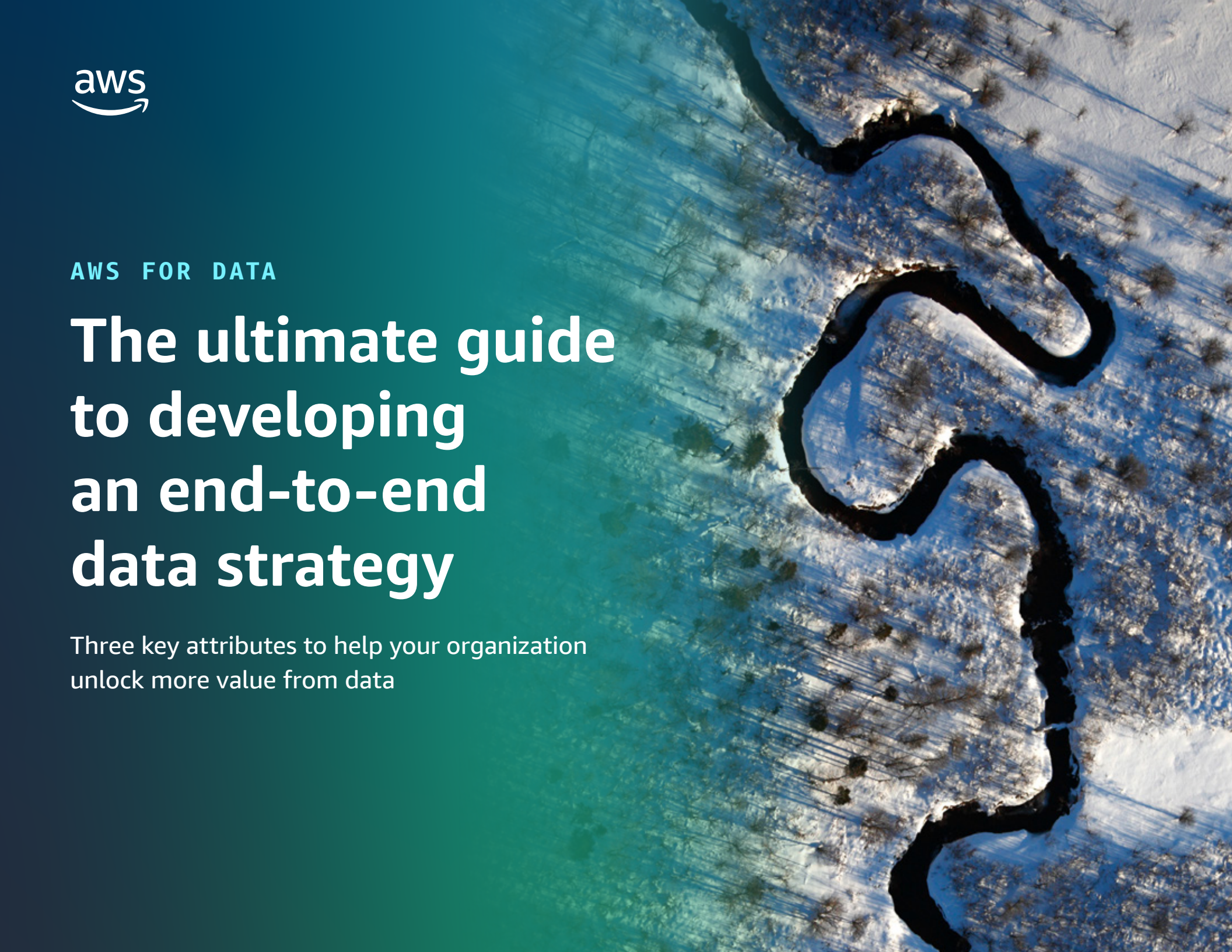


AWS FOR DATA

The ultimate guide to developing an end-to-end data strategy

Three key attributes to help your organization unlock more value from data

Table of content

INTRODUCTION

Data is the spark that leads to meaningful innovation

Now more than ever, data is at the center of every application, process, and business decision. It's the **genesis for modern invention**, and in today's fast-changing and complicated landscape, how you put your organization's data to work can be the golden ticket to accelerating innovation, and accomplishing your organizational goals. The stakes are high. According to Forrester Research, organizations that have a system to promote data-driven insights are 140 percent more likely to create sustainable competitive advantage and 78 percent more likely to fuel a revenue growth environment.[1]

With a pressing need to empower the entire organization to use data to make better, faster decisions that fuel new ideas and drive business agility, leaders are embracing a fundamental truth: The journey to innovation begins with data, and successfully becoming a **data-driven organization** begins by implementing an end-to-end data strategy.

