

Build a Modern Data Stack in 30 Minutes

Philosophy Benn Stancil

Author Chris Moffitt

Edited by Jessica Schimm

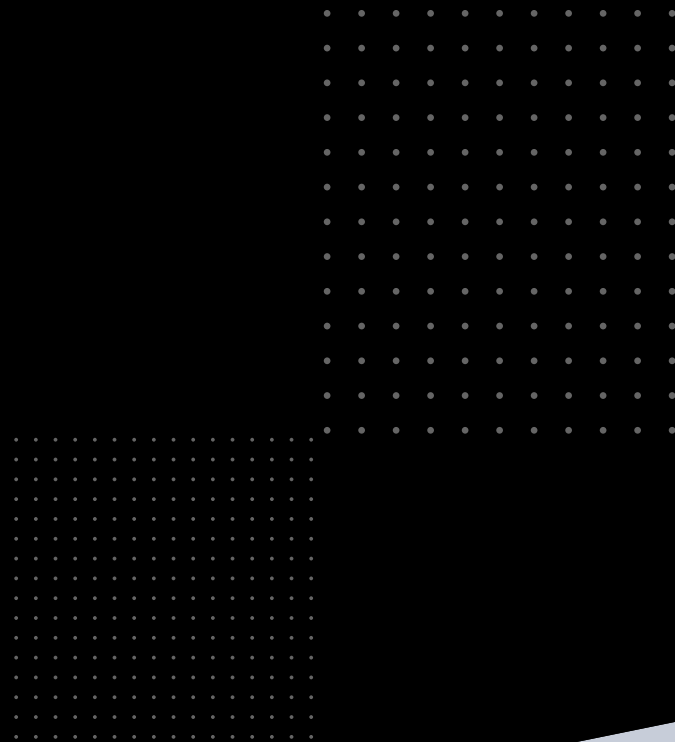
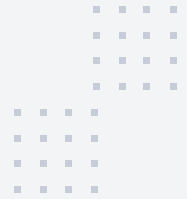


Table of contents



01	Introduction	03
02	The benefits of a modern data stack	05
	Quickly get insights to growth metrics	06
	Adapt to new data analytics and BI technology	07
	Unblock stakeholders for rapid collaboration	08
03	The components of a modern data stack	09
	Cloud-based data warehouse	10
	Data pipelines	10
	BI and analytics platform	10
04	How to build a modern data stack in 30 minutes or less	11
	Step 1: Choose a cloud-based data warehouse	12
	Popular options: Redshift, BigQuery, Snowflake	
	Step 2: Integrate data from all of your applications	14
	Popular options: Fivetran, Stitch, Segment	
	Step 3: Analyze the data	17
	Popular options: Mode, Looker, Tableau	
	Connect Mode to your cloud-based data warehouse	18
	– Online SQL Editor	
	– Notebooks	
	– Visualizations	
	Step 4: Data prep and transformation	19
	Popular options: dbt, Dataform, Airflow	
05	Get to answers faster with a modern data stack	20

01 Introduction

“ How do I deliver value quickly when I have to build infrastructure first?”

This is a common question for data teams at many early-stage startups—and understandably so. Startups [survive on growth](#), [analysis fuels growth](#), and good analysis depends on an organized, reliable data infrastructure.

For analysts to deliver value quickly and iteratively—a clear [indicator of success for data teams](#)—the right infrastructure needs to be in place first.

And while data infrastructures are complex, it is indeed possible to build your infrastructure quickly and for as little as a few hundred dollars a month.

After optimizing our own tech stack and observing the solutions at hundreds of other companies, we’ve identified a best-practice architecture that can accommodate data teams’ current and future analytics needs. We call it the modern data stack.

**We call it the
modern data stack.**

And it’s designed to accommodate best-in-class tooling for various stages of data analysis—and get you there fast.

Introduction

You can set up this modern data stack in 30 minutes

It's a bold claim, but we're standing by it. Historically, setting up end-to-end data tooling was an expensive endeavor.

However, with the rise of **a)** new cloud-based warehouse solutions **b)** free, self-serve enterprise-grade data vendors and **c)** options for open-source tooling, organizations of all sizes can quickly build a solution that dramatically speeds up time-to-insights.

