



MUTT DATA

Machine Learning for Business



Meet The Modern Data Stack

The relevance of applying Modern Data
Stack solutions to achieve actionable business insights.

The Road To Modern Data Stacks

The modern data space has evolved by leaps and bounds altering the very fabric of business analytics and data management. The mass adoption of cloud data warehouses and cloud-based tools has changed how we acquire, store, access, transform, and interact with data.

Big Data revolutionized business, making companies more and more data-driven. The infinite amount of data and its endless use cases forced the data community to develop new technologies that could address the shortcomings regarding cost, speed, connectivity, accessibility, scalability, and operationalization capabilities of on-premise databases and traditional legacy stacks.

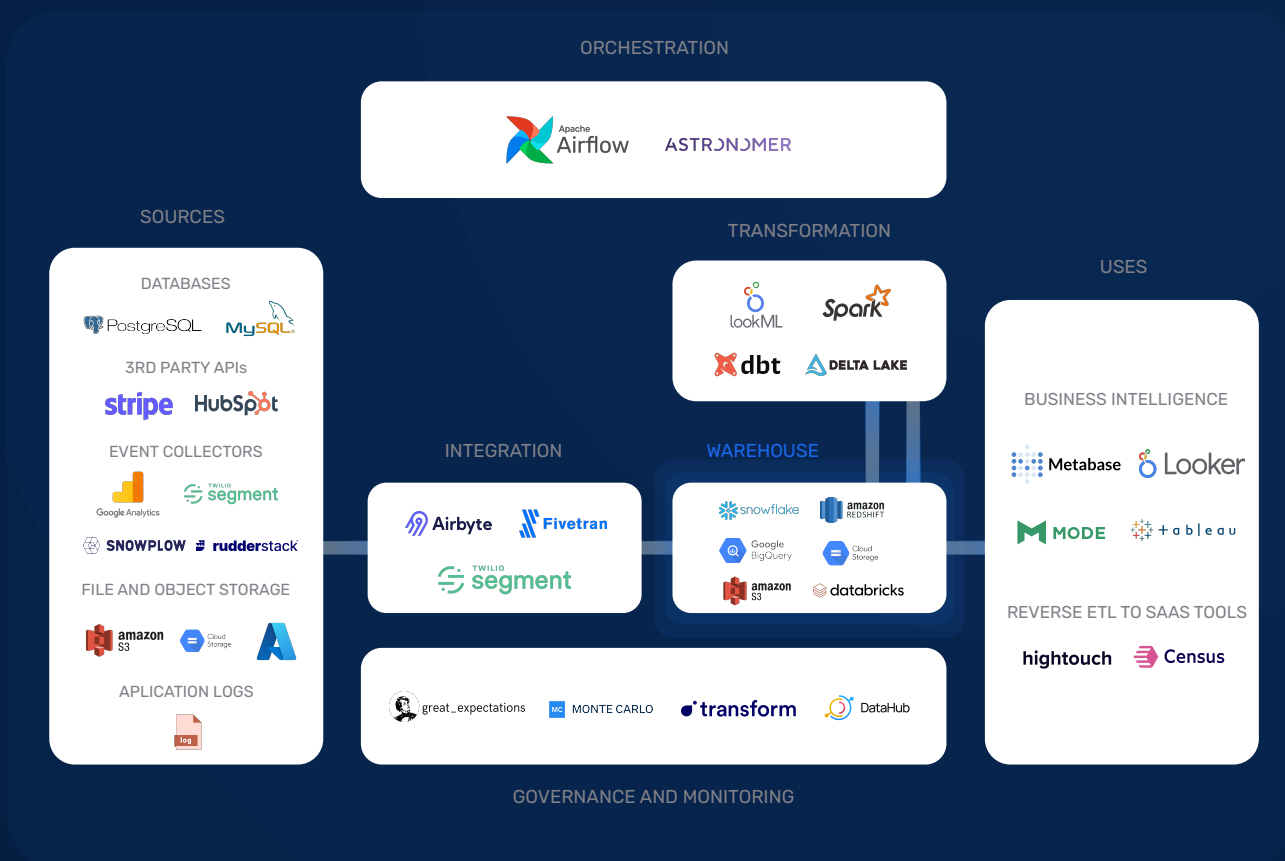
The Modern Data Stack has many benefits, but it's not only about gaining insights faster, achieving competitive advantages, guaranteeing data compliance, quality, and accuracy across the organization, or reducing costs while improving user accessibility. Although these are all fundamental ingredients to becoming a data-driven organization, there are other reasons why not having a modern data stack in today's data world can become an issue.

