.** Once you have usable data to work with, your stack should have the power to answer your questions and deliver you actionable information that you can use to improve your customer experience and drive revenue.

④ **Act on your data.** Once you know what you should do, your stack should enable you to take action. This might mean following up with and nurturing high impact customers, solving customer pain points in their engagement with you, or optimizing your website.

So how do you actually go about making this a reality? There are lots of tools out there that do different things, and it's not always clear what approach or types of tools will work best to meet your needs. Different types and sizes of businesses will be best served by different combinations of tools.

For the purposes of this ebook, we are looking at how small to mid-size businesses can build out the type of stack that will yield the greatest returns on their investment and scale with them as they grow. We recommend a cloud-native customer data technology stack for this purpose. We'll be looking at the different layers that make up such a stack, and at the notable cloud-first or cloud-available tools that you can start your technology evaluations with for each layer.

We wrote this ebook to provide a clear overview of the different layers of your customer data tech stack — the types of tools you need, how they interact with each other, and who the major players are for each layer. The goal is for you to walk away with clarity around which types of tools are important for your needs and who should be part of your considerations when creating your tech stack, based on the particulars of your company and your goals.

Reasons why you should build your own customer data technology stack

Before diving into the meat of the book, it's important to first touch on why this is necessary in the first place. The goals here are two-fold: why does customer data matter? And why should I build out my own tech stack rather than getting some full-stack solution and calling it a day?

First, customer data. There is so much that your customers tell you with each interaction they have with your business. Through their actions on your website or app, your customers are telling you things like "this signup is too burdensome and not worth the effort", "I love this piece of content and want to buy what it's pitching", "this suggested content is super helpful!", "this suggested content is irritating!", and so on. On top of this, customers engage with you in an omnichannel manner. They read and write emails, contact customer support, see your ads, engage with your social media, and more. Gathering and

compiling a clear view of your customers across these channels leads to happier customers and a better bottom line.

Second, why should I build my own tech stack and not buy everything from one vendor? The reality is, no one full-stack solution can handle this for you. Your customers interact with you in an omnichannel way, and as such, data about them lives in many different tools. Each of these tools solve a piece of the customer data puzzle. A best-of-breed approach that is targeted to your needs is much more effective than attempting to rely on one vendor to solve everything, not to mention that the one vendor approach runs the risks of vendor lock-in and the downsides that come along with that.

There are some challenges getting started with a best-of-breed approach however. There are many tools out there, each looking to handle part of your stack. They approach problems from different angles with different target customers in mind. Despite this, many will list the same benefits and give similar pitches, so how can you actually tell which is right for you?

Choosing your tools and building your stack requires 1) understanding what the end-to-end flow of customer data looks like for your business, and 2) what your primary use cases are so you can choose the tools that are best suited for your specific needs. We recommend a cloud-native technology stack for small to mid-size businesses, but the mix of tools will depend on the specifics of your business and environment.

Considerations and things to think about when evaluating tools and technologies

There are many different ways to build a stack. As you'll see, a lot of the tools we cover below have different core strengths and many play across multiple functions. In practice, this means that you'll usually have some redundant functionality across tools, or will implement two tools that do 80% of the same thing, because the 20% that they do differently are both important for your stack.

The takeaway here is that it's important to consider what factors are most important for you in your stack when choosing your tools. Your stack will end up looking different if price is your driving factor vs simplicity vs robust functionality. Stacks that need to join 80 sources of data and want them to talk to each other will look different than stacks with 5 key sources of data

WHY YOUR CUSTOMER DATA TECHNOLOGY STACK MATTERS

where the main focus is to have actionable, organized data in the data warehouse.

One more thing to keep in mind is to think about not only how to enable the analytics and visualizations you want to do, but also how the different tools you implement are going to connect and work together. Despite the language commonly used around data pipelines and data flows, data is not water. It comes in different forms and schemas, and you have to ensure that you can handle and unify all of these different data shapes and sizes together. It's not something that happens automatically.

Storing and Querying Data: Data Warehouses

If you ask people to diagram their customer data technology stack for you, you'll often see a data warehouse in the middle of the picture. There's a reason for this, as most data will end up flowing into and out of the data warehouse at one point or another. When thinking about designing your stack, starting with the data warehouse that you'll want to use is a reasonable jumping off point. Your choice of warehouse will impact some of your decisions for other layers, as a strong integration between a tool and your warehouse of choice is something you should look for when evaluating other tools. Not all tools support all warehouses equally.

The purpose of the data warehouse is to store your data in a structured manner and give you one unified foundation to model your data and run reporting against. The data warehouse options we cover here are all relational databases, meaning that they have a processor component with them that presents the data in a relational format when it's queried. One point to note

is that there are data lake options like Amazon S3 that serve a similar purpose, but store your data completely unstructured, which changes the requirements for your stack. There are also non-relational databases and database management systems. For the purpose of this ebook, we'll focus on cloud data warehouses (i.e. relational, cloud-based, columnar store offerings), as they're considered best practice for building a modern, cloud-native customer data tech stack.

Important Considerations for this Layer

Different decision factors matter more or less for different layers in your stack. As you determine which data warehouse option is right for you, there are several factors to consider that we'd recommend you keep in mind:

Who is your cloud provider, or are you provider-agnostic?

- If you use AWS, Azure, or Google Cloud Platform, each works best with their own data warehouse offering. Moving data from your cloud platform to your provider's data warehouse is simple if it's from the same provider, but requires an extra ETL step if moving to a different warehouse. Additionally, since AWS, Azure, and GCP

charge for data egress, there's a non-trivial cost involved with moving data from your cloud provider to a different provider's warehouse.

- If you're provider-agnostic, then this won't be much of a decision factor for you.

How will you most often use your warehouse?

- Different providers price in different ways. Some tools price per cluster, while others price per query. If you store a lot of data but don't query it that often, certain tools will be cheaper for you than if you don't store as much data but query it all the time. Pricing is not a straightforward determination for this layer, so having a clear understanding of your likely usage and usage patterns will help you determine which tool is cheapest for your exact needs.

How sophisticated and technical is the team in charge of your warehouse?

- Different tools offer or require different levels of configuration. For teams with less sophistication or

dedicated technical resources, a tool that handles more of the configuration for you could fit better. For organizations with a strong technical team dedicated to their warehouse, an option that offers more configuration and gives you more control could fit your needs better.

What about performance?

- Generally, the performance of the main data warehouse options is roughly similar for most uses although some providers are better as handling particular workloads (e.g. spiky workloads or consistent workloads). There are some differences when you get to the higher-end, optimized levels of usage, and for certain edge case requirements, but for many organizations, that doesn't have an impact on their current needs and likely won't for a while down the road.

- It is always worth considering your usage needs however, as you may fall into a category where you would see actual performance differences between some of the data warehouse options. Generally, we wouldn't recommend making small performance differences the leading factor for this decision however.

Modeling your data:
Beyond storage and queries

One important thing you can do within your data warehouse is model your data, turning it from raw data into fact tables.

Modeled data is extremely powerful for use in visualizations, advanced learning tools, ad hoc analysis, and getting actionable data out to other tools. In fact, it's a key functionality for any customer data tech stack.

There are a few technologies that give you the ability to create a virtualized layer of modeled data over the raw data in your warehouse. They include LookML, DBT, Matillion, and SQL Views.

We wrote a whitepaper on the power of data virtualization if you're interested in learning more.

Notable Tools in this Layer

As per the rest of this ebook, we'll be focusing our coverage of these tools towards ones that are cloud-native or at least well-suited to fitting into a cloud-native technology stack, with an eye towards the needs of small to midsize businesses.

Azure SQL Data Warehouse

Microsoft Azure SQL data warehouse is the market leader for enterprise cloud data warehousing. Launched in 2009, Azure SQL Data Warehouse is tightly integrated with the Azure cloud platform and has recently undergone a revamp to make it competitive with other cloud data warehouse providers along performance lines.

⊕ **Pros**

- Azure SQL data warehouse decouples storage and compute, so you can scale those separately according to your needs.

- Well-integrated with the Microsoft ecosystem.

- Allows for your compute clusters to be paused and resumed without affecting storage, which helps with costs.

⊖ **Cons**

- Query performance has long been a downside to SQL Data Warehouse, but earlier this year Microsoft

released a big catchup update that claims to have greatly improved performance and concurrency.

- They still have some ongoing design limitations around concurrency and real-time reporting.

Price

- Azure SQL Data Warehouse is priced based on storage and compute needs separately.

Best for

- Enterprises and large organizations.

- Microsoft shops, particularly those using Azure as their cloud provider.

- Organizations that have high storage needs but less frequent querying needs.

- Organizations that require other deployment options for their warehouse. Azure offers hybrid and private cloud deployments in addition to public cloud.

BigQuery

BigQuery, Google's cloud data warehouse, was made publically available in 2011. BigQuery handles provisioning and cleanup for you, as they operate a shared infrastructure across all their users. BigQuery differentiates around their simplicity of use.

Pros

- Well-integrated with Google Cloud Platform. Moving data into BigQuery from GCP is a simple matter.

- RESTful API calls. BigQuery is accessible via API calls, which is more familiar and easier to use for less technical users than database calls.

- Simplicity and ease of use: BigQuery handles provisioning, compaction, and other behind-the-scenes work for you.

- Makes scaling very simple, as BigQuery will scale up your clusters as required in the background.

⊖ Cons

- BigQuery offers less ability for users to configure and customize than other tools, as it handles a lot behind the scenes.

- Functionality: BigQuery makes some sacrifices on higher end functionality and performance in order to deliver a simpler and easier-to-use tool.

- Less predictable and more variable pricing than other available options.

⑤ Pricing

- BigQuery prices in an unbundled manner, charging separately for storage and for usage.

- BigQuery prices on a usage and per-query basis. You pay for the storage you use and the data you load, as well as a cost per query that you run, which is based on how much data the query touches.

- Loading, copying, and exporting data is free.

Best for

- Organizations who use Google Cloud Platform or who are provider-agnostic.

- Organizations with a lot of data they need to store, but not much need for frequent querying.

- Smaller companies or organizations with less sophisticated data teams or available technical resources, since BigQuery's simplicity and ease of use will have a bigger impact.

Redshift

Redshift, Amazon's cloud data warehouse, was one of the first to commoditize the data warehouse. Redshift offers extensive customization and configuration capabilities, leading to more potential as well as more complexity. A reasonable analogy is to think of Redshift like C++: lots of power and potential, but lots of complexity of use, since you are responsible for all of the configuration to make it run.

⊕ Pros

- Well integrated with AWS. Moving data from AWS to Redshift is simple.

- Strong performance and functionality. Redshift often delivers top results in benchmarking tests, thanks to the full freedom that users have to configure and customize it.

