

AWS FOR DATA

The ultimate guide to data cost optimization

Discover how to cut costs and get the most value from your data with AWS



Table of contents

Introduction			
Αр	Approach #1: Build a cost-effective data strategy 4		
	Modernize your data infrastructure with the cloud	5	
	Automate costly and time-consuming processes	. 11	
	Leverage serverless architecture	. 14	
	Standardize your machine learning	. 15	
	Empower business users to self-serve their data needs	. 17	
-	Approach #2: Use insights and intelligence to drive cost reductions across your business		
	Streamline operations	. 19	
	Unify customer profiles to drive marketing efficiency	. 20	
	Improve manufacturing quality and efficiency	. 21	
	Proactively detect and prevent online fraud	. 22	
	Use insights to find savings opportunities	. 23	
Со	Conclusion24		



INTRODUCTION

Cutting costs without stifling innovation

Around the world, organizations of all sizes are concerned about their growing volume of data and the costs associated with managing it. In a <u>recent survey from Deloitte</u>, 45 percent of technology leaders said collecting and protecting ever-growing volumes of data was their number one concern. Storing, processing, and analyzing large amounts of data requires significant resources, including hardware, software, and personnel. At the same time, data has immense value for an organization. It can be used to innovate, uncover new cost-efficiencies, and help organizations remain competitive.

While there is no one-size-fits-all approach to optimizing your data costs, we believe that the best approach is twofold—first, build a cost-effective end-to-end data strategy in the cloud and second, use insights and predictions derived from your data to find cost optimization opportunities in all areas of your business.

In this eBook, we explore the different ways in which organizations can reduce the costs of putting their data to work while also using data-driven insights and predictions to drive innovation, growth, and cost-efficiency. Along the way, we will also share examples of how customers have done this with Amazon Web Services (AWS).





<u>Samsung</u> lowered its monthly costs by 44 percent and applied these savings to innovative, customer-facing projects by migrating its legacy on-premises database to <u>Amazon Aurora</u>.



<u>United Airlines</u> saved over \$120 million, cut emissions by 5–7 percent, and reduced personnel hours by 1.3 million using AWS analytics and Internet of Things (IoT) to help track equipment location, routing, use, and availability.



1 Build a cost-effective data strategy

An end-to-end data strategy is essential for organizations that want to make the most of their data assets and use them effectively to drive business growth and success. This end-to-end data strategy helps ensure that data is managed effectively throughout its lifecycle, from collection to analysis to decision making. To build a cost-effective data strategy, it's important to carefully plan and consider multiple factors.

The first step is to assess the current data infrastructure and identify areas of potential cost reduction. This could mean consolidating data sources, optimizing data storage, and prioritizing solutions that offer the best price

performance. It could also mean leveraging open-source software or cloudbased solutions to reduce hardware and software costs or implementing effective data governance policies to ensure that data is managed efficiently.

By taking a comprehensive and strategic approach to data management, you can build a cost-effective data strategy that delivers measurable value and supports business growth.

We have found several areas where customers can optimize data costs. Let's unpack those.



Modernize your data infrastructure with the cloud

Moving to the cloud can reduce data infrastructure costs by removing the need to manage data centers. Customers can immediately access IT resources, such as compute and storage, without upfront costs or operational overhead, and pay only for the resources they consume. And managed data services in the cloud remove operational heavy lifting. As a result, you can focus on projects that differentiate your business rather than spend time and money on managing servers and IT infrastructure.

Carrier saw a

45%-50%

reduction in the cost of infrastructure using AWS¹



Carrier Global, a Fortune 500 company, has realized a 45–50 percent reduction in the cost of infrastructure using AWS, including a 20 percent reduction in total infrastructure overall—a direct cost savings worth millions.



Data lakes

Organizations today gather both structured and unstructured data from a variety of sources, online customer transactions, sensors on IoT devices, and industrial inventory systems, to name a few. This creates a sprawling collection of data spread across multiple services and on-premises systems. Organizations can bring this disparate data into a data lake to use it for analytics and machine learning (ML).

AWS has helped hundreds of thousands of customers build a strong and cost-effective foundation for data lakes—a centralized repository that allows you to store all of your structured and unstructured data at any scale—with services such as **Amazon Simple Storage**Service (Amazon S3), AWS Glue, and AWS Lake Formation for years. Amazon S3 offers multiple ways to optimize the costs of storing growing amounts of data. Amazon S3 Intelligent-Tiering automates data lifecycle management, moving data to the most cost-effective access tier when access patterns change. S3 Intelligent-Tiering has saved customers \$1 billion in storage costs since 2018 compared to Amazon S3 Standard. And the Amazon S3 Glacier storage classes are purpose-built for data archiving and provide customers with the highest performance, most retrieval flexibility, and lowest-cost object storage in the cloud—at about \$1 a terabyte per month.

For reporting and analysis of structured data, <u>Amazon Redshift</u>, a fast, petabyte-scale data warehouse delivers up to five times better price performance than other cloud data warehouses.

\$1 billion

The amount Amazon S3 Intelligent-Tiering has saved customers in storage costs since 2018*

*Compared to Amazon S3 Standard





ENGIE, one of the largest utility providers in France and a global player in the zero-carbon energy transition—built the Common Data Hub (CDH) on AWS to make data more ingestible, organizable, governable, shareable, and actionable across its 25 decentralized global business units for 160,000 employees in 70 countries.