The volume of data has become overwhelming and is widespread across the entire organization. Without clear processes, best practices, and modern tools to organize the flow of data, it loses its value. Disparate metrics on the same data make it untrustworthy. Expert engineers and valuable resources are wasted on endless queries, and the root causes behind final user accessibility, scalability and operationalization are not addressed leading to constant patchwork instead of long-term solutions.

What Is A Modern Data Stack

A Modern Data Stack is a cloud-based, flexible, and easily scalable data infrastructure tailor-made to the unique needs of each business. The stack is composed of a collection of different modular data components that work in unison to generate clean, actionable insights by final business users.

The purpose of this stack, which centers around an ecosystem of tools, is to support the collection, movement, storage, transformation, analysis, and operationalization of data. Essentially, to make businesses data-enabled and truly data-driven.



Why Implement a Modern Data Stack

A company's technology stack is its foundation, determining the value of its data. Data-driven companies need effective and rapid data flow to the correct teams to enable valuable insights and analytics. Without infrastructure, the right tools and processes, data loses its value.

#1: Lack Of Standardized Integration

Data incoming from multiple sources with no platform integrating them, means you have to consider each data source integration on a case-by-case basis. The more sources, the more complicated and time-consuming it gets without a standardized framework to extract and centralize this data. When different sources have contradictory information, two reports on the same data but from different sources might have completely different results. The recent surge of API SaaS Services, databases, custom made systems, and FTPs forced users to design custom integrations for almost every data source used. The Modern Data Stack provides most integrations out of the box or at least a framework or method to develop them efficiently. **Multiple siloed data sources become increasingly hard to use and might contain contradictory information**

#2: Data Quality

Traditional data tools are not concerned with Data Observability, Quality, and Lineage as key variables. There is no consistent and straightforward way to assure the quality and lineage of your data: who is generating it? Are they inserting data according to the correct schema? **No way to assure the quality of data and monitor its usage. Reprocessing old data to fix mistakes is hard.**

#3: Vendor Lock-In

Before the widespread adoption of cloud-based services, companies had to pay upfront for hardware and software when implementing a Data Warehouse. Usually this came in the form of an annual engagement with a single vendor (e.g. Oracle or Teradata).

