



How to become a data-driven startup

Unlock the value of your data with an
end-to-end data strategy on AWS



INTRODUCTION

What can today's founders do to enable a data-driven culture?

With the amount of data growing faster than ever, today's successful startups are increasingly tasked with implementing end-to-end data strategy to better inform their business initiatives. Yet, despite widespread agreement on the need to "become more data-driven," most businesses have difficulty succeeding in their data strategy initiatives.

According to a 2021 report by **NewVantage Partners**, 92 percent of leaders identified organizational culture as the main barrier to becoming a data-driven organization.¹ Culture change does not come easy, and it requires investments in the right data infrastructure, solutions, people, and tools.

Above all, achieving the changes necessary to become data-driven—cultural and otherwise—requires an end-to-end data strategy that helps you manage the growing volume of your startup's data, find insights across a variety of data types, and make those insights readily available to the right people and systems.

Furthermore, an end-to-end data strategy puts your data to work, enabling you to extract more value from your data to drive innovation, seize opportunities and, ultimately, transform your business into a data-driven organization.

By building your **data strategy with Amazon Web Services (AWS)**, you get the benefit of the most scalable, highest-performing data services, the broadest and deepest set of integration capabilities, the easiest way to connect all of your data, and end-to-end governance to free your data. AWS provides data services for every use case, so startups of any size can launch immediately.



¹ "Big Data and AI Executive Survey 2021," NewVantage Partners, 2021



In this eBook, we explain how to unlock the value of your data and explore the three pillars to achieve an end-to-end data strategy on AWS:



Comprehensive



Integrated



Governed

- 1 Comprehensive:** Equipped with the right tools, with the optimal price performance for any user, type of data, and use case
- 2 Integrated:** The ability to integrate data that is stored and analyzed in different tools and systems to gain a better understanding of your business and predict what will happen
- 3 Governed:** Governance of all your data to securely give data access when and where your users need it to speed innovation

We then explore how the most innovative startups from a variety of industries have leveraged AWS to achieve benefits such as making more-informed decisions, improving customer experience and loyalty, and building, deploying, and managing fraud detection models without previous machine learning (ML) experience.

What will you learn?

How to make better use of your data and how AWS can help your startup accomplish its goals.

Data sits at the heart of digital transformation

181 ZB

Between 2010–2020, the amount of data produced increased from 2 ZB to around 64 ZB. According to **Statista**, that figure is predicted to reach 181 ZB by 2025.

25%

In a recent survey of B2B companies, just 25 percent of those businesses reported using data on a weekly basis to understand the needs of their customers.

How AWS helps unlock the power of data

Data can be the special ingredient that unleashes new opportunities for your startup. With the power of AWS behind it, your data can help:

- **Transform customer experiences:** Connect your customer data across all points of the customer journey to deliver more optimized, personalized experiences
- **Improve decision making:** With real-time insights at your disposal, you will have the tools needed to make data-driven decisions
- **Future-proof your startup:** Take control of your data to help improve the customer experience and the demands of industry regulators
- **Optimize costs:** Streamline your operations and replace outdated legacy technologies with modern, purpose-built data systems

On the following pages, we will outline the elements of a successful data strategy and how these elements can be applied to your startup. To inspire your data journey, you will explore the success stories of startups that have harnessed the power of data and AWS, providing you with the insight and understanding of how these vastly different startups achieved smarter, faster business outcomes with strategies that maximize the ability to turn information into actionable insights.



The three pillars of an end-to-end data strategy

Your data strategy should help your startup manage, access, analyze, and act on data. What that looks like will vary from one startup to another, but AWS has identified three key pillars that should form the foundation of any end-to-end data strategy.

1 Comprehensive

In today's complex business landscape, agility is everything. For that reason, startups need to scale quickly with flexible tools that accelerate decision making and keep up with rapidly evolving customer demands. AWS offers the most scalable, high-performing service for storing, querying, and activating data. AWS is constantly evolving as a customer data platform (CDP) to enable startups like yours.