Is there a proven data strategy? The answer is "Yes"

The good news is that you don't have to reinvent the wheel. Organizations are already capturing the benefits of an end-to-end data strategy built on Amazon Web Services (AWS). For example, **AstraZeneca** is integrating and scaling its data and artificial intelligence (AI) capabilities across the business to innovate faster to improve patient outcomes. With AWS Data services, they can now run more than 51 billion statistical tests in under 30 hours, facilitating delivery of genomic insights to drug discovery projects. **BMW Group** and many more manufacturers use data to optimize their supply chains and improve production capacity. And **ENGIE**, along with others in the energy sector, uses data to find new ways to reduce costs for its customers while meeting bold and aspirational sustainability goals.

While the achievements are limitless, the central challenge is this: Many organizations are sitting on a treasure trove of data but don't know how to gain value from it. In this eBook, you will learn the fundamentals of building an end-to-end data strategy to keep up with your data needs now and in the future—enabling a sustainable advantage that comes from unlocking the value of your data.

[1] "Creating a data-driven culture," CIO.com, March 31, 2022

Key challenges and considerations

More data than ever is being generated and stored

On-premises tools and legacy data stores can't meet today's demands, organizations need new data stores that can scale and grow as business needs change—whether from the gigabytes and terabytes handled today or the petabytes and exabytes that will be managed in the future.

Data siloed across multiple sources creates productivity and cost inefficiencies

Modern organizations need to easily access and analyze diverse types of data, including log files, clickstreams, voice, and video. However, these wide-ranging data types are typically stored in silos across multiple data stores. To extract intelligence, organizations must break down these silos to unify all types of data. This important optimization of costs and operations is transforming the infrastructure from a source of complexity and expense to an engine of value creation.

The current state of decision making is unsustainable

Gartner reports that 65 percent of decisions made today are more complex (involving more stakeholders or choices) than they were five years ago.[2] To make better and faster decisions, organizations need the ability to perform analytics and machine learning (ML) operations in an agile, cost-effective way—using optimal tools and performance to scale for each use case. Organizations can no longer waste precious time constantly redeploying and reconfiguring infrastructure to scale performance and capacity.

Analytics and machine learning adoption is still impeded by a lack of skills and inertia

Many businesses are struggling to make progress with scaling analytics and ML tools. Gartner finds that organizations investing in AI moved just 54 percent of their AI proof-of-concept pilots into production.[3] A continued lack of data and ML skills and quantity or quality of data to train on are just some of the issues slowing progress in this important area. Still, the need to help business users leverage data-driven decision making is growing.

Trying to maintain data governance is a full-time job

Traditional data architectures require risky, complicated management procedures because data is accessed from so many places. Granting, tracking, auditing, and removing employee access—while simultaneously remaining in compliance with a growing number of regulations—is a full-time job. Automating these mandatory data governance tasks frees modern teams to shift their focus back to innovation.

Data is increasingly difficult to secure

There was a time when IT teams chose between making their architectures fast or making them secure. Now, they need to deliver both. Meanwhile, security attacks increased by 31 percent from 2020 to 2021, according to Accenture's State of Cybersecurity Resilience 2021 report, while average attacks per organization increased from 206 to 270 year over year.[4] How can organizations maximize privacy and security?

[2] "How to make better business decisions," Gartner, October 20, 2021
[3] "Half of AI models never make it to production," EnterpriseAI, August 23, 2022
[4] "How aligning security and the business creates cyber resilence," Accenture, 2021

BECOMING DATA-DRIVEN

Three key attributes that can help your organization unlock more value from data

According to a PwC survey of more than a thousand senior executives, highly data-driven organizations are three times more likely to report significant improvements in decision-making compared to those that rely less on data.[5]

AWS can help your organization implement an end-to-end strategy that makes data management easier at every step of the journey—from ingesting, storing, and querying data to analyzing, visualizing, and running ML models. Regardless of your business challenges, your data strategy should be:

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

A data-driven mindset may also require a broader cultural change in which both goals and decisions are supported by the data strategy.

Follow the link below to explore why data plays a vital role in enabling this cultural change. And learn why a growing number of companies are leveraging data-driven capabilities to automate a set of business-critical use cases.