A modern data stack works for startups with just a few data sources and gigabytes of data and can be adapted as the team collects terabytes of data from disparate data sources. There is no need to continually engineer the solution during this high-growth period.

This modern architecture can also be used by larger organizations that have existing data stacks and warehouses. In this case, they can benefit by introducing components of this structure as part of a longer-term migration strategy or by directly moving to this architecture as part of a wholesale migration.

No matter your company size, you can have a cloud-based warehouse with data from multiple sources that's connected to an analytics/BI platform, up and running fast.

A modern data stack works for startups with just a few gigabytes of data and can be adapted as the team collects terabytes of data.

The benefits of a modern data stack

To turn around insightful reports for your company, you need to put a data infrastructure in place quickly.

There are immediate and long-term benefits to using a modern data stack. Unlike all-in-one solutions, this architecture is designed to quickly integrate new tools (which your competitor may or may not already be using). It also lets you choose tooling with languages that your team already knows and prevents technical debt by allowing teams to swap in tools when you're limited by what you can do with your data.

Perhaps the biggest reason to adopt a modern data stack, however, is that it speeds up the time it takes to arrive at insights—a particularly critical benefit for fast-growing startups.

The problem with end-to-end solutions

While claiming to have everything you need, end-to-end solutions (with data storage, ETL, and visualization) create vendor lock-in, don't provide best-of-breed tooling, and are not guaranteed to accommodate new technologies quickly.

Some of our customers started with end-to-end solutions and still purchased Mode to perform richer data analysis and to plug in preferred tools that these solutions wouldn't allow.

The benefits of a modern data stack

Quickly get insights into growth metrics



Early-stage companies with lean teams and limited runway need to grow quickly to survive. Startups don't get the luxury of exploring multiple directions to see what works, but they can use data to prioritize where to invest next.

With the ease of setting up a modern data stack, you can get accurate growth metrics like signups, daily active users, monthly revenue, retention, quick ratio, payback period, and LTV / CAC in a shorter time period. (Learn more in our talk about [analytics for fundraising](#).)

Without a proper infrastructure in place, a simple, ad hoc question about revenue becomes time-consuming and frenetic to answer: You'll realize that financial metrics have been tracked in both Excel and in Google Sheets, and that some have been copied directly out of Stripe.

What should be a relatively simple task will be delayed by rounding up data.

Growth metrics and reports on revenue that require data from multiple sources can be a time-consuming manual data gathering process. A modern data stack automates much of this challenging work.

A modern data stack automates much of this manual data gathering process.

The benefits of a modern data stack

Adapt to new data analytics and BI technology

Another primary benefit of a modern data stack is that it can be set up with minimal engineering resources. It can also change, adapt, and grow as the business scales.

Most companies follow a similar pattern as they grow. When starting out, engineers will write simple SQL queries against their application database. This early phase is sufficient to get started but quickly starts to break down as **1)** the data needs grow more complex and **2)** the user base expands. As the business grows, the data collection includes more sources such as Google Analytics, Amplitude, Heap Analytics, and in-application reporting tools such as Salesforce.

What started out as simple SQL queries can grow into a complicated data architecture as files such as spreadsheets, CSVs, and JSON data are brought into the mix.

As more third-party applications are used, a significant amount of engineering time can be spent integrating them with the existing data.

In this ad-hoc environment, there is no single source of truth and no consistent way to incorporate new data. Thus, a significant amount of time is spent figuring out how to get the data into a suitable format that can be merged with other data sources.

Adjusting large amounts of data to fit other platforms is laborious and not an efficient use of your engineer's time.

With a modern data stack, you are able to swap data sources and tooling in and out more efficiently—minimizing the amount of technical debt you are incurring and minimizing setup costs.

With a modern data stack, you are able to swap data sources and tooling in and out more efficiently.

The benefits of a modern data stack

Unblock stakeholders for rapid collaboration

Finally, as the business grows, more users from other lines of business will need access to data so they can answer more complicated questions about the health of the company and how to grow the customer base.

A modern data stack lets you pick the best tools for non-technical users. As you want to add tooling to help different teams, you'll be able to plug them into your stack quickly, allowing richer insights to be discovered across teams.