⊖ Cons

- Complexity. Because Redshift leaves configuration up to you, it's easy to set things up incorrectly or suboptimally. You have to manage storage space on your cluster yourself, for example, including manual vacuuming to remove deleted data.

- More challenging to get data in and out of than some other options in this space. Redshift requires you to connect and access it like a database, which is more complex than other tools' API-like access methods.

$ Pricing

- Redshift prices on a flat hourly rate per GB that covers storage and usage together.

- Redshift prices on a per-cluster basis. So you pay hourly for a cluster that you provision, regardless of how much you store in it or use it. There are no separate per-usage costs.

Best for

- Organizations that use AWS as their cloud provider and want to send a fair amount of data from AWS to their data warehouse.

- Sophisticated teams or organizations with a lot of technical resources to dedicate to their warehouse that value functionality and potential performance over the complexity that comes with it.

- Organizations where control over configuration and ability to customize is important to their business.

❄ Snowflake

[Snowflake](#) is an independent data warehouse that's not tied to any particular cloud provider. It is however tightly linked to the AWS ecosystem (S3 is what underlies it). As far as complexity and what's managed for you, Snowflake generally sits between Redshift and BigQuery on the spectrum.

⊕ Pros

- Advantageous for AWS users, given the underlying links to S3.

- Offers automatic starting and stopping of its warehouses based on when you actually use them. This is great for your costs if you have spiky usage patterns.

- Concurrency. Snowflake offers virtual computation warehouses that you can spin up on demand, which is great for running concurrent queries, as well as for customizing the type of compute you need for a given task. For example, you can use a slower (and cheaper) compute resource for an overnight batch ETL dump, and faster resources for small, urgent queries.

⊖ Cons

- Geographic availability. Snowflake is available in fewer locations than other tools on this list, so if you're not located nearby, you could see more latency when using Snowflake.

- Performance and functionality aren't quite as good as some other available data warehouse options in some cases. There are [benchmarking tests](#) that show Snowflake ahead for certain use cases, particularly around spiky usages, but for steady, consistent usage, other options will top the list.

$ Pricing

- Snowflake separates out storage and compute. It charges monthly per TB of storage you use, as well as your usage of the cluster on an hourly basis for running queries.

🤲 Best for

- Organizations that want more configuration and control than BigQuery offers but want more to be managed for them than Redshift delivers.

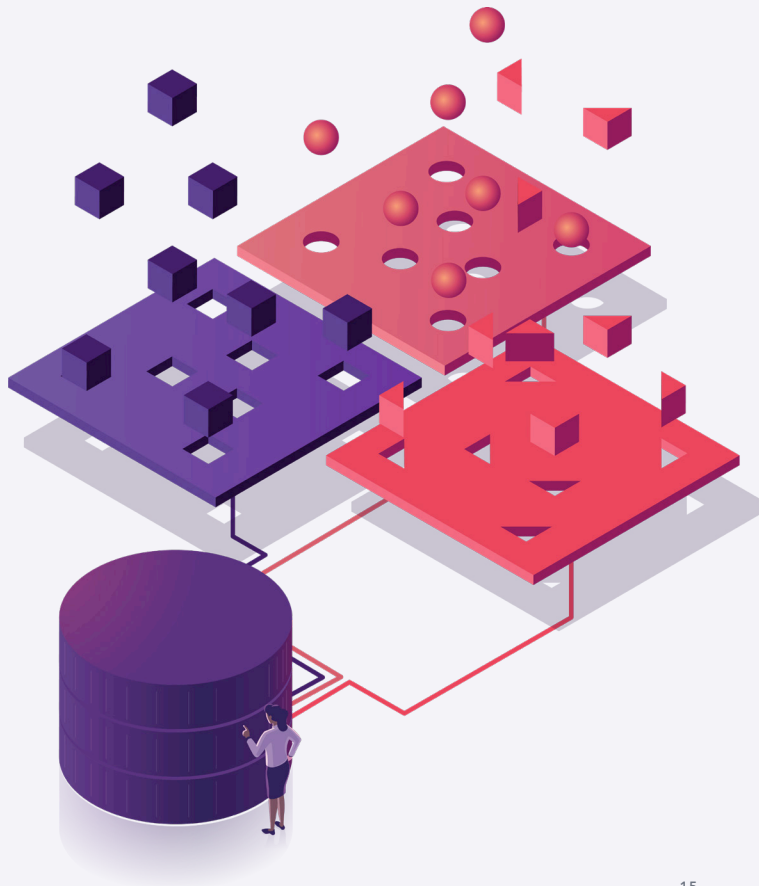- Organizations that need to run many concurrent queries.

Other Tools and Takeaways

As we're focused on cloud-native technology stacks, we limited our notable tools to options in that space. Beyond the options covered here, there is an entire category of on-premises data warehouse providers that we didn't touch on in this ebook, as well as unstructured data lakes with a querying interface built on top of them. Generally, cloud data warehouses will make the most sense for the majority of organizations outside of large enterprises (and even then), but sometimes the particulars of certain unique situations or industries merit the need for an on-prem solution or a data lake-like approach to storing and querying your data.

TL;DR Data warehouses are the hub of your customer data technology stack. They act as the single structured storage place for all your data and provide the foundation on which to model and query your data. Modeled data becomes the single source of truth for data in your organization, and querying data gives you the power to make data-driven decisions for your business. When building a customer data tech stack, you'll want to ensure that you can feed all of your important data into your warehouse, and run your reporting and visualizations on top of it.

References and Resources

- https://www.flydata.com/microsoft-azure-data-warehouse-vs-amazon-redshift/

- https://blog.panoply.io/a-full-comparison-of-redshift-and-bigquery

- https://www.periscopedata.com/blog/interactive-analytics-redshift-bigquery-snowflake

- https://news.ycombinator.com/item?id=13916030

- https://thomaslarock.com/2018/03/azure-sql-data-warehouse-costs-vs-aws-redshift/

- https://www.zdnet.com/article/azure-sql-data-warehouse-gen-2-microsofts-shot-across-amazons-bow/

- https://statsbot.co/blog/modern-data-warehouse/

- https://www.blue-granite.com/blog/is-azure-sql-data-warehouse-a-good-fit

- https://www.stephenlevin.co/data-modeling-layer-startup-analytics-dbt-vs-matillion-vs-lookml/

- https://fivetran.com/blog/warehouse-benchmark

Data Pipelines to the Warehouse: ETL/ELT Tools

In your tech stack, you're likely to have lots of sources of customer data and tools that are relevant to creating a complete 360 degree view of your customers. There are many sources you'll want to connect, as well as different formats and structures of data. This is where data pipelines come in. Depending on which tools you want to connect and what destinations you want to connect them to, there are a few different types of data pipelines you may need.

A good place to start when thinking about data pipelines is the pipelines between your tools and your data warehouse. Generally, bringing your data into your data warehouse will happen through an Extract, Transform, Load (ETL) process or tool. This is particularly true for table-based data that you send in batches, and for connecting tools that don't have an ability to send data to your warehouses themselves.

It may be helpful to sketch out the boundaries of what ETL tools are useful for and do well. The difference between the tools in this chapter, and the tools in chapter 5, for example, is that ETL tools

only replicate your data to data warehouses. They are not designed to connect and orchestrate your 3rd party tools to other tools. ETL tools are also designed to handle batch format pipeline jobs. If you need to send data in real-time, ETL tools are not your best bet.

So what kind of data would you want to use an ETL tool to send to your warehouse? There are many types and sources of data, but generally, it will include backend data like server transactional data, internal application data, shipping info, customer info, loan info, and/or direct mail info, depending on the particulars of your business. It will also include data from 3rd party sources like the tools discussed in chapter 7.

There are two common means of satisfying the requirements of an ETL process once you determine that your business needs it. One is a homegrown solution. It's not uncommon for companies to handle this layer with a customized script that they build themselves.

This has the obvious benefit of being free in terms of actual monetary costs, and gives you the flexibility to do exactly what you want to accomplish for the purpose of moving your batch data into your warehouse. However, the downside of a homegrown solution is that it requires dedicating time and engineering resources

towards building and maintaining the script. The maintenance time in particular can be substantial, since this functionality is often central to your business. If it breaks down, your business comes to a halt until you get it fixed again. This means you'll need to give it ongoing care and attention.

If you have multiple data sources that you want to send to your data warehouse in batch format, we recommend using an ETL or, as the latest thinking goes, an ELT (Extract, Load, Transform) tool that is dedicated to providing and maintaining this pipeline for you. For a cloud-native data tech stack, we recommend starting your exploration with cloud-centric tools that can operate in an ELT manner. These players are best for table-based batch data because they are schemaless and cloud-based.

For a cloud-native data tech stack, we recommend starting your exploration with cloud-centric tools that can operate in an ELT manner.

ETL vs ELT, what's the difference?

The acronyms look similar, but there are important distinctions between the ways the tools work. ETL tools are designed to take your data from point A, transform it into the format you require en route, and deposit it at point B. ETL is the way this functionality has been handled for years.

ELT tools take a more modern approach. They separate the transform part out of the equation, and focus solely on moving the data from point A to point B. The purpose is to separate out the analyst's job (modeling data) from the engineer's job (moving data). Since data storage is cheaper than engineering time these days, this way of doing things makes a lot more sense and has seen a rise in popularity.

Important Considerations for this Layer

As you get to the stage of evaluating tools in this layer, there are a few comparison factors we recommend keeping in mind:

Think about your data sources.

- If you only want to replicate one or two sources of data in your warehouse, then it may be worth looking at tools in other chapters that can capture and ETL that data for you at once.

- If you have multiple data sources that are well suited for sending in batches, then you'll want to consider one of these tools or look into building your own solution.

- If you have many data sources, then we recommend going with a tool from this chapter as opposed to building your own, unless you have a lot of technical resources to dedicate to it.

What sources and destinations does the tool support?

If the tool you're looking at doesn't support a source that's important to you, what's their timeline for building it? Fivetran and Stitch, for example, are both excellent at responding to new connection requests and prioritizing getting them live.

Security matters for this layer.

You're actually touching the data with these tools, since you're moving it around, so ensuring that your tool is secure is important. In order to be in compliance with security and privacy measures like SOC2, GDPR, and HIPAA, you'll have to ensure security for your customer data while it's in transit.

Pricing Matters.

There are some differences in pricing between the tools, so pricing is something to compare. If you're considering a homegrown solution here, consider the cost of maintenance for your own homegrown solution vs the cost for an outside tool. ETL/ELT is an ongoing process that will require ongoing maintenance and engineering focus, as opposed to a one-off project, so building it yourself is not always more cost-effective than going with an outside tool.

Support is important here.

ETL/ELT is crucial to the business. Making sure you have a tool with good support in moments of trouble is a major benefit.