For example, the **AWS Auto Scaling** feature continuously monitors how applications are used and adjusts capacity to maintain a steady, predictable performance at the lowest possible cost. It's easy to set up and can scale multiple resources in just a few minutes.



2 Integrated

The most meaningful insights come when you have a full, up-to-the-minute picture of the real-time status of your business—but that’s easier said than done.

It’s only when you connect the dots between your data sources that the full picture emerges. AWS allows you to access all of your startup’s data across ingestion, analytics, ML, business intelligence (BI), and third-party data. All AWS services can be linked together with direct integrations, giving teams access to all of the information needed much more quickly—all of which can help your startup move in a more agile way.

It’s now possible to query data in plain language without writing any code. You can even ask “why” questions to better understand what factors are influencing trends. Using past data, your users can instantly make forecasts for the future.

AWS solutions, such as [Amazon SageMaker](#) and [Amazon QuickSight Q](#), offer no-code interfaces that allow users to generate predictions without prior ML experience. Unleash new possibilities by making insights available to those who need them.

3 Governed

A fast-moving data strategy is only possible with fully reliable data governance. Big data brings a host of opportunities, but it also creates vulnerabilities. All users will need to store and locate data securely and act on information they can trust. AWS provides complete end-to-end governance, allowing your teams to move safely and confidently.

These three pillars provide the foundation for a successful end-to-end strategy. On the following pages, we highlight a number of related startup success stories that demonstrate what this modern data approach looks like in reality.

Optimizing operational data costs to meet customer needs

About Sayurbox

Sayurbox was founded in 2016 with a mission to use technology to improve Indonesia's agricultural supply chain. The startup works with 8,000 farmers across the country to bring fresher, more affordable produce to consumers in Jakarta.

The challenge

Scalability and speed of deployment are key to enabling growth for Sayurbox. Daily orders had recently spiked by nearly 300 percent, and the company expected growth to continue as online grocery shopping became widespread in Indonesia. In 2019, Sayurbox migrated from another cloud service provider to AWS to reduce overhead and take advantage of managed offerings while supporting its data requirements.

[Read the full story ›](#)



The AWS solution

[Amazon Elasticsearch Service](#) helps with monitoring and analytics activities performed in Sayurbox's operations and warehouse management systems to assess demand levels and match them to supply sources. Customers can order on the Sayurbox website and app or even via WhatsApp. The startup also uses [Amazon Relational Database Service](#) (Amazon RDS) to automate inventory management.

The results

With scalable managed services on AWS, Sayurbox sustains lower data infrastructure costs, which is vital for maintaining its business model, while automating inventory management to match demand with supply. Customers can save about 30 percent on supermarket purchases, and farmers earn up to 20 percent more for their harvest.

sayurbox

"With AWS, it's been easier to scale up our resources and track spending as we go."

Nilesh Kumar
VP of Engineering, Sayurbox

AWS services used

[Amazon Elasticsearch Service](#)

[Amazon RDS](#)

[Amazon DocumentDB](#)

Increasing the use of backend services by up to 60%

About OneFootball

OneFootball is the world's most popular digital media platform for soccer enthusiasts, reaching 85 million monthly fans in 15 languages with 24/7 news, live streams, scores, stats, and highlights on more than 200 leagues and competitions worldwide.

The challenge

To make backend data more available to stakeholders, OneFootball made the decision to build a data lake. But to get up and running quickly, the organization needed to create a data extraction system on its own using existing frameworks. All backend data exposed through APIs was extracted through scripts that would comb through data and drop it into **Amazon Redshift** every night. The OneFootball infrastructure team managed to extract, transform, and load (ETL) frameworks individually using different blueprints. This ultimately increased technical debt and the amount of maintenance the team had to manage.