Read *5 Key Elements of a Data Driven Culture* eBook ›

Comprehensive

Integrated

Governed

[5] "Data and Intuition: Good Decisions Need Both," Harvard Business Publishing, January 6, 2023

1

Comprehensive

Equipped with the right tools, with optimal price performance for any user, use case, and data type

Businesses need to build future-proof data strategies that can meet their needs now and in the future. It takes more than just a single data lake, data warehouse, or business intelligence (BI) tools to harness data effectively. It requires an end-to-end data strategy with a comprehensive set of tools that accounts for the scale and variety of data and the many purposes for which you want to use it. In fact, 94 percent of the top one thousand AWS customers use more than 10 AWS Database and Analytics services.

Building with a cloud provider that innovates to continuously bring you all the data tools you'll need and more with the right price performance for your use case ensures that you have a data strategy that grows with you. AWS has the broadest and deepest set of data capabilities to support any data workload or use case. From databases for applications to storage for data lakes to analytics to ML and end-user tools, AWS provides the right capability in each area, so you don't have to compromise on performance, cost, or results. AWS is continually accelerating its pace of innovation, so you will never outgrow AWS for your data needs.

COMPREHENSIVE

Scaling data-driven applications

Build future-proof applications on a modern data infrastructure for the best price and performance for your use case at scale. AWS databases include Amazon Aurora, which provides the performance and availability of commercial-grade databases at one-tenth the cost. More than a hundred thousand AWS customers use Amazon Aurora for their relational databases. AWS offers eight purpose-built database engines for use cases such as graphs, streaming, and documents, each uniquely designed to provide optimal performance for your applications, transforming the economics of database ownership.

Powering data and data warehouses

Organizations use data lakes and data warehouses to unify datasets and apply analytics and ML to aid in decision making and create new data-driven customer experiences.

AWS has been helping hundreds of thousands of customers build a strong foundation for data lakes with services such as Amazon Simple Storage Service (Amazon S3), AWS Glue, and AWS Lake Formation for years. Customers also rely on Amazon Redshift, a fast, petabyte-scale data warehouse delivering up to five times better price performance than other cloud data warehouses. Amazon Redshift helps you securely ingest, combine, and run historical, real-time, and predictive analytics on all of your data in just a few clicks, with super-fast query results.

Providing analytics for all use cases

True agility helps organizations adapt quickly to changing business needs. To power these rapid actions, AWS analytics services enable your organization's teams to ingest, combine, and run historical, real-time, and predictive analytics on all of your data. This includes services for SQL querying, log analytics, streaming, and Apache Spark. For big-data querying, Amazon EMR supports more big-data frameworks than any other provider and gets you up to two times faster time-to-insights. AWS customers achieve more than three times performance with Apache Spark when they run supported and AWS-optimized runtimes for Amazon EMR, AWS Glue, and Amazon SageMaker. To make decisions in real-time, you'll need streaming data services such as Amazon Kinesis Data Streams (Amazon KDS), which allow you to build applications for high-frequency event data such

Data strategy tip

Achieve a cost-effective data strategy without sacrificing performance. Enable your organization to maximize its current capabilities by optimizing cost:

Samsung

Samsung saved 44 percent on monthly operational costs, and an additional 22 percent on maintenance fees when migrating to Amazon Aurora PostgreSQL.

Carrier

Carrier connected its cold chain logistics network to help its customers optimize cold chain operations, decrease their energy use, and enhance their outcomes with a reduction in costs, delays, cargo loss, and spoilage in transit.

United

United Airlines created an intelligent airport with more than 20,000 sensors producing data to drive real-time insights, optimizing ground equipment capabilities, resulting in $120 million in savings for equipment that was no longer required.

as clickstream data, and gain access to insights in seconds. Amazon Kinesis Data Firehose simply and reliably loads data streams into data lakes, warehouses, and analytics services—no extract, transform, and load (ETL) or cumbersome data preparation required.

Deploying data science and machine learning

ML adds intelligence to existing processes, automates time-intensive manual tasks, and accelerates innovation with the creation of new products and services. With AWS, you have access to the most comprehensive set of AI and ML services. With Amazon SageMaker, used by tens of thousands of customers, you can build, train, and deploy ML models for any use case with fully managed infrastructure, tools, and workflows.

Enabling data insights throughout the organization

It's no longer just data-savvy individuals who can rapidly extract valuable, relevant insights from data to help inform decision making. ML-powered BI solutions such as Amazon QuickSight, enable easy connectivity to data sources. Business analysts can utilize this data to showcase fresh trends and predictive insights on interactive BI visualizations and dashboards.