Notable Tools in this Layer

As per the rest of this ebook, we'll be focusing our coverage of these tools towards ones that are cloud-native or at least well-suited to fitting into a cloud-native technology stack, with an eye towards the needs of small to mid-size businesses. These tools are priced on a $-$$$$ scale based on available information.

Alooma

Founded in 2013, Alooma aims to be a modern ETL and data pipeline. They differentiate by offering the ability to handle event streaming data as well as batch data, and by the ability to optionally do transform work on the data en route like a traditional ETL tool.

⊕ **Pros**

- Capable of handling event stream data as well as batch data.

- Offers the ability to transform the data en route in their code engine via Python. This means that Alooma can function more like a traditional ETL tool if required for certain transformations, although it's not a full-fledged traditional ETL tool as far as additional functionality and features go.

- Offers lots of customization ability and lots of features beyond data pipeline services.

⊖ **Cons**

- More expensive than other tools in this area.

- The transform functionality works in their Python-lite code engine, so users of this functionality must know Python.

💲 **Price**

- $$$

- Alooma is priced based upon rows loaded into your data warehouse (so a per-event basis).

- Alooma is priced more granularly (per million events) than other tools on this list.

- They offer a free trial, but no ongoing free version is available.

Best for

- Larger organizations.

- Organizations that need a lot of customization for their ETL/ELT functionality.

- Organizations that want to move both batch and event streaming data to their warehouse with one tool.

- Organization that can get value out of an extensive range of features at this layer.

Fivetran

Started in 2012, [Fivetran](#) positions themselves as a tool for syncing and replicating your business data into your data warehouses quickly and painlessly. They operate largely as an ELT tool, focused on moving data from place to place instead of on the transform step. They do offer some lite transform capabilities in the form of data cleaning through normalizing your schemas.

⊕ Pros

- Robust set of integrations and connections. Also happy to build what's missing, so you can copy or import to lots of warehouses.

- They have good functionality with webhooks for event collection, so you can build your own connections with tools like Zapier if needed.

- Great support. Fivetran has a top notch support team.

- Reliability: You know your data will end up in the warehouse. It will often need a bit of transformation work, but it's there and ready for you to transform or use directly.

⊖ Cons

- Can take up extra storage space beyond what you may initially expect. The reason for this is that Fivetran doesn't actually delete data in the warehouse when you tell it to delete data. Instead, it archives it. This means that it only appends, never modifies. Some people like this fact, since it lets you look at state change over time, but it means you have to customize your queries to ignore rows that you've deleted.

- Doesn't have much transform functionality. This enables it to move data easily and effectively, but you'll have to work in a separate tool to transform the data to your required schema.

- Can't really control what's synced from your data sources. It's more all or nothing than other tools on this list. You can choose what sets of data are synced, but you can't select only certain columns within those sets.

$ Price

- $$-$$$

- Fivetran prices based on the data sources and connections you need.

- They offer a free trial, but no real ongoing free version.

Best for

- Users or organizations who want more of an on/off switch for their data pipeline. It's great for marketers, analysts, or operations users.

- They target their product more for the line-of-business side, not the developer side.

- Also better for more mature organizations, given the more premium pricing range.

Stitch

Founded in 2016, Stitch positions themselves as a developer-focused ETL and data pipeline tool. More generally, they operate as an ELT tool, focused on moving data from place to place, and separating out the transform steps, although they do offer integrations with transformation tools to do the transform step as well if that's a preferable method for you.

⊕ Pros

- Stitch is one of the most open source-focused tools in this space. They developed Singer, an open source ETL tool, and enable the open source community to develop new integrations.

- Stitch exposes log files to you so you can get better visibility into your ETL process.

- They offer more granular control for ingesting data than most other options. You can choose which columns you want to import into your warehouse rather than have it be all or nothing, but this also comes with more configuration. This

is better if you have a smaller data warehouse and want specific data to be synced only, or you're being very careful with your storage usage.

- Reliability: You know your data will end up in the warehouse, ready for you to transform or use directly.

- Stitch has a great support team.

⊖ **Cons**

- Fewer commercially supported sources than some of the other tools in this space.

- More granular control means that it's easier to accidentally create discrepancies with your data or tables.

- Doesn't have any built-in transform functionality. There are good reasons for this (it enables Stitch to move data easily and effectively), but you'll have to work in a separate tool to transform the data to your required schema.

$ **Price**

- $-$$

- Stitch tends to come out the cheapest of tools on this list (but not always). They price on the volume of rows you need to replicate.

- They have a free version that can satisfy small teams or users (up to 5 million rows per month).

Best for

- Developers. Stitch's target market is people who want to dive more deeply into their ETL/ELT tool. This makes it well suited for data engineers and developers who are comfortable with working in more technical and backend-oriented tools.

- Often slightly cheaper than other options, so better for smaller companies or those who are more price-sensitive.

- Those who want more granular control and configuration over their ETL/ELT tool.

Other Tools and Takeaways

In addition to the tools we've covered above, other tools that may be worth considering for you in this layer are Blendo and Flydata. Singer is a strong open source option for this layer, and Segment and Heap can handle this functionality for certain sources and data types as well.

TL;DR A common data format that you'll often want to replicate in your data warehouse is table-based data. Generally, this is data that you'll want to send in batches over a fixed interval, as opposed to data that you'll want to stream continuously in real-time. ETL and ELT tools are designed to handle this data and bring it easily and quickly into your warehouse. ETL tools transform the data into your desired schema en route, while ELT tools let you do the transform step after the data is in your warehouse.

References and Resources

- https://www.stephenlevin.co/segment-vs-fivetran-vs-stitch-which-data-ingest-should-you-use/

- https://fivetran.com/docs

- https://www.alooma.com/blog/etl-tools-modern-list

- https://blog.modeanalytics.com/etl-tools-analytics-stack/

- https://statsbot.co/blog/etl-vs-elt/

Capturing Customer Interactions: Customer Behavioral Data Tools

The second major type of data that you'll want to replicate in your data warehouse is event or event-like data. This is often client-side, 1st party data on how your customers actually interact with your site, product, or application. The tools in this layer are focused on capturing that data to different degrees, and some are also good at sending it to your warehouse. The data that's generated and captured in this layer can give you insights like "how are your customers interacting with the shopping cart?" and "are customer starting to fill out your signup forms, but giving up part way through?"

There are several different perfectly valid ways to group the tools in this chapter and in the next chapter. There are a few tools in this chapter that can be used to do the functionality of the tools in chapter 5, and there are some features of a few of the tools in chapter 5 that are relevant for chapter 4. In essence, the separation between these two layers is fuzzy. We've decided to separate them based on whether or not they offer a visualization and analysis layer.

The tools in this chapter have well-fleshed out analysis layers, while those in the next chapter focus more on taking in data and pushing it to other systems and tools. Between the tools in these two chapters, you should walk away with the ability to a) capture customer behavior data on your website, product, and mobile app, b) connect that data to your warehouse and enrich it with other data sources (such as business systems and marketing channels), and c) distribute that data to other relevant tools in your stack in order to trigger actions (such as sending out an email after someone makes a purchase).

Unlike some of the other layers, the tools here have similar goals, but very different approaches and ways of accomplishing those goals. One type of tool on the following list is the aggregate data tracker. These are the oldest tools, your Google and Adobe Analytics, having been created in the 1990's. They're focused on capturing aggregate metrics like pageviews and bounce rates. The second type of tool on this list are event-based and user-centric. These are your Mixpanels and Amplitudes. They were developed in the late 2000's and early 2010's. These tools saw the change in the web from static pages to dynamic applications and built themselves around being event-based. This opened up the

possibility of capturing individual user behavior through tracking predefined actions that users did on your website. The other tools on this list, like Heap, have built upon the idea of user-centric data and analytics to address some of the downsides of the original event-based approach. These tools focus on improving the ease of setup, making data capture automatic, adding virtualization, and/or improving the connecting capabilities of the late 2000's tools.

"Aggregate metrics like bounce rates are too low value for today's business needs. When thinking about your stack, start with the business questions you need to answer, and then tie those to the tools that can answer those questions. Some of those tools will provide behavioral data like Heap. Some will be data warehouses and business intelligence tools. Some will be multi-touch attribution or lead scoring tools. Regardless, it all starts with the business questions you want to answer. By focusing on the business questions you need to answer, you won't get bogged down in the metrics or visualizations provided by the tools vendors early in the process."

RYAN KOONCE,
CEO, MAMMOTH GROWTH

Important Considerations for this Layer

When evaluating the options in this layer, there are some important considerations and factors that merit your attention. To help determine which customer behavioral data tool is right for you, you'll want to take a moment to clarify what it is you want to accomplish with your analytics and customer behavioral data. The goal for most businesses should be a clear 360 degree view of their customers, which requires strong and complete first party data. Beyond that, the type of data you need and the actions you want to take with that data are influenced by the particulars of your industry and business.

Price is a factor here.

There are lots of differences between the tools in terms of total expense and the way they price. It's important to compare price on a process and overhead basis vs the straight price tag of a tool, given these differences.

The technical sophistication of your company and the amount of data and data engineering resources you have matters.

Some tools can be best utilized by more technically sophisticated users and organizations, while other tools are well suited for less technical or more line-of-business users.

The kind of business that you are makes a difference as far as what grouping of tool would work best for you.

Different businesses have different needs in terms of what customer behavior is relevant to them and the kind of questions they most need answered.

- For example, companies with low Lifetime Value (LTV) customers or content-focused sites may be better served with aggregate tools, while companies with high LTV customers, SaaS products, or ones that can see tangible benefit from increases in conversion rates may be better served with more modern, user-centric tools.

- Consider the industry you're in, how much and how fast your business changes and grows, and the sorts of information you may want to know about your customers.

What does the reporting functionality look like?

How much do super advanced features matter to you as opposed to focusing on having trustworthy, clean, and complete datasets?

Is mobile a primary part of your business?

Certain tools on the following list are better at mobile than others.

What kind of questions do you want to ask?

If your questions are clearly defined and will be the same for the foreseeable future, then the downsides of some tools on this list won't be a problem for you. If your questions change and evolve with your business, then instrumentation becomes a much bigger factor for you.

Notable Tools in this Layer

As per the rest of this ebook, we'll be focusing our coverage of these tools towards ones that are cloud-native or at least well-suited to fitting into a cloud-native technology stack, with an eye towards the needs of small to mid-size businesses. These tools are priced on a $-$$$$ scale based on available information.

Adobe Analytics

Adobe Analytics, formerly Omniture, was founded in the late 1990's in the same generation as Google Analytics. Acquired by Adobe in 2009, Omniture has the largest wallet share of the analytics market, and is especially prevalent in large enterprises and international companies.

⊕ **Pros**

- Great partner and channel presence. There are lots of Adobe certified partners and consultancies all over the world.

- Endless features. If you can instrument it, you can do nearly anything in Adobe Analytics.

- Good at dashboarding and reporting.