[Read the full story ›](#)



The AWS solution

Using the prefabricated blueprints in **AWS Lake Formation**, OneFootball has been empowered to put the data in the right shape to be consumed by Amazon Redshift with very low overhead. It allows multiple company stakeholders to ingest the information they need on the fly and handpick the tables they want to replicate in the data lake—making data more accessible and giving data engineers more time to innovate.

The results

Since implementing AWS Lake Formation, OneFootball has significantly reduced lead times for loading data from operational databases to the centralized data lake, and coverage of backend services as part of its data lake has grown by 60 percent. This ultimately helped the organization boost weekly active analytics users—a key performance indicator—increasing usage of the analytics platform by 40 percent.



“Building something like this ourselves would have cost us time and caused headaches. If the team needs to ingest new data, instead of creating a complex project, we set up a blueprint and schedule that data to be available daily in the data lake.”

Stephan Durry
Head of Data & Insights, OneFootball

AWS services used

[AWS Lake Formation](#)

[Amazon Redshift](#)

[Amazon Simple Storage Service \(Amazon S3\)](#)

[Amazon Kinesis Data Streams \(Amazon KDS\)](#)

Leverage real-time data to improve customer experience, speed to market, and sales turnovers

About Bizzy Group

Bizzy Group is one of Indonesia's largest B2B ecommerce platforms. By serving the sales and distribution process for fast-moving consumer goods (FMCG), Bizzy connects retailers and FMCG brand owners across Indonesia on its digital platform.

The challenge

The company wanted to provide customers with insights gathered from sales patterns and consumption trends—but its legacy distribution platform could not keep up with the growing demands and fast-paced nature of the business.

The AWS solution

Bizzy engaged AWS and [Source](#), an [AWS Partner Network](#) (APN) Advanced Consulting Partner, to build a data lake and automate data pipelines that would allow Bizzy to

store data at scale and run different types of analytics to generate more timely insights from its data. The data lake centralized 13 data sources on [Amazon S3](#), which are now ingested in real-time. The data can also be used on [Amazon Redshift](#), a data warehouse, to enable in-place querying and analytics. This allows Bizzy to reduce the time to generate sales and inventory reports from three days to one hour.

The results

The platform now collects hourly data on key information such as delivery status, stock status, payments, and sales performance. With a high level of data transparency, Bizzy is able to improve the experience of its retailers and brand owners on its platform. By acting on insights provided by Bizzy's real-time data, brand owners may, for example, experience a higher speed to market and more sales turnovers.

[Read the full story ›](#)



“With the capabilities of our new data infrastructure, our teams have been able to uncover insights that were not easily observed and make data-driven decisions to scale the reach of Bizzy Digital.”

Glenn Lai
COO, Bizzy Digital

AWS services used

[Amazon S3](#)

[Amazon Redshift](#)

[Amazon Athena](#)

[AWS Lake Formation](#)

Boosting sales with real-time, personalized customer recommendations

About Pomelo

Pomelo Fashion is a global fashion ecommerce service that creates personalized, seamless customer experiences to ensure its customers discover new items.

The challenge

The algorithm for displaying Pomelo's fashion items relied on old data streams with limited inputs and spotty accuracy, slowing down this fast-growing startup. Pomelo needed to increase the quality of recommendations on category pages for customers and the relevancy of the products, as well as boost customer engagement and convert sales more efficiently.

The AWS solution

Using **Amazon Personalize**, Pomelo developers can build applications with the same ML technology used by

Amazon.com for real-time, personalized recommendations. It does this by sorting products on category pages based on individual shoppers' preferences, which are used to predict which products they are most likely to find interesting. **Amazon SageMaker** builds additional ML models for forecasting to ensure customers don't see the same recommendations.

The results

Pomelo Fashion significantly boosted sales, with return on investment (ROI) increasing by 400 percent within a month. The personalized ranking algorithm is now being used across nearly all Pomelo categories. This has helped increase gross revenue from category pages by up to 15 percent, click-through rates from category to product pages by up to 18 percent, and add-to-cart clicks from the category page by up to 16 percent and unlocked an 8 percent gain in incremental gross revenue.