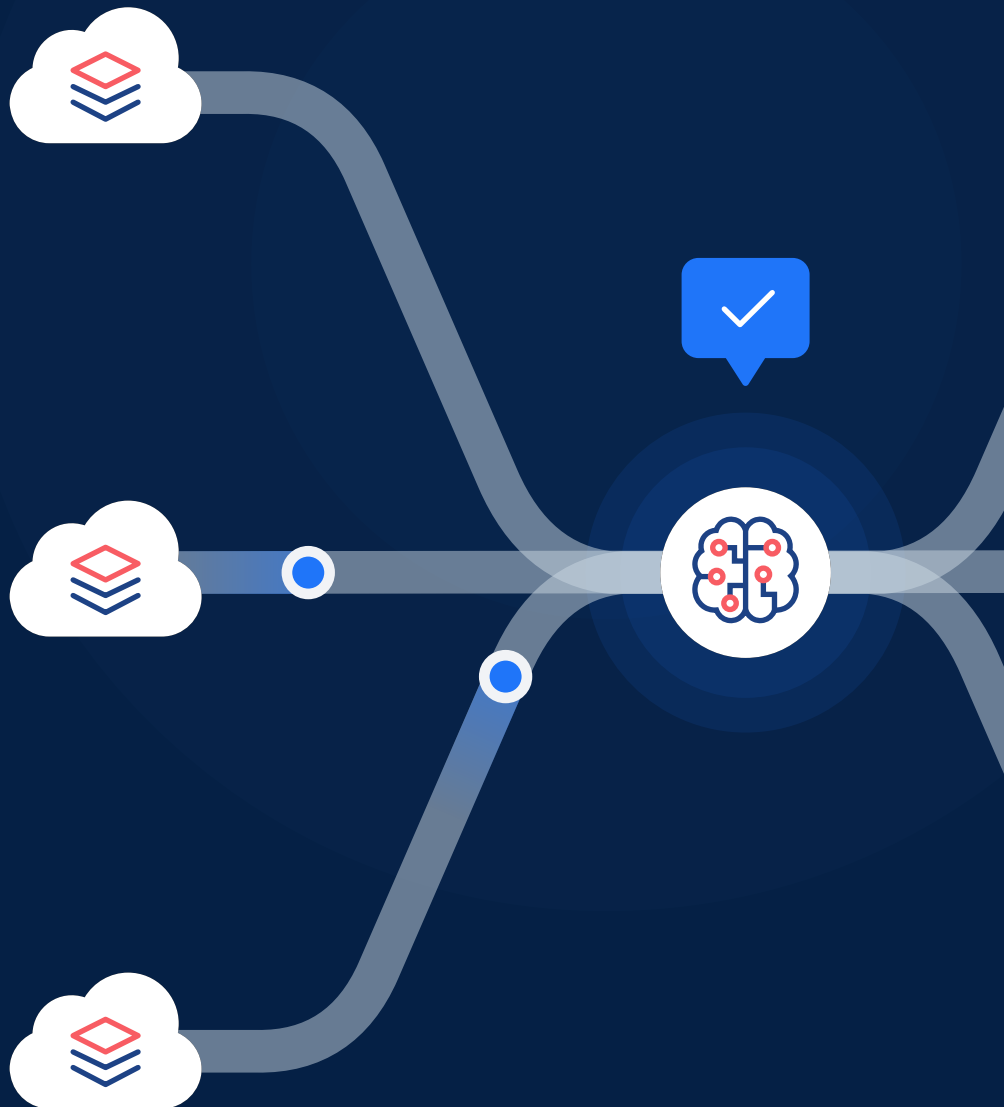
Additionally, traditional legacy solutions create a sunk-cost mentality as opposed to a pay-as-you-go mindset. Using proprietary solutions to solve data needs can lead to vendor lock-in. Many traditional licensed options are limited to the vendor's ecosystem, this can bring about some issues.

- **Pricing:** Traditional vendors force users into hefty upfront payments as opposed to pay-as-you-go models.
- **Functionality:** What can be done with data may be limited. This makes adapting to new or more demanding business needs hard.
- **Hard Exit:** Switching vendors becomes a complicated affair. If one vendor solves all components, switching one component could mean having to switch out an entire solution that might account for all the data infrastructure.
- **Risk:** If a component doesn't work out as expected, it might be necessary to migrate most of the architecture.

Vendor lock-in leads to high costs and reduced flexibility.

Classdojo saw a 30% reduction in data pipeline processing and data delivery times.

[Explore the full case here](#)



Game Changing Benefits



Modular

Modern Data Stack tools mitigate the issue of vendor lock-in, as they're designed to integrate with a greater variety of tools instead of a specific vendor ecosystem. They can be swapped out for different tools that although provide similar uses might be a better fit for the specific goals of a company. Essentially, each business can choose what they consider to be the best tool given their circumstances and constraints.

Modern Data Stack tools are designed to work easily with other tools. Integrations are usually facilitated between cloud services, tools, and platforms.



Scalable

Modern Data Stacks are cloud-based, enabling them to scale as needed to accommodate any surge in data volume, velocity, or needs without the associated costs or downtimes involved when scaling Legacy Servers and licenses. Cloud-based stacks avoid up-front payments with pay-as-you-go models designed around facilitating scalability.

Modern Data Stacks also leverage the declining costs of cloud storage and computing and are faster to set up and iterate when compared to traditional legacy stacks. In simpler terms: scaling is simplified.