⊖ **Cons**

- Adobe Analytics is the hardest tool in this group to properly instrument, by a large margin. Adobe implementations essentially require a large, full-time team solely dedicated to instrumentation and maintenance in order for it to work at a usable level.

- The core technology is around aggregate pageviews and traffic metrics, so today's user-centric needs are not a core element of Adobe Analytics. They've been stapled onto the product recently as user behavior analytics rose to prominence, but the functionality is not strong.

- Weak ecosystem. Adobe doesn't integrate well with many tools or datasets (even other Adobe tools).

- Mobile support and functionality isn't as good as other tools on this list.

Price

- $$$$

- Adobe Analytics is essentially exclusively enterprise-focused, and it's priced to match.

Best for

- Older, well-established companies that don't have internal technical expertise and have a lot of budget to bring in and maintain one of the many available Adobe partners.

- Large companies based primarily outside of the US that don't have an in-house analytics team and can take advantage of Adobe's international strength.

Amplitude

Founded in 2012, [Amplitude](#) originally aimed to build upon the path Mixpanel paved and deliver a Mixpanel-like product with a more modern architecture and a cheaper price point. Today, Amplitude focuses on delivering product analytics for product managers.

⊕ Pros

- Strong mobile analytics history that translates into good cross-platform capabilities.

- Amplitude has a rich UI with a wide range of reporting features and advanced analytics functionality (such as Compass, their predictive retention tool), particularly for product managers.

⊖ Cons

- Instrumentation is difficult to set up and extremely challenging to maintain over time. Amplitude is a [manual tagging](#) tool.

- No retroactivity. If something is not tagged or the tracking code breaks, then your data is lost.

- Integrations in Amplitude are limited by platform or pricing tier. They only have a direct integration with Snowflake on the data warehouse side.

💲 Price

- $$

- Amplitude prices based on the number of events you track.

- They have a free tier that supports up to 10 million actions per month.

Best for

- Product managers who have a good grasp on what they want to track about users in their product over the next few years.

- Those who need advanced reporting functionality as a core business driver.

- Mobile app companies.

Google Analytics

Formerly known as Urchin until Google acquired it, Google Analytics (GA) was one of the first analytics tools on the market, coming on the scene in the late 1990's. As a first generation tool, they were built in the aggregate analytics era. They are well designed for aggregate metrics, but less well suited for event and user metrics. Google Analytics is essentially two separate tools: GA Free and GA360. GA Free is their free version, which provides analytics value for a limited set of use cases, while GA360 is their premium version, which acts as a full-featured tool more akin to Adobe Analytics.

⊕ Pros

- Great at measuring "publisher metrics", such as pageviews and bounce rates.

- Best-in-class integration with Google Ads, good integrations with other Google tools like DoubleClick, and a few proprietary integrations (such as with Google Search Console).

- Strong capabilities for measuring your SEO and ad performance, thanks to the above integrations.

⊖ Cons

- Not well suited for tracking user-centric activities. Over the years, GA has added functionality like event tracking and attribution, but they are stapled on features that don't mesh well with the core architecture, so the capabilities of these features is weaker than what other tools on this list can offer.

- Sampled data. GA Free samples your data for analytics, presenting conclusions often drawn from less than 0.1% of available data. Depending on what you're analyzing, this can give mostly accurate results, but it's very dependent on what you want to find out. GA360's sampling threshold is much higher, and in essence doesn't appear as sampled for most use cases and customers.

- Inflexible data model. GA Free can only support a max of 20 properties and GA360 only 200, both of which becomes strictly limiting very quickly for user-centric businesses.

- Weak mobile support and functionality.

- Proprietary practices make it hard to integrate GA with other tools. For example, GCLID automatic tagging is available within GA to ease the pain of manually tagging

ad campaigns, but it's proprietary, so no other tool will understand the GCLID tags.

- Data ownership: with GA Free, you don't own your data. It belongs to Google. This means you can't integrate GA Free with BigQuery, for example. This is not an issue for GA360.

Price

- Free or $$$$

- GA has two different versions: Free or GA360. GA360 is a flat annual $150k fee per year, and brings additional functionality and reporting capabilities beyond GA Free. For example, GA360 greatly raises the sampling threshold, as opposed to the heavy sampling that GA Free does, and gives you back ownership over your data.

- GA Free is the best free tool in the space for handling your ads and SEO performance, given the integrations it has with other Google tools.

Best for

- Tracking your SEO, blog, and advertising performance. GA has tight integrations with the Google suite of products,

including exclusive integrations with Google Search Console. GA Free is something that many businesses will find to be useful. Since it's free and can provide this functionality, there are good reasons to have it, despite the fact that you'll need a separate tool in this layer to bring in the user-centric data.

- Startups with no money. GA Free will at least give you some basic functionality until you reach enough maturity to make analytics and 1st party data a priority.

- Media and blog sites. If you're exclusively content-driven and care about traffic instead of user behavior, then what GA is able to track will be well suited for your needs.

Heap

Founded in 2013, Heap is infrastructure for customer data that automates away the pain of data. The Heap platform automatically captures every interaction from web, mobile, and other customer touchpoints and makes that data available retroactively. Heap's data virtualization technology enables your team to spend their time discovering new insights and taking action–not tagging, building pipelines, or wrangling data.

⊕ Pros

- Autocapture. Heap automatically captures all of your customer behavior data on your website, product, and/or mobile app so you have a complete dataset to work with.

- Retroactivity: Heap is retroactive, so when you decide to define a new event, the data for that event is present from the time you installed Heap up to real-time.

- Ease of set up and instrumentation. Setting up Heap is a 10 minute process of adding a Javascript snippet to your website or iOS/Android SDK to your app. Heap starts automatically capturing data right from there.

- Virtual Events. Heap delivers a virtualization layer on top of the raw data it captures, so you can define, change, and update events without touching your code.

- Sources: Heap acts as your customer data infrastructure, allowing you to connect your various 3rd party tools or databases with Heap to enrich your user-level properties and event data.

- Data out: Heap supports direct integrations with Redshift, BigQuery, and S3, as well as other warehouses through an ETL tool.

⊖ Cons

- Fewer advanced reporting features than some other tools on this list.

- Autocapture is more than you need if all you want to capture are traffic metrics.

ⓢ Price

- $$

- Heap tiers are split based on the level of support you need, as well as around a few advanced security features.

- Heap prices based on the number of user sessions, as opposed to the number of events.

- Heap offers a free version for use with analyzing up to 5,000 user sessions per month.

Best for

- Organizations that want a complete picture of user identity and a resilient data foundation.

- Those who want to ask analytics questions and iterate quickly, where speed of getting answers has a meaningful impact.

- Organizations that grow and change, and need agility in their web and mobile analytics tool to let them redefine events or ask new questions without waiting weeks for new data.

- Organizations that value deep integrations for their behavioral analytics tool with the rest of their data stack, and value the ability for their tool in this layer to send data to their warehouse.

Mixpanel

Founded in 2009, Mixpanel was created in response to the changing nature of the web at the time. Mobile had begun to arise as a legitimate presence, and the web had begun to shift from static pages to applications and dynamic user interactions.

Mixpanel helped usher in the idea of user-centric analytics, as opposed to traffic-based analytics that had existed previously. This came along with the introduction of events and the first forays into tracking those.

Pros

- Mixpanel delivers a wider range of functionality than many tools on this list. They've pursued a more all-in-one strategy, bringing in things like A/B testing and notifications into their product for mobile apps.

- Solid mobile capabilities and history.

- Mixpanel is a simpler tool to use than some others on this list.

⊖ **Cons**

- Instrumentation. Mixpanel is a manual tagging tool, which comes with loads of maintenance, hardcoding, and scope challenges.

- Reporting. For a manual tagging tool, they have fewer reporting features than other options on this list.

- Getting data out of Mixpanel is usually a challenge. They only offer an API for this purpose, instead of any native integrations or managed ETLs.

⑤ **Price**

- $$$

- Mixpanel offers a free version for up to 5 million event instances per month.

- They offer two paid tiers: one is a cheap, limited startup version, and the other is the full enterprise version.

Best for

- Those who want a jack-of-all-trades tool for mobile that can do their A/B testing, their notifications, etc. along with user analytics.

- Those who value an easier learning curve, once they get past the implementation challenges.

Other Tools and Takeaways

The tools covered above are some of the most notable tools in the space. They provide aggregate views or 1st party, frontend data on what your users are doing on your website or app that you'll feed into your data warehouse. However, there is a whole spectrum of tools in this space. Beyond the tools listed here, other options include Kissmetrics, Coremetrics, Pendo, and many others that look to capture some degree of customer behavior, aggregated web data, or specific analytics data. A category we don't cover here for example is "experience analytics" tools that capture session recordings of what your customers do on your site, such as their mouse movements and typing.

TL;DR An important type of data that you'll want to capture is 1st party "event-like" data about your customer behavior. This data is central to understanding individual users and cohorts, and their interactions with your site or app. The tools in this layer are focused, to different degrees, on gathering that data, organizing it, reporting on it, and sending it to your warehouse or other locations.

References and Resources

- https://heapanalytics.com/blog/analysis/manual-tagging

- http://blog.oribi.io/mixpanel-vs-heap-analytics/

- https://practicoanalytics.com/amplitude-vs-mixpanel-pros-and-cons/

- https://medium.com/@jasons_eid/reviewing-the-pros-and-cons-to-google-analytics-556edff7dfa8

- https://chrislema.com/why-i-love-heap-analytics/
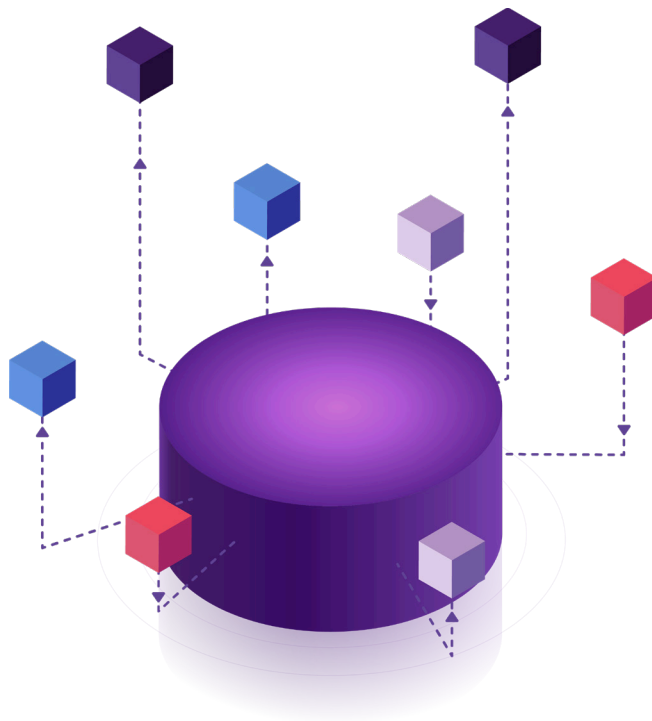
Connecting and Distributing Data: Customer Data Platforms

Your relevant customer data lives across many tools, and the actions that you want to take in response to new customer behavior will happen across many tools. Likewise, when you get relevant new data about a customer cohort from your customer behavior tools, you'll likely want to adjust the knobs and levers you have in your company for interacting with that customer cohort. To move all of this data around, format it, and make it accessible to all these different tools is where a Customer Data Platform (CDP) comes in.