QuickSight Q uses ML, allowing users to query their data in plain language without writing a single line of code. Business users can even ask "why" questions to better understand factors that are impacting data trends. They can also forecast metrics by stating something such as, "Forecast sales for the next 12 months" to receive an immediate response based on the insights of past data and seasonality. A visual point-and-click interface enables business analysts to generate accurate ML predictions without prior experience. In just a few clicks, analysts can import data from various sources, automatically prepare data, and build and analyze ML models.

Boosting data proficiency

Having employees who can use data effectively will help your organization achieve its data objectives. Invest in educating and upskilling your workforce in data, analytics, and ML with AWS Training.

aws

94%

With the industry's broadest and deepest set of data services, 94 percent of the top 1,000 AWS customers each use 10+ AWS Database and Analytics services.

Scale data-driven decision making throughout your organization

- **Amazon QuickSight:** Meet varying analytic needs from the same source of truth through modern interactive dashboards, paginated reports, embedded analytics and natural language queries
- **Amazon SageMaker Canvas:** AWS no-code interface that enables business analysts to generate accurate ML predictions without prior experience
- **Amazon DataZone:** Simplifies governed access to data for business users
- **AWS Training and Certification:** More than 150 professional development courses related to data, analytics, and ML

CUSTOMER STORY

ADP makes 312 trillion decisions a month with analytics processes

ADP helps more than 900,000 businesses manage the 70 million people through its people and payroll process. That management generates a massive amount of data. In fact, ADP processes more than 2.5 petabytes of data with more than 25 billion individual data points represented. ADP uses Amazon Redshift and Amazon Neptune to perform aspects of its overall data processing. These AWS Data services help companies measure, compare, predict, and apply insights about their workforces. Using AWS services, ADP also enables organizations to create Pay Equity Dashboards, helping more than two-thirds of companies show improvement in pay equity.

"Now is the time to use data to help people to understand what actions we can take to create a more diverse, more equitable, and a more inclusive work environment and to build the future we all want to create."

Jack Berkowitz, CDO, ADP

Read more about ADP's experience ›

CUSTOMER STORY

BMW Group democratizes data usage at scale

BMW Group moved to an AWS-based centralized data lake for its agility, flexibility and its ability to process terabytes of telemetry data from millions of vehicles daily. Anonymized data from vehicle sensors and other sources across the enterprise is now easily accessible for internal teams who create customer-facing and internal applications. Building up a human-readable data catalog and clearly displaying data resources proved essential, boosting the productivity of data analysts, data scientists, and engineers.

“We are just starting our journey with AWS, and we look forward to helping our business fulfill its strategy of driving innovation into the future.”

Kai Demtröder, VP of Data Transformation, Artificial Intelligence, Data, & DevOps Platforms, BMW Group

Learn how leading organizations are capturing the benefits of an end-to-end strategy built on AWS ›

2

Integrated

Break down silos so data can be put to work effectively

Opportunities to transform your business with data exist all along the value chain. But making such a transformation requires you to see the full picture of your customer and business. With data spread across multiple departments, services, on-premises databases, and third-party applications, you need to be able to easily integrate data across silos to get the best insights. Companies have various approaches to how they are unifying data—data mesh, lake house, data fabric, and so on—but typically, it involves a data lake as a foundational element. Data lakes allow you to collect, store, organize, and process valuable data from your data silos and make it available to analytics, visualization, and ML tools in a governed way.

INTEGRATED

Zero-ETL

Many organizations have multiple data lakes in addition to data warehouses, analytics tools, ML tools, and SaaS applications. Integrating data across silos requires complex ETL pipelines, which can take hours, if not days. That's just not fast enough for modern decision making. Organizations should adopt technologies that automate or eliminate ETL where possible.

AWS is investing in a Zero-ETL future, allowing organizations to automatically integrate all of their data. This includes bringing ML to the data source with SageMaker integration into Amazon Redshift, Amazon Aurora, Amazon Athena, and Amazon Neptune, integrating Amazon Aurora and Amazon Redshift for real-time analytics; and providing a direct integration between Amazon S3 and Amazon Redshift for real-time data streams. In addition, you can run queries across data stored in operational databases, data warehouses, and data lakes to provide insights across multiple data sources with no data movement using Amazon Athena and Amazon Redshift.

Analyzing all of your data and third-party data

To break down data silos, you can't have connections to only some of your data sources—you need to be able to seamlessly connect to all of them, whether they live in AWS or in external third-party applications, on premises, or even in another cloud environment. No matter where they live, with AWS, you can automatically integrate hundreds of data sources across AWS and third parties.