By using a modern data architecture, companies can be confident that their important and varied data can be captured, stored, and efficiently analyzed by all members of the team.

In sum, the modern data stack offers many benefits:

- Faster data insights, by leveraging known technologies such as SQL, JavaScript, Python, and R
- A single source of truth to consolidate information
- A variety of tools to choose from
- Rapid collaboration between data teams and business users in analytics platforms

A modern data stack lets you pick the best tools for non-technical users.

03

The components of a modern data stack

A modern data stack is built on the foundation of three main components: a cloud-based warehouse, data pipelines, and an analytics platform.

This architecture leverages the availability and scalability of today's cloud-based platforms, which can handle large datasets that have predictable and manageable cost structures.

The process of setting up these services is simple, future-proof, and requires limited technical input from engineers—but we'll get to that further down. First, let's take a look at the pieces of a modern data stack.

The process of setting up these services is simple, future-proof, and requires limited technical input from engineers.

The components of a modern data stack

A modern data stack is built on the foundation of three main components.



Cloud-based data warehouse

The first component is a cloud-based, warehouse that serves as the central location to collect all of the organization's data.

Tools:

Redshift, BigQuery, Snowflake



Data pipelines

The second component is one or more pipeline services to ensure that data is fed seamlessly into the warehouse with minimal engineering effort.

Tools:

Fivetran, Stitch, Segment



BI and analytics platform

The third component is a powerful data science platform that can take advantage of the consolidated data warehouse to analyze data.

Tools:

Mode, Looker, Tableau

04 How to build a modern data stack in 30 minutes

To help you set up your data stack, we've created a step-by-step guide with tool recommendations. We'll review options for cloud-based warehouses, pipeline tooling, and walk through Mode's features that will help you




Follow alongside Mode's co-founder Benn Stancil as [he sets up this data stack in a live demo.](#)

Note: With an introduction and Q&A, this video is technically an hour long. But the entire process—from setting up a data warehouse to analysis—is roughly 26 minutes. We've cut to the corresponding section for each step.

Step 1 Choose a cloud-based data warehouse

The first step is setting up a cloud-based warehouse—the core of the modern data stack. Before cloud-based solutions, organizations would have to spend significant time and resources engineering an environment to store, process, and query data. As the data needs become more complex, organizations would spend more and more time trying to manage this environment.

Today, with cloud-based solutions, many of these performance and data-scaling details can be adjusted through an admin interface with little to no involvement from the engineering team.

→ Follow along – [Set up your data warehouse](#) 

Popular options: Redshift, BigQuery, Snowflake

We'll be focusing on three common solutions for the cloud-based data warehouses: Redshift, BigQuery, and Snowflake. Each of these solutions have several common beneficial characteristics:

Column-oriented databases that are optimized for efficient storage and fast querying of petabytes of data. These solutions are not a replacement for operational databases, such as PostgreSQL or MySQL, but are purposely designed for warehousing and analyzing large amounts of data.

Flexible billing approaches that allow you to tailor the cost structure to your unique needs.

A rich ecosystem of third-party support and connectors for each of these solutions.

SQL-first platforms in contrast to a noSQL solution like Hadoop which requires bringing a new query language into the ecosystem. Redshift, for example, is based on PostgreSQL, which is well-understood and adheres to most SQL standards.

Cloud-based warehouses

The table below summarizes a few of the key aspects of each of these solutions.

Cost Factors	Redshift	Snowflake	BigQuery
Separated storage and compute	No	Yes	Yes
Query speed	Scalable	Scalable	No flexibility
How cost is computed	The cost depends on size and number of nodes not on usage. Offers usage-based option on top of data stored in S3.	The cost depends on the uptime, not the amount of data you use or how many queries are running concurrently.	The cost depends on the total amount of data scanned across all of your queries. Query runtimes don't impact cost.

*We use Snowflake for sharing features, lower maintenance, and scalable compute, but recommend Redshift for early-stage companies because it's easy to set up (no sales process necessary) and has a pay-as-you-go option.

Once you choose a cloud-based solution, it's easy to set up a brand new warehouse. In fact, once an organization is set up in Redshift, any new database can be set up via a single screen and a handful of configuration options. As needs change, the database can be scaled without extensive involvement from the engineering team.