This layer is new to the market, and product-market fit is still being shaken out. The exact definition of a CDP is far from settled, and the requirements for what your CDP must do are not fully conceptualized either. Many tools on this list evolved from different spaces and use cases to make a claim on the CDP space. Given how diverse and unsettled this space is today, it's worth evaluating the tools that we cover below on their own merits as opposed to head-to-head. The breadth of what can be considered a CDP is important to keep in mind,

but it doesn't mean that you should ignore this layer. Some of the tools in other layers can handle some of the functionality that CDPs can, particularly around connecting your tools to your warehouses, but in general there are a few factors you'll want to look for in your CDP. You'll want to make sure that they can connect your different tools together, and that they can route and distribute data around your technology stack appropriately.

The overall value for a CDP is around omnichannel orchestration.

Generally speaking, a CDP should be able to collect data across your important sources of customer data, resolve customer identity from these different sources, enable you to segment customers any way you want, and send those segments to destinations that let you take action (such as send emails, show targeted ads, etc.).

Given how unsettled this space is today, many of tools that call themselves CDPs are really only focused on part of the above set of functionalities. It's not uncommon for organizations to use multiple tools from the CDP space because of the widely different strengths and narrow set of capabilities that they have, or to not use a tool from this layer altogether. Your particular business needs will determine whether you need multiple tools or none at all from this layer.

Important Considerations for this Layer

When looking to implement a CDP, there are a few factors worth exploring that you should keep in mind.

Sources and destinations.

Your CDP must be able to pull data from the sources most relevant to you and push it into the destinations that are most relevant to you.

Price.

These are not cheap tools, so the costs involved for this layer must be a consideration for you and the size of your business.

What's the primary source of your data?

For mobile-first businesses, certain tools are stronger, while others have more of a history with web-based organizations. Since the tools in this layer came to be a CDP in different ways, they have

very different core strengths. Companies with small data files, that are web-first, or that are quite technical all point towards different tools.

Vendor Lock-in.

If you implement one of these tools, how easy is it for you to change tools or move in a different direction?

Ease of implementation and ongoing maintenance.

Many of these tools will have a long initial implementation period, but some of them are easier to maintain as your data stack changes via one-click integrations with popular tools, while others require lots of dev work to set up new integrations.

"All decisions should be based off of customer data. We're moving from a product-based world, where we analyze product performance, to a customer-based world, where we analyze customer experience and relevance. The only way to understand that and to operate against that is to have a platform that collects the data about customer events and interactions, unifies it, and makes that data accessible to everyone in the organization."

JONATHAN MENDEZ
PARTNER, ARKLE ADVISORS

Notable Tools in this Layer

As per the rest of this ebook, we'll be focusing our coverage of these tools towards ones that are cloud-native or at least well-suited to fitting into a cloud-native technology stack, with an eye towards the needs of small to mid-size businesses. These tools are priced on a $-$$$$ scale based on available information.

AIQ ActionIQ

Founded in 2014, ActionIQ is focused on building database technology. They aim to tackle the infrastructure for marketing activation for larger enterprise-level companies. Their target audience is enterprise marketers who are looking to better implement their campaigns.

⊕ **Pros**

- Their biggest strength is their database. This is important for enterprise customers with large files of customers with lots of attributes and dimensions to handle.

- They claim to operate on the original structure of the data emerging from any source rather than transforming it into a common structure.

- Offer some segmentation, measurement, and testing capabilities for campaign orchestration.

⊖ **Cons**

- ActionIQ is designed specifically for large enterprises, so much of their functionality and benefits are not needed or not well suited for small-to-midsize businesses.

- They provide no information about sources or destinations, so make sure to verify that they support your particular needs

- ActionIQ doesn't focus on data collection or identity resolution, but more on the segmentation and campaign orchestration side.

Price

- N/A

- ActionIQ keeps their pricing structure and info close to the chest. You'll have to contact them to learn more.

Best for

- Large organizations that have huge customer data files and large databases that they want to activate their campaigns against.

Lytics

Founded in 2012, Lytics has made their mark by focusing on orchestrating customer experiences and predictive audiences. They position themselves for marketers and their main way of achieving connection and distribution is through sending audience segments between tools.

Pros

- They offer machine-learning driven orchestration to try and automate campaign orchestration for marketing teams.

- Have built many of their core functionalities atop open source projects.

- Lytics' UI is one of the best in its space.

Cons

- Lytics is targeted only for marketing use cases. It's not a particularly technical tool, so it's not well suited for broader uses and teams.

- Less focused on collecting data, and more on segmenting and campaign orchestration.

Price

- N/A

- Lytics keeps their pricing structure and info close to the chest. You'll have to contact them to learn more.

Best for

- Marketers. Lytics has filled the space left by mParticle and Tealium as they've started to move more towards developer audiences, and oriented their product towards marketing users and use cases.

- Larger organizations and teams that are focused on understanding their customers' experience and on their marketing campaigns.

mParticle

Founded in 2013, mParticle was born from the mobile SDK world. They began as a data integration tool for mobile app developers, and have expanded to work across platforms as a CDP.

Pros

- Their API connection layer (inputs and outputs) is strong, especially for mobile-app based businesses.

- Good at understanding intricacies of app-based business and the data connections that are specific to that type of technology, and customer facing interfaces.

Cons

- Their path and core strength comes from the mobile world, so the ability to handle more web-based data is newer and less strong than some web-native tools.

- No retroactivity. Once you decide to start syncing data with mParticle, you don't get historical data along with it.

- Brittle data layer. The data layer is not dynamic, so if what you want to collect changes, you'll have to manually update the tool. This is fine if the data you want to collect from your tools is unchanging, but runs into issues with data around customer behavior.

💲 Price

- N/A

- mParticle keeps their pricing structure and info close to their chest. You'll have to contact them to learn more.

🤝 Best for

- Mobile-first organizations and companies where mobile data is critical to the business.

Segment

Founded in 2011, Segment was created to simplify the process of integrations and connecting the APIs from different tools. They work cross-platform today to pull and push data from tool to tool. Compared to other tools in this layer, they are most focused on infrastructure, achieving their connection and distribution goals by acting essentially as an ELT tool for streaming data. There are one of the few tools in this space that can operate as an ELT tool and push data into your warehouse as well.

⊕ Pros

- Segment offers one-click push integrations for some popular tools, as well as the ability to push data into popular data warehouses.

- Segment offers a limited form of replay with their historical data replay feature.

- They can capture some event interactions directly, unlike many tools in this space. This makes it possible to use them for collecting some customer behavior data as well.

- Segment offers some intelligence on top of being a data pipeline. They offer Personas, for example, which can give you some ability to orchestrate actions for segments of users in your tools.

⊖ **Cons**

- Segment doesn't enable different tools to actually talk to each other. It enables them all to talk to Segment, so it operates a middle man. Outside of data collected through Segment's APIs, many of their 3rd party sources don't integrate with each other and are not available in Segment's streaming destinations.

- Less intuitive UI than other tools.

- Brittle data layer. The data layer is not dynamic, so if what you want to collect changes, you'll have to manually update the tool.

- Event capture happens via manual tagging and instrumentation.

- Their ability to push customer segments to destinations for taking action upon is less robust than some other tools in this space.

$ **Price**

- $$

- They offer a small free tier, then two paid tiers that differ based on the number of users you want to track per month, as well as their advanced set of features.

Best for

- Segment is more developer-focused than other similar offerings, so it's great for that user set.

- Those who want more of a full stack tool that can connect tools together for taking actions as well as connect them to your data warehouse for storing data, and that can handle many different types of data.

👓 Simon Data

Founded in 2014, Simon Data is targeted at marketers and non-technical users. Their focus is around letting their users build audience segments for use in their campaign tools and pushing out those segments to tools that can act on them.

⊕ Pros

- Complex campaign orchestration is their focused value-add over some other options available. They're excellent at sending your defined segments to tools that take actions based on those.

- Simon Data brings testing and reporting capabilities to help marketers be data-driven. This includes dashboarding and A/B testing capabilities.

⊖ Cons

- Their focus is less on the data infrastructure and more on the campaign orchestration component. This means their scope is around the martech stack and marketing use cases.

- They don't address data collection needs or customer identity resolution, like some tools in this space do.

💲 Pricing

- N/A

- Simon Data keeps their pricing structure and info close to their chest. You'll have to contact them to learn more.

👥 Best for

- Marketing teams who want to orchestrate complex campaigns and value the activation capabilities that a CDP can do.

- Those that need their testing and reporting capabilities to be unified with their campaigns.

Snowplow

Founded in 2011, Snowplow is focused on enabling your most technical users to set up the exact data collection implementation that best matches the needs of your particular business. Snowplow is a bit of a hybrid tool. They are focused on event data collection, pushing that data into the warehouse, but they've also released (in beta) the ability to push that data directly to other tools for the purpose of taking actions.

⊕ Pros

- If properly set up, the final dataset out of Snowplow is very well structured for use beyond reporting use cases. They offer a lot of flexibility for structuring your data the way you want, instead of pushing you towards any particular format or organization.

- Excellent for collecting data with varied event collection support. Snowplow can collect events from most any type of platform.

- Snowplow offers enrichments for your data from 3rd party sources, such as Marketo.

- Snowplow brings a fairly unique functionality to the market, with the ability to act as or support CDPs, analytics tools, and ETL tools.

⊖ Cons

- Very complicated set-up. Perhaps the hardest tool in this space to set up and get working in the first place.

- Primarily works only within the Amazon ecosystem (e.g. Redshift, Kinesis, and AWS). Although, it's important to note that they recently added support for Snowflake as well.

- Snowplow is largely a manual tagging tool on the data collection front. You have to decide what you want to track and tag your specific events before you see any data.

- As a hybrid tool, they are not as strong in each individual focus area as targeted tools. For example, they have fewer sources and destinations than most tools in this chapter, no reporting functionality when compared to customer behavior tools in chapter 4, and are only focused on event data when compared to ETL tools in chapter 2.

$ Pricing

- $$

- Snowplow has three tiers: a basic tier, a premium tier, and an enterprise tier. They split based on support levels, the number of events you want to track, and some premium features.

- Snowplow also offers an open source version of their product and a few other open source functionalities.