Modular components, decreasing costs, and growth benefits of cloud storage and computing combined with the possibility of customization and tailor-made solutions make the Modern Data Stack much more future-proof than a traditional legacy stack.

Cost and Overheads

A Modern Data Stack reduces the need for data engineers to work on maintaining pipelines, data normalization, or querying. Costly hours are reallocated to new projects or insights.

Moreover, the total cost of ownership of deploying infrastructures on the Cloud is significantly cheaper than on-premise traditional solutions. With traditional solutions, infrastructure and other resources generally need to be estimated and paid upfront as capital expenditure (CAPEX), while ongoing maintenance and monitoring need to be owned by the client. However, Cloud services allow pay-as-you-go models and serverless models that reduce fixed and operating costs. Downtime costs when scaling or changing infrastructures are much lower.

Cloud elasticity and resource availability allow for virtual environments where you can replicate a dev environment and test changes simply without impacting the original. With Cloud Computing and storage costs declining this trend will continue to favor Modern Data Stack solutions in terms of costs.

Modern Cloud-based Data Stacks benefit from savings and cost reductions due to declining computing and storage costs, the possibility of re-allocating human resources by avoiding manual processing, and pay-as-you-go infrastructure solutions with no downtime.



Single Source Of Truth

The lack of a centralized data source where any analyst or user can cross-check data leads to users creating custom processes to do so. These processes might not be correct and may produce poor data quality or inaccurate or unmatched data between teams.

Analysts sometimes end up with contradictory results depending on the data source they're looking at. A Modern Data Stack can join multiple data sources into a Single Source of Truth (SSOT), implemented as your Data Warehouse, which takes care of the quality of your data. Making sure there are no repeated rows, incorrect values, or inconsistencies

A Single Source of Truth avoids the common pain point of having different tools reporting different numbers. One single source of truth means metrics are the same across the organization because the same data is used based on the same definitions.

Working with multiple data sources can cause insight and metric inconsistencies due to maintenance issues. Implementing a SSOT allows analysts to have a single source that connects all of these sources and assures the consistency of its data.



Data Governance, Monitoring & Quality

Data is produced, consumed, and used by people. For each dataset, it's important to understand ownership, domain, and flow of data. This is what Data Governance is all about: managing your data and avoiding chaos.

Processes should be clear: Who do I have to ask permission to access a certain asset? If there is no clear process to do this, a person might get unnecessarily blocked for a long period of time.

Data Monitoring adds an extra layer of understanding to your system: Which process is running that incredibly compute-intensive query? Can it be optimized? Else, should it run at a different time? Is it a critical part of the data landscape?

And there is no knowing what we will be able to do with data as technology continues to progress. In our imaginations, the sky's the limit. In the real world, we need privacy regulations and guidelines. Such examples might be the GDPR and CCPA. To face compliance and anti-abuse concerns, teams need to implement fine-grained control of policies regarding data access and storage.

Only some users need to access sensitive information for their day-to-day activities. Most users only need a subset of data. For this reason, assigning specific roles and permissions to different users is key to achieving adequate governance over data.

You can understand the impact your data is having, where it's coming from, and assure its quality. See who is accessing what and limit access to sensitive information.



Low Barrier Business User Focus

Many tools that make up a Modern Data Stack are accessibility-focused. For certain parts of the stack, usually associated with reporting and analytics, little to no code is necessary. The technical know-how required to access data by final business users (marketing, sales, etc.) is reduced, enabling them to easily and routinely consult data to generate insights.

The possibilities for integrating Business Intelligence, marketing, or sales tools are endless.

Furthermore, modern data stacks reduce time-to-market, or in this case, time-to-report drastically. Essentially, this means increasing the value of your data which is only as valuable as the time it takes to become an insight that can be turned into a competitive advantage.

Modern Data Stack enables final business users to easily and routinely consult data to generate insights through low technology barriers, a focus on accessibility, and self-service capabilities. The reduced dependency on data engineers removes bottlenecks and empowers Marketing and Sales teams.

Classdojo saw a 10x increase in Data Warehouse Efficiency.

[Explore the full case here](#)



What Challenges Is Your Company Facing Today?