While the initial setup can be a bit daunting due to each solution having unique configuration options, these options can be updated as needs change. For many organizations, the default settings are good enough to get started.

Because you are not purchasing any specific hardware or software, the scaling is done virtually and through a similar configuration-driven approach. Use the quick setup wizard to configure the cheapest option and then adjust as needed.

Another important consideration when choosing a cloud-based provider is how it integrates with other services. Setting up a database is important but not useful if it is difficult to get data loaded into the system. Fortunately, these solutions have enough market share that most vendors integrate with each of them. Let's take a look at the integration setup.

Step 2 Integrate data from all of your applications

Once you've set up the data warehouse, you will need to load data across multiple sources into it. While each of these warehouse solutions supports custom data-loading approaches to get you up and running, several vendors offer scalable ways to more easily load diverse data streams into your warehouse.

→ Follow along – [Set up your data pipelines](#) 

Popular options: Fivetran, Stitch, Segment

We recommend selecting a data pipeline management tool that will capture your data from various systems (e.g., CRM, billing systems, inbound marketing platforms, etc.) and store it in your data warehouse.

Many of our customers have had good experiences with tools like Fivetran,

Stitch, and Segment for integrating applications and funneling the data to the warehouse. In addition to capturing application-level data (for example, using Segment's JavaScript API), most organizations will want to integrate their own data with third-party sources like Salesforce.

Once you've set up the data warehouse, you will need to load data across multiple sources into it.

Data integrations

Instead of trying to build custom integrations, these data pipeline tools provide ready-made data integration solutions for many applications.

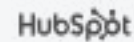
Analytics

Google Analytics, Heap Analytics, Amplitude



CRM

SalesForce, Hubspot, Insightly



Advertising

AdWords, Facebook, Twitter, LinkedIn, Pinterest



Customer Success

ZenDesk, Slack, Intercom



Email Marketing

MailChimp, Drip, Mailjet



Payments

Stripe, Square



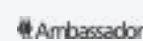
Performance Monitoring

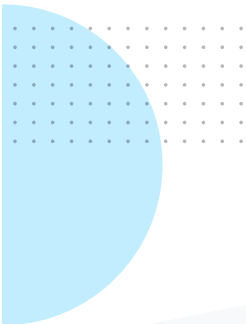
Sentry, Keen IO



Referrals

SaaSquatch, Ambassador





Integrate data from all of your applications

One benefit is that you'll only need to configure the data warehouse connection one time. Once connected to your warehouse, you can connect third-party sources easily, often through simple forms and point-and-click flows.


As you add additional data sources, you can continue to utilize the existing infrastructure to funnel data into the warehouse.

Data pipeline tools

The table below summarizes a few of the key aspects of each of these solutions.

Key aspects	Fivetran	Stitch	Segment
Transfer data into cloud-based warehouses	Yes	Yes	Yes
Included Integrations	About 126	About 70 integrations through "Singer Taps"	About 41
Trial Pricing	Free trial available	Free plan	Free plan

Once connected to your warehouse, you can connect third-party sources easily.



Step 3 Analyze the data

Now that the data is available, you need a tool to analyze it. Mode unites an analyst-centered workflow with modern business intelligence, which allows teams across your company to generate richer insights.

Companies with rich and varied data needs (especially enterprise companies) need quick flexibility for exploratory analysis—and Mode lets teams move the fastest and most flexibly out of the current BI and analytics tools on the market.

→ Follow along – [Set up and integrate Mode](#)



Popular options: Mode, Looker, Tableau

Other modularized BI and analytics tools can be used in this stack, but come up short. Looker and Tableau require data teams to model data before doing analysis—Looker through LookML, and Tableau through extracts. And while working with modeled data is generally a best practice, it can sometimes slow you down in critical, exploratory analysis.

This is a problem because even the most well-modeled data can exclude important cross-sections of data. One of our well-known customers felt the bottleneck of modeled data when they were losing revenue and needed to find the root cause of the bug quickly. Because they were able to do exploratory analysis on unmodeled data in Mode, they were able to fix the error in their app, preventing hours more of revenue loss.