Best for

- Engineering teams. Snowplow targets their product and messaging towards the most technical users in an organization.

- Highly technically sophisticated organizations who find the high set-up and instrumentation pains worth it to get actionable data for uses beyond reporting.

- Organizations who want more of a full-stack solution that can handle data collection, event data ETL, and a bit of basic orchestration in one tool.

Tealium

Founded in 2008, Tealium built up their technology as a web-based tag management solution. From there, they've expanded to handle mobile, client-side, and server-side connections, and moved into the CDP space with one of their offerings, AudienceStreams. The way they address the connection and distribution of use cases is through enriching cookies and creating a Unified Data Object that speaks to all the tools.

⊕ Pros

- Strong at handling web-based data.

- Tealium is great for client-side interactions and tag management.

- They have the largest library of integrations and an easier to use UI than some other tools in this space.

⊖ **Cons**

- No retroactivity. Once you decide to start syncing data with Tealium, you don't get historical data along with it.

- Brittle data layer. The data layer is not dynamic, so if what you want to collect changes, you'll have to manually update the tool. This is fine if the data you want to collect from your tools is unchanging, but runs into issues with data like customer behavior.

⑤ **Price**

- $$$

- Tealium prices based on visits per month, with four tiers. They don't offer a free tier.

▦ **Best for**

- Better for companies whose main touchpoints are on the web (from a technical perspective) because of the way Tealium has built their technology.

- If you're looking for a uniform data layer for your technologies, Tealium is well suited to create and deliver that.

- Larger organizations and enterprises.

- Those who need tag management capabilities. Tealium is a tag management provider first and foremost, so if you use a manual tagging tool from the customer behavior chapter, Tealium could be useful to help you with the instrumentation challenges present in those tools.

Arm Treasure Data

Founded in 2011, Treasure Data aims to be a full stack solution for martech stacks. They are targeted towards marketers and position themselves as a CDP that brings data warehousing and BI reporting to the table. They started out as a database management solution before moving towards the CDP space, so they bring some added storage capabilities that other CDPs don't. They were recently acquired by Softbank-owned ARM.

⊕ Pros

- Treasure Data provides a full stack solution, which includes storing your data in a data lake, which can replace data warehousing, and reporting, which can replace BI tools.

- They do have the ability to ETL to Redshift if you want your own data warehouse, or export out to BI tools if you want a full BI tool.

- They're strong with collecting data and resolving user identity.

⊖ Cons

- Their focus today is really on data pipelines as a service, so their ancillary elements (data storage, reporting, data capture) are not as strong or well fleshed out as comparable vendors in those spaces.

- UI is fairly minimalistic, which may pose challenges for less technical users

- Aren't as strong on segmenting users and sending those segments to tools for activation purposes as some other tools in this space are.

- Data management is a challenge. No ability to delete records or update tables for the data that you collect and store with Treasure Data.

$ Price

- N/A

- Treasure Data keeps their pricing structure and info close to the chest. You'll have to contact them to learn more.

Best for

- Organizations with marketing-led data collection and user identity resolution needs.

- Less technically sophisticated organizations that don't have the resources to focus on building out their own stack and infrastructure.

Other Tools and Takeaways

As the CDP space is very new and not well-defined, there are lots of tools that partially fit into this category or that are entering the scene every day. Some other tools that are worth a mention in this chapter are AgilOne, Blueconic, Optimove, and RedPoint, but there are more than 60 tools that claim some connection to the CDP category in one way or another, so this merits further exploration based on your specific needs.

TL;DR   The purpose of this layer in your stack is to bring omnichannel orchestration to the different technologies and tools in your stack where customer data and customer interactions occur. What exactly denotes a CDP is not currently well-defined, nor are the capabilities that a CDP should offer, but putting that aside, the goal when designing your stack is to make sure that no important data gets stuck and siloed. This can be addressed by using a tool like Heap or an ETL tool to send all your important data into your data warehouse, or with a tool from this chapter to collect and send data between your other tools and technologies. Your exact needs will depend on your business requirements and the other decisions you make for your stack.

References and Resources

- https://www.targetmarketingmag.com/post/what-is-a-cdp-how-does-it-differ-from-other-customer-data-tools/

- https://www.gartner.com/smarterwithgartner/marketers-to-guide-customer-data-platforms/

- https://www.cdpinstitute.org/cdp-basics.html

- https://www.martechadvisor.com/articles/data-management/investing-in-cdp-choosing-the-right-vendor/

Visualizing and Exploring Data: Business Intelligence Tools

Business Intelligence (BI) and data visualization tools operate as the output of your stack in many ways. Once you've captured, connected, and organized all your various elements of customer data, you're ready to visualize and analyze it. That's where your BI tool comes in. BI tools generally sync up directly to your data warehouse and generate the graphs and charts that makes understanding and visualizing your data easy when you query it.

BI tools serve three main purposes:

(1) Last mile modeling: BI tools are excellent for transforming fact tables (modeled data in your warehouse) into visualization tables. This sort of last mile modeling is usually done in BI tools, because the questions you want to ask here are hard to do in SQL. Generally, you'll be converting dimensions and measures into tables for visualization. Some BI tools can handle earlier stages of modeling through the tool, while others leave it to you to do so via SQL.

(2) Visualization: BI tools make amazing ranges of graphs and charts based on your visualization tables. These are consumable by people in your organizations for making data-based decisions.

(3) Sharing and democratization: BI tools are useful for embedding saved dashboards and making them widely accessible to the appropriate people and teams in your organization. Most BI tools will have sharing, permissioning, and versioning functionality as well. The strength of this functionality becomes more and more important the larger your organization gets and the more you deal with sensitive data.

"The requirements gathering step often gets short-shrift, but it's vitally important. The requirements for a tool should not be sourced from the data team alone. The ideal state of play is one where you have dashboards that answer 80% of the needs of all of your stakeholders. One way to achieve that is to ask stakeholders how they're currently using data and what would make their lives easier. From that, create a wishlist that goes beyond the basic requirements to compare vendors against. The goal is to bring down the walls and make data accurate and easily available."

LISA PENELTON
SVP OF MARKETING SCIENCE, CRITICAL MASS

Important considerations for this layer

BI tools differ around the last mile modeling function, sharing and democratization, target user groups, and a bit around visualization capabilities. When considering which tool is right for you, here are some questions to ask yourself:

Who do you want to be your primary users of this tool?

Data scientists and analysts well versed in SQL, or marketers and other business users who need more of a drag-and-drop type UI?

How does the tool address modeling?

Is it more of a full stack solution that does modeling in the warehouse, or is it only focused on visualizations?

What extent do you need more visualization features vs a lower price point?

Do you have several business-critical usage needs, or will a more narrowly-focused tool suit your requirements?

Who will be consuming your dashboards and visualizations?

Do you need very specific visualizations and dashboards, or are your needs more generalized for this function?

Notable Tools in this Layer

As per the rest of this ebook, we'll be focusing our coverage of these tools towards ones that are cloud-native or at least well-suited to fitting into a cloud-native technology stack, with an eye towards the needs of small to mid-size businesses. These tools are priced on a $-$$$$ scale based on available information.

Looker

Founded in 2012, Looker made their mark as a full-stack BI tool for a more technical audience. These days, they are an excellent user-friendly option for a mix of both technical and non-technical users. A standout for Looker is LookML, which provides a virtualization and modeling layer for users to model data in the warehouse.

⊕ Pros

- LookML. LookML is far and away a standout differentiator for Looker. It's extremely useful and brings an entire modeling layer that most other BI tools don't offer.

- Looker enables you to build more than reports. You can build data applications that can serve the exact needs of your organizations. For example, you can build a data application in Looker that allows people to build self-serve reports of their own.

- Robust feature set, with offerings like versioning and fine-grained controls and configurations.

- Strong functionality for non-technical users with drag-and-drop reports.

⊖ **Cons**

- A bit challenging for analysts to run ad hoc analysis, thanks to aspects like version control.

- More challenging set up than other BI tools, due to the need to set up LookML as well.

- Some limitations around visualizations when compared to the enterprise-level offerings in this space, but generally, they'll satisfy the vast majority of needed use cases.

ⓢ **Price**

- $$$

- Looker customizes pricing based on the specific needs of your organization. Generally, they price based on number of users, scale, and database connections.

- Often, Looker is the most expensive of the options that an SMB would consider.

Best for

- Organizations that want to do data modeling. LookML makes that very easy.

- Companies with a mix of analysts and non-technical users who want to use their BI tool, and organizations that want company-wide reporting.

- Organizations looking for a tool that will scale with them as they grow from a mid-size business to an enterprise.

Mode

Founded in 2013, Mode is focused on serving the BI functionality needed by analysts and data scientists. They've chosen to not focus as much on business users or company-wide reporting, and instead optimize their tool for their target audience. This lets them be a smaller, cheaper, lightweight tool.

⊕ Pros

- Excellent sharing and collaboration capabilities. You can build notebooks of analyses in Python that are easily shareable, for example.

- Delivers Python, SQL, and R functionality all together.

- Great layout and usability for querying.

- Offers a lot for users who are well-versed in SQL. Mode is feature-rich for complex data analysis needs.

⊖ Cons

- Dashboarding is not as advanced as other options in this space.

- Company-wide reporting is not their focus. It's difficult for business users or non-SQL-familiar users to do much with Mode.

- Search and easily repeatable queries can be a challenge.

⑤ Price

- $

- Mode is one of the least expensive BI tools in the space.

- They offer an excellent free tool for individual users.

👐 Best for

- Analysts and data scientists who want to share their reports with other analysts or data scientists.

- Analysts who often need to run a series of analyses and present a case for them.

- Analyst-driven organizations who can have their needs met with a less expensive, lightweight BI tool.

Periscope Data

Founded in 2012, Periscope Data sits between Mode and Looker in the BI space, as far as their approach to BI and their target user group. They focus on empowering data professionals, and differentiate around the fast querying speed.

⊕ Pros

- Good dashboarding functionality.

- Fast querying: Periscope caches queries you run in an optimized Redshift instance, so that when you repeat those queries in the future, they run much faster.

- They have a good data lake integration.

- Offer Python, SQL, and R functionality together.

⊖ Cons

- No modeling layer beyond SQL querying.

- Difficult to use for users that don't know or are less comfortable with SQL.

💲 Price

- $$

- Pricing is not publicly listed for Periscope, but they take into account a number of factors, such as the number of rows of data, the number of users, and your sharing requirements.

Best for

- Those well-versed in SQL and technical users.

- Organizations that rank query speed in their BI tool as very important.

- Those whose main goal is consuming dashboards, as opposed to enabling dashboard exploration for non-technical users.

Power BI

The other major BI tool in the enterprise space besides Tableau, Microsoft Power BI, has the distinction of being the most white-labeled BI tool out there. Founded in 2013, Power BI is well integrated in Microsoft's stack, and also shares a partnership with Adobe.