No Data Stack

“We want our data to work for us, we need to design a working data architecture.”

- The company lacks the manpower and expertise to plan and develop the necessary initial Data Architecture.
- Duplicated data and multiple definitions for the same concept or KPI are causing issues. What appears to be the same data is generating different business metrics because it's being computed from different tables or sources.
- Data teams need to integrate data from multiple sources to generate basic insights, losing valuable time and resources. Common questions are hard to answer, and basic insights hard to come by.
- The team is having a hard time integrating new data sources as well as maintaining existing integrations functioning correctly. This might become visible by an overburdened engineering team drowning in a backlog of queries and fixes.
- There are no quality controls in place on ingested data. When issues arise, errors are hard to identify, monitor and reprocess.

How Can Mutt Data Help You

- Develop a strong tech foundation to scale your organization based on a platform tailored to your particular business goals. Proactively preparing your capabilities to scale avoids headaches in the future.
- Develop a robust Single Source Of Truth with data quality, consistency, and visibility in mind that will become the centerpiece of your Modern Data Platform.
- Develop and implement the essential components of the Modern Data Stack on top of your newly implemented single source of truth: Storage, Ingestion, and Transformation.



What Challenges Is Your Company Facing Today?



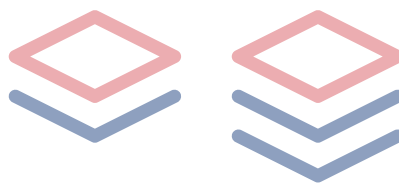
Basic Data Stack

“We have an operating stack but some of the nice-to-have components we’re missing are becoming a need.”

- We have difficulties handling complex or real-time data scenarios.
- We have a difficult time creating dev environments to test changes in processes and pipelines.
- We have issues with performance and maintenance in a scenario of rapid growth.
- We have a hard time achieving proper visibility on what data domains are available and how these can be used.
- We don’t have alerts or notifications on what’s going on. We have issues, failures, and inconsistencies that lead to reactive fixes instead of planned solutions before said errors occur.
- We have no quality controls in place. Users are losing trust in the data sources.
- We have technical debt.

How Can Mutt Data Help You

- Mitigate operational pain points and align the stack to growth best practices by gradually developing the rest of the components of the modern stack, which we’ve previously mentioned: Orchestration, Metadata, Monitoring, Quality, and Governance.
- Evolve your modern data platform using best industry practices and DataOps methodologies to become a self-serve solution that empowers engineers and analysts to make changes efficiently.
- This might not all be necessary at once, but it is possible to gradually add some of these components, tools, frameworks and functionalities as the need arises.



What Challenges Is Your Company Facing Today?



Predictive Stack

“We’re ready to take the next step into the world of Machine Learning and upgrade our stack to enable predictive analytics.”

- We’ve evolved and gradually built on our Modern Data Stack. We’ve achieved a deterministic vision of what’s going on, but what about predicting the future?
- We re interested in forecasting use cases such as operational KPI, sales turnover, and energy demand forecasting as well as anomaly detection, custom product recommendations, discounts, and offerings.
- We need automated models to consider past customer behavioral data to be ahead of changing consumer patterns to adjust marketing efforts accordingly based on rational predictions.

How Can Mutt Data Help You

- Design, development, and implementation of Model Training Workflow: Developing the tooling required to gather assets and build machine learning models from scratch, including the development environment needed to iterate wisely.
- We build operations that ensure the delivery of newly promoted models (promotion is an automated process by which a new model replaces the current production model). This is a necessary step to use different models and metrics tailored to each biz case.
- Best practices and expertise in the efficient development of a centralized feature store. This system enables all data science teams to store, update and serve their features for any machine learning model used across the company.



What Challenges Is Your Company Facing Today?