Increasingly, organizations are also harnessing third-party data to deepen insights by joining this third-party data with their own data. AWS Data Exchange enables AWS customers to access third-party data through files, tables, and APIs from more than 300 data providers and more than 3,500 data products, all from one place. Third-party data from partners and customers are also being used, which increases the need for comprehensive governance policies to protect the data. Data clean rooms—protected environments where multiple parties can analyze combined data without ever exposing the raw datasets—have emerged as a solution. AWS Clean Rooms helps companies and their business partners securely analyze and collaborate on their datasets—without sharing or revealing the underlying data.

Connect with hundreds of data sources

- **Amazon AppFlow:** Integrate data lakes and data warehouses with 50+ sources of data
- **AWS Data Exchange:** Access 350+ third-party providers and 3,500+ public data products
- **Amazon SageMaker Data Wrangler:** Build ML models with 40+ data sources with a single click

Data strategy tip

Pursue Zero-ETL

Avoid building and maintaining complex data pipelines to perform ETL operations. Instead, Amazon Aurora now supports zero-ETL integration with Amazon Redshift, to enable near real-time analytics and ML. Transactional data written into Amazon Aurora is available in Amazon Redshift almost immediately. Analyze data from multiple Aurora database clusters in the same new or existing Amazon Redshift instance and derive holistic insights across many applications.

Goldman Sachs

aws

CUSTOMER STORY

How data delivery enables Goldman Sachs to work smarter

Goldman Sachs has seen demand for financial market data increase exponentially, typically doubling year over year. AWS Data Exchange is a managed service that allows data processes to be centralized, streamlined, and automated wherever possible. Migrating existing market data feeds to AWS Data Exchange can improve the overall efficiency of consuming third-party data, enabling more time to be spent on the value-add analytics of the data and less on wrangling the data to get it ready for use. The frictionless data delivery through AWS Data Exchange is critical to future-proofing the organization.

"AWS Data Exchange is a key component of Goldman Sachs's financial cloud strategy because it reduces friction for sourcing financial data from new and existing third-party providers, and allows us to focus on delivering our core services and differentiated data analytics to better serve our clients."

Marco Argenti, Co-CIO, Goldman Sachs

CUSTOMER STORY

ENGIE accelerates its zero-carbon transition

A global utility company in the process of a zero-carbon transition, **ENGIE built its Common Data Hub data lake on AWS**. ENGIE was supported by AWS Professional Services in the design and implementation of the solution and the formation of an internal service team to oversee the platform. With more than a thousand projects worldwide currently on the Common Data Hub, this cohesive solution eliminates silos, empowering every department with equal access to a common data framework.

"We were convinced that AWS was a good solution for many reasons, including the cost model—and especially in terms of data storage."

Gregory Wolowiec, Technology Team Leader at ENGIE Data Programs

Explore how AWS customers integrated their data, breaking down data silos, and creating transformative solutions for their organizations ›