⊕ Pros

- First class integrations with Microsoft tools and products.

- Strong visualization capabilities.

- Good cost-to-value ratio for an enterprise BI tool. It's easy to start small and scale up with Power BI.

⊖ Cons

- Support is weaker than what other tools in this space offer.

- Vendor lock-in: Doesn't integrate as smoothly with tools outside of the Microsoft ecosystem.

- Collaboration and sharing functionality is not as developed as other tools in this space.

$ Price

- $$

- Power BI is free for basic use and has a relatively affordable Pro license.

- They price on a per-user basis outside of the Premium version, which can be frustrating for SMBs.

- For high-end BI functionality in terms of users and size of datasets, more similar to what some other tools in this space deliver, you'll need to go for Power BI Premium, which ups the costs by a decent margin.

Best for

- Great for organizations steeped in the Microsoft ecosystem, as Power BI integrates exceptionally well with other Microsoft sources.

- Larger organizations who want to dip their toes in the BI space without wanting to invest a lot of time and money in a tool at the moment.

- Large enterprises and older, established companies.

Tableau

Founded in 2003, [Tableau](#) is far and away the tool with the largest market and wallet share in the space. Their focus has been on the visualizations aspect of BI and they've spent the most effort to push the envelope on that front of any tool in this space.

⊕ Pros

- Offers the most powerful and beautiful visualization capabilities of any tool in this space.

- Great drag-and-drop capabilities make for an excellent UX for business users.

- Strong data governance.

- Great community knowledge, given the size of Tableau's market and how long it's been around.

- Good for cross-database needs. If you don't have a data warehouse and instead have several disconnected SQL databases, Tableau is great at extracting the data from these.

⊖ Cons

- Not well suited for handling complex production data or ad hoc analysis. Performance and complex calculations are a weakness.

- Tableau does a little bit of the data modeling that some other BI tools in this space do, but not to the same extent as the modeling leaders.

- Collaboration and sharing functionality is not as developed as other tools in this space.

ⓢ Price

- $$$

- The price range for Tableau depends on the type of deployment you want to go with. Assuming a cloud-based SaaS model, as fits the cloud-native stack we're laying out here, Tableau is pricy, but not a giant leap more than some other options. On-prem and other deployment methods are a different ballpark as far as price goes.

- Tableau prices on a "per-seat" basis, which is often a bit frustrating to navigate for SMBs.

Best for

- Business users and organizations driven more by the needs of their business users as opposed to analysts and data scientists.

- Organizations that need some on-prem capabilities or flexible deployment options.

- Large enterprises.

Other Tools and Takeaways

There are many solid BI tools available beyond what we've covered on this list. Some examples include Chartio, Domo, and Google Data Studio. There are tools beyond this more suited for on-prem deployments as well, and a wide range of tools more specifically targeted for enterprise organizations, such as Qlik, Birst, and Thoughtspot.

TL;DR BI tools are the visualization layer for your customer data tech stack. They provide last mile modeling and visualizing of the data that you make available to them. With a well-served data warehouse, your BI tool can deliver dashboards of your entire customer dataset that you can use to make data-driven business decisions. Most BI tools will have similar functionality on the visualization side of things, but will see some differences around their last mile modeling capabilities, and their sharing, permissioning, and versioning functionality. These tools also often segment themselves along the technical proficiency of their target users.

References and Resources

- https://blog.aptitive.com/power-bi-vs-looker-vs-tableau-a-ctos-guide-to-selecting-an-analytics-bi-platform-5edc519f2d12

- https://blog.fishtownanalytics.com/is-looker-the-right-business-intelligence-tool-for-my-company-afc1f750a0f9

- https://www.encorebusiness.com/blog/power-bi-vs-tableau/

- https://www.betterbuys.com/bi/looker-vs-tableau/

- https://www.stephenlevin.co/advanced-analytics-part-3-data-visualization/

Other Tools and Customer Data Sources

An overview of the customer data technology stack wouldn't be complete without including 3rd party customer data. This data enriches your event-based and backend data with further information about your customers.

3rd party data lives in an expansive range of tools that are particular to the specifics of your business, so we won't attempt to do a notable list of tools here. Some examples of these include A/B testing tools like Optimizely, e-commerce platforms like Shopify, marketing automation systems like Marketo, advertising tools like AdRoll, ESPs like SailThru, and multi-touch attribution tools, like AttributionApp. B2B and B2C organizations will have different sets of tools as well. B2B organizations will likely want to bring in data from tools for CRM, like Salesforce, or for lead scoring, like MadKudu, for example.

The important consideration for this layer is how you're going to connect tools that hold customer data to your warehouse effectively and understanding which tools hold the data that makes a difference for your understanding of customers.

Much of this data will be used for enrichment purposes. This means enriching your customer behavior data that you're capturing in the Customer Behavioral Data layer with additional customer information and actions. Customer touchpoints are omnichannel these days, so to get a complete 360 view of your customer, you need to incorporate data from email interactions, web and mobile, direct mail responses, live chats, etc. The tools in this layer are where a lot of those interactions happen and potential enrichment data is stored. To connect and unify this information, tools from the CDP chapter or some from the Customer Behavioral Data chapter that enable you to enrich your behavior data directly are generally needed.

TL;DR You have a lot of 3rd party customer data in various tools that exist for business purposes. These are tools that do things like A/B testing, personalization, your email service provider and other messaging (texting, push notifications), or your e-commerce platform. To make your customer data tech stack most effective, you need to bring this data into the picture and unify it in your data warehouse. Many of these tools can also take actions based on the customer identities you create in your stack, so sending data and updated customer segments to them is also important.
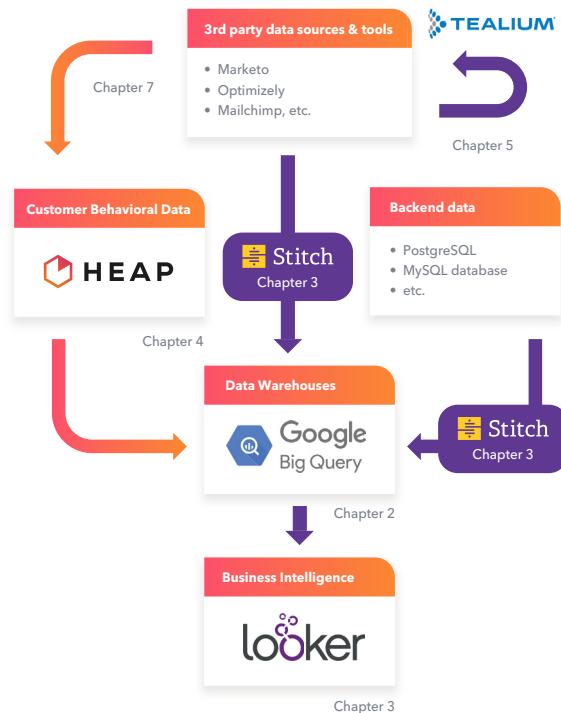
How it Comes Together

We've talked about the different layers of your customer data technology stack and what each layer does, but now it's time to talk about how these tools work together.

Implementing tools in a vacuum does you no good. To figure out the best ways for these puzzle pieces to come together, you'll want to keep in mind your goal for creating out this stack in the first place. You may have specific outcomes you want to achieve, but generally, the purpose of this stack is to give you a clear 360 degree view of your customers that you can use to drive customer interactions and business decisions across your organization. To achieve that, you'll generally want to ensure that you're capturing all the data that make up this picture, unify it all in your data warehouse, and leverage it for actions and interactions.

To sketch out the general way a stack works together, it's easiest to think about the way data flows and comes together. Here's an example of how the different chapters we've talked about can flow into one another and work together:

When thinking about the connecting and interoperating aspect of your stack, it's important to dive into how well the tools you're evaluating plug into your stack. Just because someone lists an integration on their website, doesn't mean it's a good one. For example, you'll want to make sure that your customer behavioral data tool has a deep and well-supported integration with the data warehouse you're using.

Regardless of how you want to set up your data pipelines, the key is to ensure that data is not siloed and that all your relevant data ends up unified in your data warehouse. Depending on the particulars of your business needs, the ways you interact with your customers, and the technologies you currently have, you may not need a tool from a particular chapter we've written about here. Certain tools have functionalities that allow them to work across a couple layers. Or you may find that you'll want multiple tools from a chapter to work in tandem to take advantage of their unique strengths.

Real world examples of customer data technology stacks

Here at Heap, we have some awesome customers. When we told them about this ebook, they were more than willing to offer a look into their real-life customer data technology stacks to share with you as examples. We've decided to highlight companies of different sizes, industries, and level of available resources. Here are a couple ways that the tools and stacks we've been talking about in this ebook actually get implemented in real-life:

App Annie

App Annie delivers the most trusted mobile market data and insights for your business to succeed in the global app economy. Over 1 million registered members rely on App Annie to better understand the mobile app market, their businesses, and the opportunities around them.