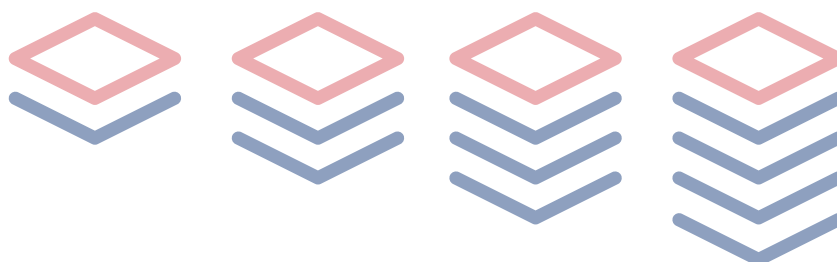
Expert Stack

“We’ve achieved a data-driven, ML-enabled Modern Data Stack for predictive analytics. Now want it to predict behaviours in real time.”

- Many of today’s marketplaces, fintech or delivery applications need to generate recommendations and decisions that react to customer behavioral data in real-time.

How Can Mutt Data Help You

- We have an expert, in-house, multi-disciplinary team of #DataNerds specialized in mathematical optimization, applied machine learning, computer vision, and audio and text processing.
- We have worked on multiple Modern Data Stack implementations that required Real-Time Predictive analytics. Including projects in e-commerce, three-sided delivery app marketplaces, and finance.



Where To Begin: Next Steps

Modern Data Stacks can solve many of your current pain points. Maybe you've even identified which stage your business is at in its data maturity journey. How do you go about implementing an effective and robust Modern Data Stack?



Step 1

Find A Partner

Implementing a Modern Data Stack that will work for your company is challenging. Our first recommendation is to find an expert partner that knows what to look out for. Thoughtfully planning every detail of your Modern Data Stack will save you headaches in the future.

Step 2

Discovery & Planning

This phase is crucial. Take the time to understand your current technology and business context and identify your data tooling and capability maturity. **How?** Ask as many questions as you possibly can!

- **Context:** Where are we today, where do we want to be? Is it feasible considering time and budget? What are the main hurdles? How do we adjust, and how do we bridge the gap? Do we have technical debt?
- **Team:** Who are the main stakeholders? Who will participate in this project? Who owns the data? What skill sets do we need?
- **Data:** What does the current infrastructure/architecture look like? How is data generated? Where does it come from? In what formats? What's its quality? Who has access to it? Who should have access to it?
- **Business:** Who are the final business users? What are their common use cases? What business processes are missing or modified? How can the Modern Data Stack improve operations?

These are just some of the questions that need to be answered to define a clear document and roadmap with S.M.A.R.T. objectives, milestones, requirements, and use cases.

Step 3

Implementation

Once you have a complete understanding of the current business problem, status, infrastructure and needs to close the gap between where the organization is and where it wants to be you design and implement the solution based on everything you've learnt during the discovery. Here are some tips to look out for:

- Create consistent processes across the organization on how to handle data throughout the entirety of the Modern Data Stack.
- Know where to look, consider the following hurdles as you develop your solution: Idempotence, Data Completeness, Atomicity, Schema Evolution, Data Latency, Backfill-ability, Duplicates, Cross System Interfaces, Partitioning, and GDPR Compliance.