In sum, you want the option to model your data, but not be forced to model it to access it.

Moreover, if an analytics or BI application has a built-in modeling layer, that layer is only useful to that application. To move fast, we recommend modeling your data in a way that's agnostic to the tools that use it, which we'll talk about in the next section. Mode is designed to be the BI and analytics layer on top of this application-agnostic data governance layer.

Designed for speed and a quick setup, our pricing model allows teams to get started without spending a lot of money. Mode lets you:

- Work with raw data stored in the data warehouse
- Share analysis through Slack and email integration or URLs
- Use a freemium version without initial setup costs

Connect Mode to your cloud-based data warehouse

The next step is to connect Mode to your data warehouse. Once Mode is connected, you can start building dashboards, models, and share reports.



Online SQL editor

Move faster in a shared, cloud SQL editor

Start working with your data in Mode's collaborative SQL editor.

Teams can use SQL on their raw data in an interactive environment and can reliably visualize millions of rows of data.

In addition, Mode improves team efficiency by including additional features:

- Schema browser
- Auto complete for tables and column
- Query history
- Language for templating queries



Visualizations

Instantly visualize query results

Use visualizations to understand data and quickly act on trends or anomalies. You can instantly visualize query results on millions of rows (larger than Excel) and build custom dashboards to tailor their look and feel for your own organization's needs. With Mode's visualizations you can:

- Choose from different chart types, preset and custom color themes
- Build custom data visualizations with an HTML editor
- Add narrative text for your audience



Notebooks

Focus on results, not setting up Notebook environments

SQL results can be seamlessly integrated into Python and R Notebooks for complex analysis and modeling—all in a familiar format. Each Notebook environment includes popular libraries in Python or R for:

- Statistics and exploratory analysis
- Machine learning
- Data visualization
- Web scraping and natural language processing
- Ability to install custom packages using pip or install

Step 4 Data prep and transformations

One of the last steps you'll want to take is to add a modeling layer for business logic. We recommend a modeling layer that's decoupled from other parts of the stack (as opposed to one that's built into your database or BI application). This makes your modeled data accessible to all the tools that might use it.

Popular options: dbt, Dataform, Airflow

Both dbt and Dataform work in modern data stacks. They're both SQL-based, provide integrated development environments (IDEs) for model development and testing, and enable analysts to model and transform data directly in their data warehouse.

Dataform has a managed service option, while dbt remains mostly self-serve with an active Slack community. Many companies also use Airflow, an open-source technology developed at Airbnb, to manage data transformations. While Airflow can be used in this way,

Airflow's primary benefit is orchestration, not transformation. It's more flexible, but can be challenging for analysts who are comfortable with SQL.

We use dbt in combination with Mode's own Definitions (Definitions are saved SELECT statements that act like centralized and shareable views). We keep core business concepts centralized in dbt, like the canonical definition of an active user and calculate the litany of metrics that sit on top of that concept, like daily and weekly active users, in Mode definitions.

Data transformation tools

The table below summarizes a few of the key aspects of both tools.

	dbt	Dataform	Airflow
Supported databases	Google Big Query, Redshift, Snowflake, Postgres	Google Big Query, Redshift, Snowflake, Microsoft Azure	Any database backend supported as a SQLAlchemy backend
SQL-based	Yes	Yes	No, Python-based
Pricing	Free for developer seats. Partially hosted version starts at \$50/month	Free trial, then \$550/mo	Free

Get to answers faster with a modern data stack

With a modern data stack, you can go from having no infrastructure at all, to having shareable dashboards that are running on all of your live, game-changing data in 30 minutes or less.

This solution would not have been feasible a few years ago. With today's cloud-based solutions, early-stage companies can build scalable data infrastructure that will enable the organization to quickly derive insights on all their data.

By creating the modern data stack, data teams can spend more time analyzing their data and less time engineering their data processing pipelines.

New companies have goals to meet, and it's accurate, accessible data that will help them get there.



Contact us

Email hi@modeanalytics.com

Call 650.993.9125

YouTube youtube.com/mode-analytics

LinkedIn [@modeanalytics](https://www.linkedin.com/company/modeanalytics)