For their customer data tech stack, App Annie captures user activity in their platform and on their site with Heap, which they feed into their Amazon Redshift data warehouse. In the warehouse, they also bring in 3rd party data from tools like their CRM and marketing automation tool with Pentaho. In the warehouse, they use Pentaho to transform event tables into fact tables that they can then pull and visualize with Tableau. They also use Heap SQL to visualize and query Heap data directly in Tableau as needed.
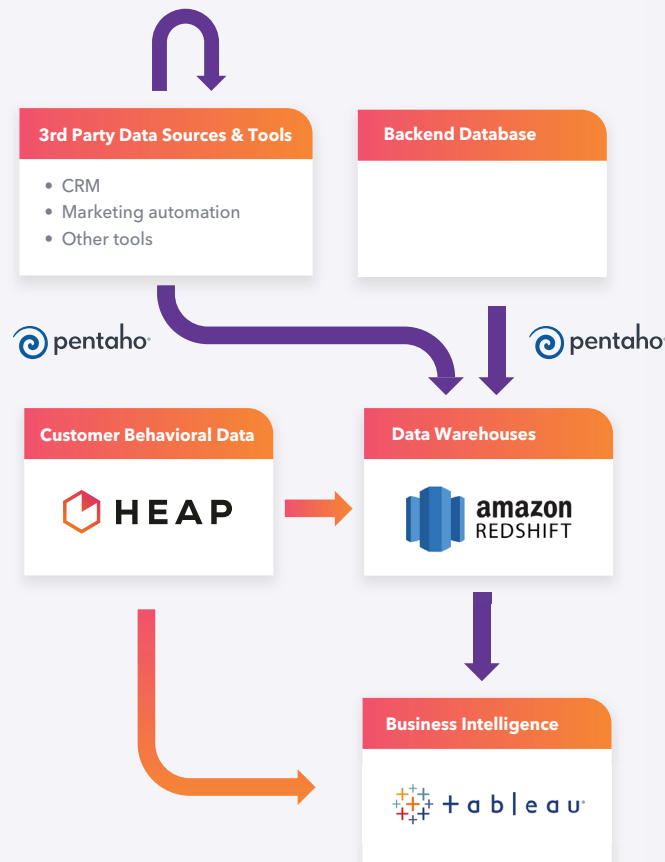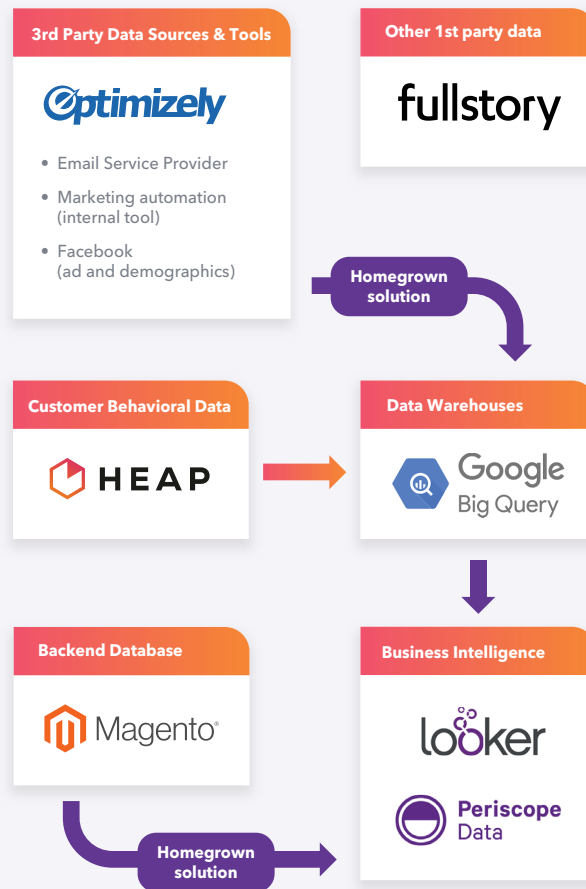
βetabrand

[Betabrand](#) is a crowdfunding platform for fashion and apparel. Their users submit ideas and solicit feedback from the community, and the best ones are chosen for crowdfunding. If enough people buy the prototype, they make a full production run.

For their customer data tech stack, Betabrand uses BigQuery as their data warehouse. They use Heap to gather customer behavior data and for performance monitoring on their platform. Heap then sends that data to their warehouse through its direct integration with BigQuery, and they've built homegrown solutions for sending data from their primary 3rd party data sources to BigQuery as well. They use Looker and Periscope as BI tools, and connect them to BigQuery and directly to a Magento MySQL database. Additionally, they use Fullstory for session replay analysis, which lives in isolation from the stack.

Betabrand is at a size and level of technical sophistication internally where they decided that it makes sense for them to handle most of the ETL and Connector & Distributor layers with their own internally built and maintained solutions.

3rd Party Data Sources & Tools

Optimizely

- Email Service Provider
- Marketing automation (internal tool)
- Facebook (ad and demographics)

Other 1st party data

fullstory

Homegrown solution

Customer Behavioral Data

HEAP

Data Warehouses

Google Big Query

Backend Database

Magento

Business Intelligence

looker

Periscope Data

Homegrown solution

slab

[Slab](#) is a modern wiki that helps you capture, retain and share your team's knowledge and expertise. It features an intuitive interface, blazing fast search and numerous integrations like Slack, GitHub and GSuite to tie everything together.

Slab's team has domain experience in productivity and collaboration, having founded and sold a YCombinator-backed document collaboration company, shipped products at Airbnb and Dropbox, and maintains one of the most popular [open source projects](#).

For their customer data tech stack, Slab uses Redshift as their data warehouse, and populates it via Segment for their 3rd party data sources and via Stitch for their backend PostgreSQL database. They use Mode for visualizing and querying the data

in their warehouse. They also use Heap's free tier to capture and analyze customer behavioral data from their website and Fullstory for session replay analysis, neither of which they send to their warehouse.

As Slab is an early stage startup, they've looked to find tools with a good price-to-value ratio to fill the different layers and that enable them to focus their time and resources on their product. Today, they've built up a mix of tools that fit their current needs and price points, but that also offer them the ability to scale up over time.
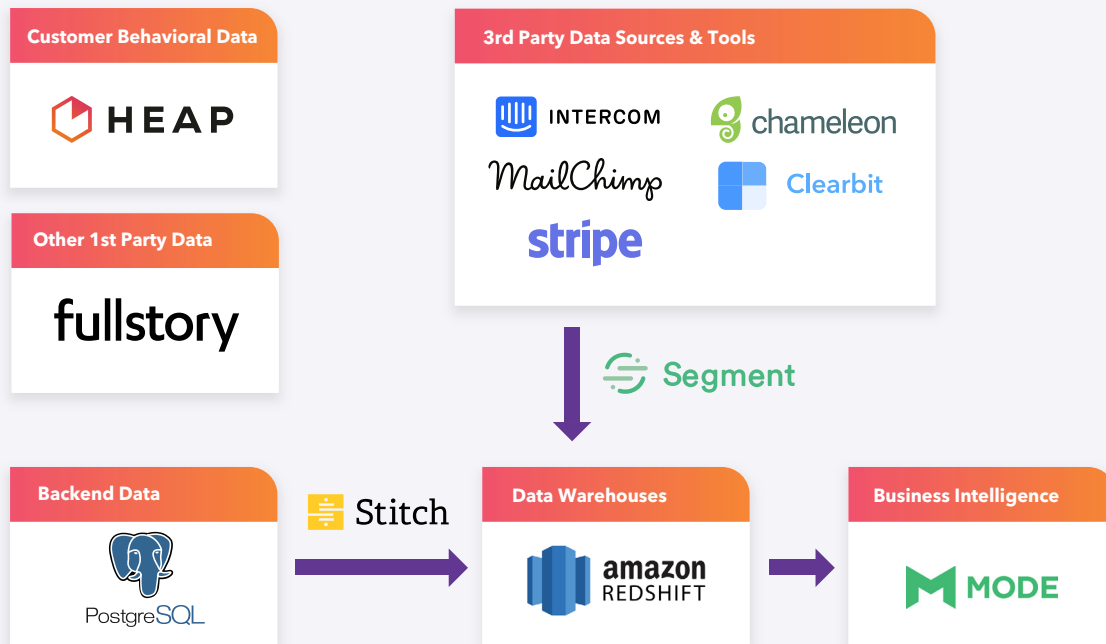
Conclusion

In the beginning of this ebook, we talked about starting with the goals you want to accomplish when thinking about how to design your stack. A strong, well thought out stack should not only empower you to accomplish those goals, but should leave you with a certain set of beneficial outcomes.

- **First,** it should free up your data engineering team to focus on core business initiatives, as opposed to spending all their time on maintenance and upkeep.

- **Second,** the output from your stack should be a complete and actionable customer behavior dataset that you can use to accomplish tangible business activities and goals.

- **Third,** your tools and layers should work well together. The integrations and way data moves through your stack is often overlooked but extremely important, and can save you many headaches in the future.

- **Fourth,** it should set you up well for the future. While you shouldn't necessarily be paying for a bunch of scale and performance that you don't need, you should be building a technology stack that gives you the power and freedom to scale up as your business grows and needs change. If you build your stack around the available integrations for a small free tool, and then find that you'll need to replace it in 18 months, you'll have backed yourself into an awkward corner. Build your free tools and tools that you'll need to replace beyond a certain scale into your other stack decisions, and not vice versa, or at least ensure that your free tools offer the option to scale up to more robust plans.

Throughout this ebook, we've presented a guide on how to approach and build your customer data technology stack in a cloud-native manner. In this ebook, we've outlined the layers of a modern customer data technology stack, and taken a look at some of the notable tools in each layer. We explored data warehouses, ETL/ELT tools, customer behavioral data tools, CDPs, BI tools, and other sources of data, as well as how these different tools come together to create a stack. Each business is different, as are the needs for your business. For some organization, you'll be able to skip a layer or two of the stack we've outlined here. Other organizations will need or want to use multiple entrants from a layer at the same time.

In the end, you'll want to ensure that you choose a mix of tools and technologies that integrate well together, that fit your budget and your business goals, and that enable you to have a unified understanding of your customers and their journey.

HEAP

Heap automatically captures every customer touchpoint and automates away the pain of data. Other analytics tools require you to tag events upfront and manually instrument tracking code. Instead, Heap automatically captures everything: clicks, taps, swipes, form changes, and more. Get answers in seconds and make decisions faster.

Behavioral Segmentation

Complete history on every user. Don't miss out on unknown unknowns. Automatically capture every event and easily build segments based on behavior.

Unified Customer Profile

Connect anonymous and cross-platform behavior to a single identity. Heap's best-in-class identity resolution creates a unified view of your customers across mobile, desktop, email marketing, and more.

Retroactive Funnels

Dynamically change your funnel events and go back in time with retroactive data that's instantly available. Discover where users drop off and compare how cohorts convert.

Clean Schema

Data teams spend up to 80% of their time cleaning and organizing data. Heap's structured user-event schema remains constant through naming convention changes and event combinations. Any changes automatically update in Heap's reports or your data warehouse, which means less time organizing and more time gathering insights.

If you're interested in learning more about Heap, or if you have any questions or comments regarding this ebook, feel free to email us at **sales@heapanalytics.com** or sign up for a free trial at **heapanalytics.com/signup**.

Credits and Thanks

Written by: Josh Dreyfuss

An effort like this doesn't happen in a vacuum. I'd like to thank everyone who contributed their time and expertise to making this ebook what it is. Having people share their direct experiences with the different tools throughout this book and their approaches to building customer data technology stacks made this book much more thorough than I would have been able to create on my own. It's impossible to thank everyone, but I'd like to call out a few names.

Thanks to:

- Ravi Parikh, Neal George, Taylor Udell, Nick Erdenberger, James Katz, Kevin Moyer, and Chrix Finne for sharing their knowledge, experience, and expertise with the different tools and layers of the customer data tech stack.

- Shawn Hansen, Brian Tecklenberg, Charlie Liang, and Dan Robinson for sharing their worldviews and helping to make this ebook happen.

- David Saxon for helping me find the best experts around to interview.

And a special thank you to our external experts who contributed:

- Anvisha Pai from Slab, Seamus James from Betabrand, and Ishpreet Singh from App Annie for their willingness to share their stacks and real world deployments with me.

- Jonathan Mendez from Arkle Advisors, Ryan Koonce from Mammoth Growth, and Lisa Penelton from Critical Mass for sharing their agency and consultancy experience with different tools, layers, and how they fit into the stack.

HEAP

HEAPANALYTICS.COM