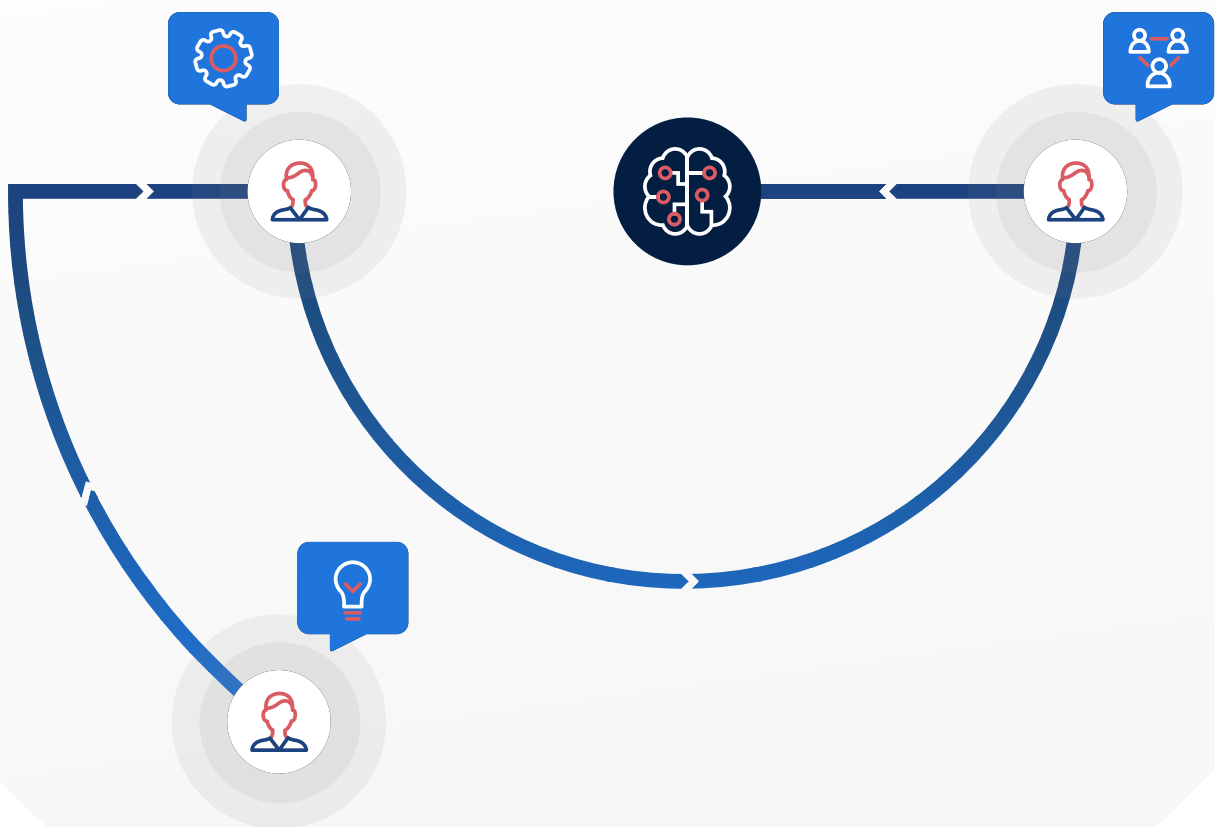
Step 4

Hardening

Understand that Data Stacks are ever-evolving: as your needs continue to change, so will your stack. The Modern Data Stack lifecycle has different phases, which means implementing different monitoring, reporting, and control variables to make sure your system is always up to date:

- Is your stack still fulfilling your business needs? Are there new use cases? Do you need new data? Are insights being generated easily? Is basic data easy to come by across the organization? Is Data still healthy?
- Are we ready for machine learning based capabilities? Do we need predictive analytics?
- Are there processes, rules or issues that can be further optimized? Are we still using the best tools and practices for the job? Are we missing further components regarding governance, monitoring, or quality?

The catch is to set processes in place to constantly monitor the health of your stack and its data to make sure your business doesn't lose its competitive edge.



What Can We Do For Your Stack

Modern Data Stacks, Machine Learning ,and AI implementations can be quite challenging and failure-prone. Companies spend significant amounts of time and money implementing these solutions. Industry statistics show AI implementations have been an issue since 2019, and they're still one of the most common pain points in business today.

We've climbed the Modern Data Stack mountain on many occasions. We are experts at planning, organizing, developing, and nurturing the necessary teams, capabilities, tooling, frameworks, and best practices for the climb.

Years of experience have allowed us to develop a sixth sense. We know where to look, we predict unforeseen consequences and prepare your company for different tracks and needs you will need down the line in years to come.

Most artificial intelligence (AI) projects fail. About 85% never reach deployment.

Gartner



Strategic Consulting

We don't believe in one-size fits all solutions: Every business is unique. Working with Mutt means working with an expert team of in-house data engineers, scientists, mathematicians, and business #DataNerds that take the time to understand your specific needs, goals, and constraints.

Our team will analyze where you stand today, what your options are, and the best way to reach your goals, making the best possible use of your resources.



Accelerated Time To Value

Our experience building and implementing Modern Data Stacks for different companies in different industries have led us to a collection of tried and best practices and baseline structures that allow us to leapfrog your developments shortening time to value.