[Read the full story ›](#)



Pomelo.

“When you think of e-commerce, you think of AWS. New services are always coming out on AWS, and support is very good.”

Shane Leese

Business Intelligence Director, Pomelo Fashion

AWS services used

[Amazon Personalize](#)

[Amazon SageMaker](#)

[AWS Lambda](#)

Increasing engagement by personalizing customer interactions

About Desygner

This simple-to-use content creation tool empowers anyone to create and share content. Desygner offers multiple graphic design apps to consumers and web-based tools for customers who prefer to design from their computers.

The challenge

After their first use of a product, users lapsed into inactivity. Desygner wanted to increase engagement by personalizing its communications to improve the quality and frequency of email communication with users after their initial sign-up. Off-the-shelf external email marketing tools were cost-prohibitive and had limitations for personalizing email at scale.

The AWS solution

Amazon Pinpoint is a scalable marketing communications service with a feature set including campaign metrics, message personalization, and audience segmentation. Its ability for segmentation enables the creation of unique customer journeys within a marketing automation funnel, and it has a proven high value-to-cost proposition. Amazon S3 stores customer data and follows the AWS Shared Responsibility Model for secure operations in the cloud.

The results

Desygner's user retention rate has more than doubled, from 3.1 percent to 6.6 percent, using the funnel approach, while the email open rate increased significantly, from about 3 percent to 30 percent, and the bounce rate dropped from 6 percent to 1 percent. Personalized content has increased in frequency, with 25 times higher email volumes being generated than before.

[Read the full story ›](#)



DESYGNER

“One thing that drove us to Amazon Pinpoint was segmentation and the ability to create unique customer journeys within a marketing automation funnel.”

Mariana Aguiar
Founder & CPO, Desygner

AWS services used

Amazon Pinpoint

Amazon S3

Maximizing the value of data for better decision making

About Eko

Eko democratizes banking and financial services by helping low-to-moderate-income workers digitize their earnings. Primarily offering financial services and solutions that facilitate payments and money transfers, it has expanded into lending and insurance.

The challenge

Eko's dataset was complex, where one database alone could contain more than 800 tables. Spread across multiple legacy databases and with its data and analytics team experiencing long waits for data retrieval, a new analytics platform was needed.

The AWS solution

Working with Oneture, Eko built a data lake on the AWS Cloud to gain a unified view into diverse data sources. [Amazon EMR](#) is used for big data processing, and [AWS Glue](#) prepares and loads data for analysis. [Amazon Athena](#) is at the core of its analytics pipeline and is used to run serverless queries from data stored in [Amazon S3](#).

The results

Data is retrieved faster than ever, taking just 3.8 seconds to run key reports. This near real-time data provides better visibility into its daily cash flow. Customized dashboards have led to improved decision making and productivity gains from empowered teams. Eko has achieved a comprehensive level of customer segmentation based on the volume and frequency of transactions and greater personalization by leveraging customer data.

[Read the full story ›](#)



“With our platform running on AWS, we can identify the customer segments that consistently deliver high transaction volumes and focus our marketing efforts accordingly.”

Sheekha Verma
Data Science Manager, Eko

AWS services used

[Amazon S3](#)

[Amazon EMR](#)

[Amazon Athena](#)

[AWS Glue](#)

Using data to help customers reduce their energy costs by 30%

About Enerbrain

To meet sustainability goals, Enerbrain devised a digital solution to harness the huge volumes of data generated by building management systems (BMSs). The solution helps create benchmarks to reduce energy use, improve internal comfort levels, and feed into environmental, social, and governance (ESG) reporting.

The challenge

Enerbrain needed to handle and analyze large volumes of data generated by numerous devices, such as environmental sensors and actuators retrofitted at customer sites. It needed to do this without investing in on-premises IT equipment. To accommodate different vertical markets, including schools, airports, and large industrial buildings, it needed to create different models.