3

Governed

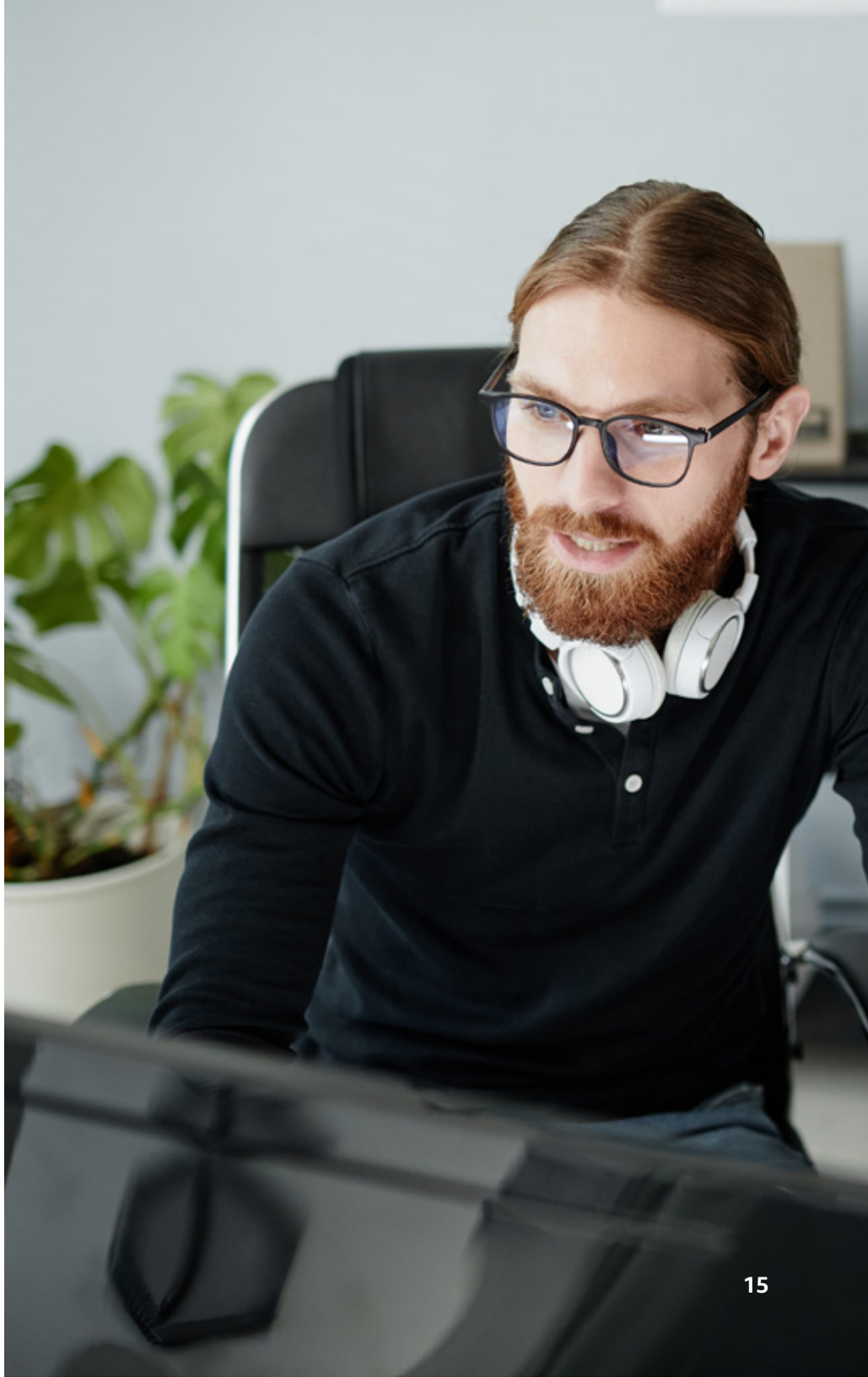
Free your teams to move faster with governed data access across the data lifecycle

Beyond being comprehensive and integrated, it's equally important to ensure that your users can access data where and when it is needed with the right level of control. With the right data governance strategy in place, you can move faster to empower users with the data access they need—when they need it.

As more data migrates to the cloud, driven by the cloud's near-infinite scale and horsepower, it's imperative that enterprise data governance models evolve in lockstep. IT and business leaders need up-to-date policies to protect data as it moves back and forth among different repositories and to accommodate changing privacy and data security regulations about where data can be stored.[5]

"The key to good governance is figuring out how to define access, then getting out of the way. By that I mean creating exception processes, rather than taking an approach that anytime you need data, you have to ask someone in a central organization for it."[6]

Rahul Pathak, VP for Analytics, AWS



[6] "A unified approach to data governance," CIO Magazine, August 2021

GOVERNED

Simplifying data access permissions

Implementing a successful governance strategy continues to present a unique set of challenges. It's time-consuming and challenging for organizations to provide internal or external consumers with their data with the right level of access to specific datasets. They often engage in heavy lifting, such as manual scripts or investigating individual data clusters, to figure out which consumers have access to what data.

Manual work can also lead to costly data quality issues across different teams and departments. Without centralized governance tools, data gets locked down in siloes, which means you won't be able to access and analyze all the data you may need to solve problems or identify large areas of opportunity.

Developing a data governance strategy

A new AWS/MIT survey of more than 350 data professionals shows that data governance is the top priority of chief data officers (CDOs), with more than 50 percent of CDOs noting "establishing clear and effective data governance" as their leading responsibility. Governance is also an area CDOs spend much of their time on, as more than 66 percent of survey respondents said data governance initiatives are a top focus.[7]

Without a governance approach that supports innovation, organizations will find it hard to be data-driven and, ultimately, to remain competitive. After all, the more time workers spend grappling with data, the less time they spend innovating with it.