Knowledge Transfer & Best Practices

We focus on delivering a robust modern data stack with tech capabilities that last. We don't simply deliver a product and leave our clients to their own devices. We transfer our knowledge on best practices and processes, upskilling the client's team so they can sustain the solution in time and adapt or scale the system if necessary.

Reduce Data Processing Pipeline Implementation from weeks to days.

Get Started Today

Is your organization struggling with your current data stack? Are nice-to-have components becoming a necessity? As your business grows, so will the number of sources, incoming volumes of data, and the complexity of its use cases.

Developing and implementing a tailor made modern data platform can be the gamechanger you need to boost your business. Why? The foundation of any data-driven organization is to have a healthy flow of, and easy access to trusted, integrated, and consistent data. The Modern Data Stack enables companies to manage their data from ingestion to visualization saving time, money and effort.

Diagnose the state of your stack:

- Do you trust your data? Are errors hard to identify and reprocess?
- What's the state of your data infrastructure, is it easy to monitor? And how fast can the data pipelines recover from errors?
- Are data engineers over burdened with maintenance issues and business queries and demands?
- Does your business have a centralized self-service source of insights that business users can easily access to make decisions?
- Are machine learning teams sharing a centralized repository of user data or does each project need to generate user features from scratch?
- Are KPIs across teams and departments easy to answer, is the information readily available? Are basic insights easy to come by?

Find out more about what a robust, tailor made Modern Data Stack can do for your specific use cases and goals.

GET IN TOUCH WITH AN EXPERT TODAY

Success Case: Classdojo

ClassDojo is a global community of 50M+ teachers and families who come together to share kids' most important learning moments in school and at home—through photos, videos, messages & more.

Challenge

Classdojo needed a **self-serve Modern Data Platform** to empower engineers and analysts to make changes efficiently. The company faced technical debt, ETL issues, duplicated extraction processes, troubling data ingestion, and a general lack of testability with no monitoring in place in input and output tables.

The goal was to **give every vertically-integrated team ownership over its metrics, tables and data.**

Solution

Mutt Data deployed an expert, hands-on team of High-Performance In-House Data Engineering Muttters' to **design and implement a Modern Data Stack, its tooling, pipelines, components, and integrations** tailor-made to their needs, that could serve the whole company.

[EXPLORE THE FULL CASE](#)

ClassDojo

"Mutt's expertise in building Data Ops platforms, Apache Airflow, data warehouses, and best practices for developing data pipelines led us to a cost-effective, top-of-the-line data architecture."

Dominick Bellizzi
Chief Technology Officer



30%

reduction in data
pipeline processing
and data delivery time

10x

increase in Data
Warehouse efficiency

Success Case: Ank

Ank is an Argentine fintech that developed a product for money transfers between bank and virtual accounts. It allows you to manage several accounts from different banks from a single app and facilitates quick transfers between them.

Challenge

Ank's team needed to fulfil three main objectives: create an automated flow of reconciliation transactions to **reduce manual work** whilst improving the accuracy and trustworthiness of the process. Second, increase the frequency of the reconciliation process to **obtain updated data**. Finally, **implement a batch data processing architecture** based on best practices and technologies that could serve as a model for future pipelines. Together, this needs could be solved by the **implementation of a Modern Data Stack**.

Solution

Mutt Data **developed a system for automatic reconciliation** on an hourly, daily and weekly basis. The system extracts the information from the different sources, centralizes it in a Data Lake in Amazon Web Services, performs the relevant reconciliation checks and, if it detects inconsistencies, corrects them automatically or creates tickets for cases that require manual attention.

[EXPLORE THE FULL CASE](#)



"In Mutt Data we found a partner with extensive experience building data systems from design to production. They worked seamlessly with both our technical and business teams.

Their know-how on building modern data architectures on Amazon Web Services (AWS) was fundamental in developing a reconciliation system which has proven essential to efficiently scaling our operations.."

Gustavo Arjones
Chief Technology Officer



The client saw manual conciliation processes reduced to under 3% of their total monetary transactions.