[Read the full story ›](#)



The AWS solution

As scale was needed for the initial proof of concept, Enerbrain built its architecture on [Amazon Aurora](#). Fast data analysis was performed using [Amazon Athena](#) and [Amazon S3](#). To help the company train its algorithms to deliver the best energy use strategy for each customer, [AWS Lambda](#) and [Amazon SageMaker](#) ML capabilities were key differentiators for the startup. They allowed Enerbrain to easily learn from customer data and use this to help provide high levels of energy efficiency and optimal comfort levels—down to specific areas within a given building.

The results

Using AWS compute and ML solutions, Enerbrain has helped its customers reduce their energy costs by up to 30 percent and has drastically cut carbon dioxide by 9,600 tons in the buildings it manages in a single year—the equivalent of the annual carbon emissions of 1,894 passenger vehicles. AWS guarantees uptime and availability, allowing Enerbrain to spot any red flags to make needed adjustments.



“It’s essential for us to scale easily so we can quickly manage more buildings. Since developing our platform on AWS, we’ve been easily able to increase that number from 10 to 450 in a relatively short time.”

Marco Martellacci
CTO & Co-Founder, Enerbrain

AWS services used

[Amazon Aurora](#)

[Amazon S3](#)

[Amazon Athena](#)

Detecting fraud in real-time to protect your organization

About Truevo

Truevo makes simple, intuitive, and user-friendly payment products that allow its clients to receive payments effortlessly so they can focus on growing their businesses.

The challenge

The conventional limitations of on-premises or software-as-a-service (SaaS) offerings had curtailed operational success at Truevo. To gain greater control of systems and processes, it was seeking out a flexible solution to respond to bad actors and detect fraud in real time.

The AWS solution

With **Amazon Fraud Detector**, the startup is no longer bound by the conventional limitations of on-premises or SaaS offerings. The solution has equipped Truevo to deploy

rule detections when odd activity is noticed and to respond and adapt to constantly changing regulatory and schematic requirements. AWS ML-enabled services can help the company meet its needs on demand, and Truevo can use the AWS rules-only option while easily scaling to full ML capabilities.

The results

Amazon Fraud Detector has increased flexibility to respond to bad actors and provided greater control over systems and processes. This enables Truevo to operate with greater confidence knowing that the organization can detect fraud in real time. Adapting to an ML-powered service has saved Truevo 3–6 months in development and enabled its team to deploy its first prototype model within 30 minutes.

[Read the full story ›](#)



truevo

“Overall, we are operating with greater confidence in our ability to detect fraud in real time. We are better equipped to deploy rule detections when we notice odd activity that we may not fully understand, but need to stop. We are able to respond and adapt to ever-changing regulatory and scheme requirements allowing us to stay on top of our game.”

Charles Grech
COO, Truevo

AWS services used

[Amazon Fraud Detector](#)

[AWS ML tools](#)

Scaling with a proper data infrastructure

About claimsforce

claimsforce is the future of claims software, creating great claim experiences through the perfect combination of data and user-friendly software. The organization differentiates itself by using data to tackle difficult problems, such as fleet routing or recommendations, helping maximize the experience for its customers.

The challenge

With no dedicated infrastructure in place to process the data and an ever-increasing amount of data stored in application databases and external sources, performing tasks on this distributed data was becoming more and more difficult. A new data infrastructure was needed to accommodate the growing volume of data and access insights to benefit customers.

The AWS solution

Lake House architecture combines both data lakes and data warehouses and can accommodate rapid changes in requirements, leading to changes in available data and its

structure. claimsforce utilizes Amazon S3 and AWS Glue for data storage and migration and the Amazon Athena query engine to provide a single source of truth. Amazon DynamoDB is the main data source, and AWS Lambda gathers the data. Amazon SageMaker is utilized to build a disposition engine to help recommend the best damage adjuster for a specific claim. Amazon Redshift provides valuable insights about the efficient use of its client's resources, while Amazon QuickSight provides dashboards and analyses for its customers and for internal use.