AWS is investing across the data journey to enable end-to-end data governance with less effort. AWS Lake Formation makes it easy to govern and audit the actions taken with data in your data lake on AWS S3 and AWS Lake Formation can also be used to govern data sharing in Amazon Redshift. Amazon DataZone is a new data management service to catalog, discover, share, and govern data so that everyone in the organization can act on data. And for your ML models, Amazon SageMaker has features to help you govern and audit the end-to-end ML development cycle.

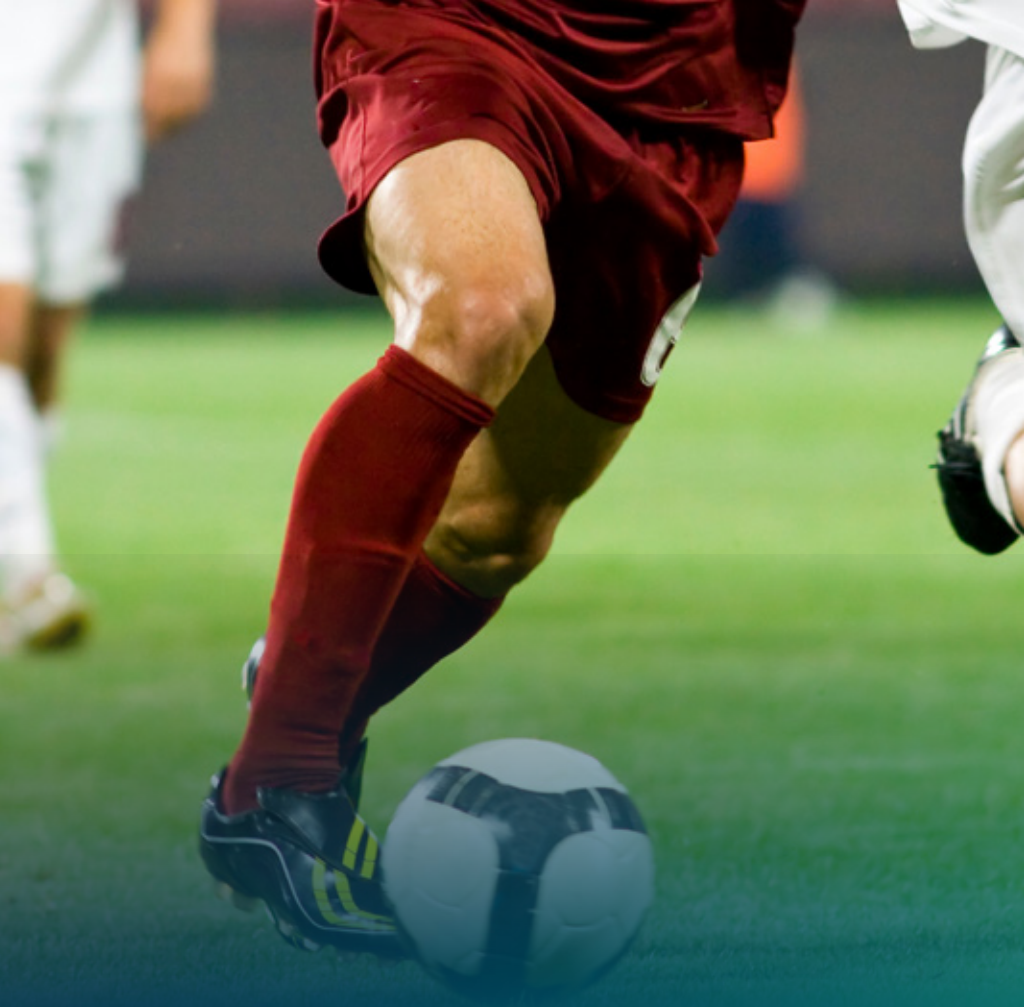
Govern holistically with AWS

- **AWS Lake Formation:** Makes it easy to govern and audit the actions taken with data in your data lake on Amazon S3
- **Amazon DataZone:** A data management service to catalog, discover, share, and govern data

Data strategy tip

A **Splunk report** on the economic benefits of mature data practices reinforces the importance of access. By focusing on things such as data classification and data sharing across business areas, "data innovation leaders" enjoy greater brand loyalty and higher customer satisfaction and operationalize more of their data. Data leaders are "4.6 times as likely to drive more than 20 percent of their revenue from new, innovative products and services" as data beginners.[8]

[7] "The Best Offense Is a Great Defense," TechCrunch Brand Studio, by AWS, 2022
[8] "The Economic impact of data innovation," Splunk.com, 2022



ONEFOOTBALL

aws

CUSTOMER STORY

By simplifying governance, OneFootball saw a 40% increase in the utilization of its analytics platform

OneFootball has grown rapidly to become one of the world's most popular digital media platforms for soccer ("fútbol") enthusiasts. To better use data for the benefit of the company and 70 million fans of "the beautiful game," OneFootball built a nimbler solution on AWS in just a few days. Since integrating data from its inefficient backend databases into its cloud-based data lake, OneFootball has radically simplified data ingestion and eliminated legacy ETL workloads altogether. Beautiful game, indeed.