The results

With the seamless scalability of AWS, there are no issues with scaling infrastructure, helping to avoid downtime and maintenance efforts. AWS infrastructure and services enable claimsforce to provide its partners with valuable insights and improve internal efficiency and flexibility. This allows the focus to stay on the customer, as speed, convenience, and reliability are essential in cases of flooding and other natural disasters.

[Read the full story ›](#)



“After just a couple of months of using our Lake House, it’s clear that it’s had a positive impact on our business, allowing us to deliver the best possible customer experience.”

Robert Kossendey
Team Lead Data Chapter, claimsforce

Johannes Kotsch
Data Engineer, claimsforce

Hans Hartmann
Data Engineer, claimsforce

AWS services used

Lake House architecture

Amazon S3

AWS Glue

Amazon Athena

Amazon DynamoDB

AWS Lambda

Amazon SageMaker

Amazon QuickSight

The power of real-time data processing with no downtime

About Choco Up

Choco Up is the largest revenue-based financing investor in Asia. Its financing platform serves digital merchants, providing brands and startups with a funding avenue that was previously unavailable to them. Its quick financing helps modern small businesses grow.

The challenge

To ensure its data pipeline could handle large volumes of data at speed and maintain stability, Choco Up turned to AWS for its big data architecture. Choco Up clients make payments on funding based on a share of their revenue, so it was imperative that the startup could generate reliable 6–12-month revenue projections based on all available data.

The AWS solution

Choco Up utilizes [Amazon Elastic Container Service](#) (Amazon ECS) for container orchestration and [AWS Lambda](#) to

deploy infrastructure as code. It also leverages [Amazon Redshift](#) as a data analytics warehouse for scenario analysis at scale. With access to extensive proprietary data on its database, Choco Up can benchmark each client to evaluate credit risk. [Amazon SageMaker](#) helps its credit risk team perform further validations before a final funding decision is made.

The results

Since launching, Choco Up has seen no downtime on its platform. With the power of near real-time data processing, the organization can make a preliminary offer of funding as soon as a client completes their application, making this a key differentiator compared to financing from a bank. Decisions are made from a pure performance perspective, eliminating any bias, while predictive modeling allows the credit risk team to reduce underwriting risk.

[Read the full story ›](#)



“AWS is renowned for its data and analytics infrastructure, which provides a wide range of choices for building and deploying microservices.”

Lewis Pong
VP of Product, Choco Up

AWS services used

[Amazon ECS](#)

[AWS Lambda](#)

[Amazon Redshift](#)

[Amazon Sagemaker](#)

[AWS Activate](#)

CONCLUSION

Next steps

Suddenly, data is exciting. Because the moment you discover how an end-to-end data strategy can help your teams deliver successful data-driven outcomes, you will want to begin doing just that. Established on an infrastructure foundation that is comprehensive, integrated, and governed, there are no limits to what your startup can do.

The flexibility and scalability of AWS services provide a more agile way to scale with evolving demands, eliminating drag and giving you the foundation for an effective data strategy that not only works today but can also power your startup well into the future.

As the AWS startup customers featured in this report have demonstrated, data has enormous power to deliver positive customer experiences, optimize processes, create products, overcome unique challenges, and open revenue streams. Its comprehensive suite of services applies to all workloads, all types of data, and all outcomes.

With our professional services and partners, AWS can help you create a culture that is driven by data—one in which you can uncover insights through real-time intelligence to make more accurate predictions and smarter decisions that deliver a consistent competitive advantage across your startup.



Gartner's 2021 Magic Quadrant for Cloud Infrastructure and Platform Services identified AWS as the highest-placed leader across both "ability to execute" and "completeness of vision."

[Get started with data solutions from AWS ›](#)