With AWS Lake Formation, OneFootball could simplify security management and governance at scale, so they could make data more accessible across the company. This ultimately helped the team see a substantial growth in weekly active analytics users and other internal key performance indicators, increasing usage of the analytics platform by 40 percent.

"Now we can spend more time understanding business problems rather than maintaining different types of database extractions. Seeing more and more people across the organization make use of analytics on a daily basis is a great achievement. Having all relevant data sources reliably integrated was a prerequisite."

Stephan Durry, Head of Data & Insights, OneFootball

Pinterest

aws

CUSTOMER STORY

Pinterest puts customers first with governance

A scalable, automated fine-grained access control (FGAC) system built using Amazon S3 ensured Pinterest's growing data wouldn't outgrow the company's existing controls. FGAC controls the access to data and is based on multiple criteria offering options such as role-based access control plus security for petabyte-scale datasets. It also enabled creators and businesses on the platform to self-identify as a member of an underrepresented group while ensuring that sensitive data wouldn't be used for any other purpose such as advertising.

"Customer-facing impacts of Pinterest's governance efforts include using self-identifying data in a "very controlled way" to support Black-owned businesses for Juneteenth. Creators can also add badges to their profiles—which allows creator content to appear in themed spaces on Pinterest—to show that businesses are owned by someone who identifies with an underrepresented group."

David Chaiken, Chief Architect, Pinterest

Learn how AWS customers have empowered their employees with the right governance strategies ›

Making security more strategic

AWS has prioritized security since day one—with continuously protected, high-performing, resilient, and efficient infrastructure for your workloads and applications. World-class security experts who monitor the AWS infrastructure also build and maintain a broad selection of innovative security services—which can help simplify the complexities of your own security and regulatory requirements.

AWS Security services and solutions can enable a mix of important advantages:

- **Getting to insights faster** – Provide the right level of access to your resources at all times while maintaining confidence that your data is protected. AWS Security is built with performance in mind, so you get maximum protection and data governance that doesn't slow you down.
- **Reducing downtime** – Tougher, more modern cloud security helps keep your enterprise moving, so you don't have to stop analyzing data to perform a discrete security process—it can be integrated into every step along the way.
- **Staying within your budget** – AWS keeps security cost-effective and scales with the evolving needs of your security risks and requirements, protecting your organization's investments and its commitment to data initiatives.
- **Keeping your focus** – From infrastructure to services, AWS is secure by taking security into account at every step along the way, so you can spend more time transforming data into better decisions that drive business results and less time worrying about security and governance.

A history of unmatched reliability and security

Amazon S3

Store and retrieve any amount of data with the best security

AWS Lake Formation

Build a secure data lake in days with fine-grained access control

Multi-AZ Regions

Ensure seamless failovers if an Availability Zone is disrupted

CONCLUSION

The next wave of innovation will be driven by data

Leaders and other decision makers looking to join the next wave of reinvention need to be tenacious about getting to the truth. They also need the essential tools to stay agile enough to pivot when needed to act on new opportunities. Simply stated, that means becoming data-driven.

Organizations that are data-driven seek the truth by treating data not as the sole property of siloed departments but also as an organizational asset for all to use. Realizing a modern data strategy for your organization is possible, no matter its size, location, or business needs. AWS provides the most comprehensive set of services over the entire end-to-end data journey for any workload, type of data, and desired outcome.

Learn more about why AWS is the best place to unlock value from your data and turn real-time insights into meaningful innovation. And explore how we can help your teams with infrastructure, tooling, and implementation support via the world's leading professional services and partner network. When it comes to data, AWS customers know how to do it better.

Discover the top data use cases to maximize business value

In *Maximizing Business Value with Data: 6 Data-Driven Use Cases for Business Leaders* eBook explore AWS customer use cases and learn how you can take advantage of data to improve customer experiences, optimize and reinvent supply chains, improve decision making, future-proof applications, and more.

Learn more about reinventing your organization to be data-driven ›