The CDH is built with an Amazon S3 data lake and uses a variety of Amazon analytics and ML solutions, including Amazon Redshift, to query semi-structured data. ENGIE's CDH made it possible to address the challenges of predictive maintenance, a key use case, by using ML models to prevent equipment malfunction and better plan interventions. The CDH provided the necessary data to train the ML models, resulting in the development and training of more than 1,000 prediction models for a variety of equipment, such as valves, pumps, ventilation, air conditioning, and heating systems.

ENGIE predicts that around 10,000 pieces of equipment will be connected and benefit from predictive maintenance in the next few years, with potential savings of €800,000 per year across the business.

Database for applications

Databases are the foundation of your applications. But for most companies, database management requires heavy lifting. You can free your teams from time-consuming database management tasks, such as provisioning, patching, and backups, by migrating to AWS fully managed database services, which offer continuous monitoring, self-healing storage, and automated scaling to help you focus on application development.

Amazon Aurora, a relational database built for the cloud, delivers the performance and availability of a commercial-grade database at one-tenth the cost. Over a hundred thousand AWS customers use Aurora for their relational database needs. For customers that need a relational database, AWS offers a clear economic advantage: According to a recent IDC study commissioned by AWS, a large organization with an average annual revenue of \$1 billion will realize business value worth an annual average of \$9 million (\$4.4 million per 100 Amazon Aurora databases). Additionally, with a 32 percent lower cost outlay on databases, the organization will see 13 months of payback and 439 percent of a three-year ROI. Additionally, Amazon Aurora I/O optimized, a new configuration, helps customers with I/O-intensive workloads predict database spend. With Aurora I/O optimized, customers simply pay for their database instances and storage with no charges for read and write I/O operations. This improvement in price performance can result in up to 40 percent cost savings when I/O spend exceeds 25 percent of their total Amazon Aurora database spend.

To help you move to managed cloud databases, such as Aurora, to start realizing cost savings quickly, AWS offers the <u>AWS Database Migration Service</u> (AWS DMS). AWS DMS is a managed migration and replication service that helps move your database and analytics workloads to AWS quickly, securely, and with minimal downtime and zero data loss. AWS DMS has successfully and securely migrated over 800,000 databases to AWS over the years.



The Pokémon Company International (TPCi) is a perfect example of how a purpose-built database can help you reduce costs. After the launch of Pokémon GO in 2016, the number of users ballooned to more than 300 million in two years, causing daily database management issues. The company migrated to Amazon Aurora with PostgreSQL compatibility and moved global configuration and time-to-live (TTL) data to Amazon DynamoDB. Since the migration, TPCi has significantly reduced the time and money spent on databases. "Using Amazon Aurora, we went from 300 nodes to 30, and we are no longer paying for database licenses," says Jeff Webb, former development manager at TPCi.



Choosing the right database for an application is critical to ensure that the application performs well and meets the needs of its users. It also ensures you get the best price and performance for your use case at scale. AWS also offers eight purpose-built databases designed to provide customers with optimal scale, performance, and cost. You can choose the right database for a variety of workloads, including high-traffic web applications, caching and geospatial applications, content management, high-scale industrial applications, fraud detection, IoT applications, and systems of record.



Dropbox is a global collaboration tool and file-sharing service. By mid-2018, the rapidly growing cold metadata—data that is accessed infrequently but needs to be stored durably and available instantly—was less than two years away from overwhelming Dropbox's datastore: Edgestore. Using Amazon DynamoDB and Amazon S3, Dropbox built Alki, a scalable cold metadata solution that would cost-effectively store metadata. Alki was rapidly prototyped and deployed on AWS within just a year, saving Dropbox millions in expansion costs. Additionally, Alki significantly lowered the "per-user gigabyte" costs" for Dropbox users. For reference—Dropbox's Edgestore would cost users 5.5 times more than Alki per user gigabyte per year.



Analytics

AWS analytics services are purpose-built to help you quickly extract data insights using the most appropriate tool for the job and are optimized to give you the best performance, scale, and cost for your needs. For example, <u>Amazon Redshift</u> helps you securely ingest, combine, and run historical, real-time, and predictive analytics on your data in just a few clicks with super-fast query results. In addition to five times better price performance than any other data warehouse, Amazon Redshift also uses performance telemetry from AWS's large customer base to optimize performance for real-world customer workloads—delivering up to seven times better price performance for high-concurrency, low-latency query workloads.

Amazon EMR, a managed big data service compatible with Apache Hadoop and Apache Spark, enables you to run big data applications and petabyte-scale data analytics faster and at less than half the cost of on-premises solutions. Amazon EMR offers a number of pricing options that can help businesses optimize their analytics costs based on their usage patterns.



Magellan Rx saw a

20%

reduction in operational costs on Amazon Redshift

Magellan Rx MANAGEMENTS

Magellan Rx Management (Magellan Rx) is a division of Magellan Health Inc. and a next-generation pharmacy organization. In 2020, Magellan Rx migrated to AWS and now hosts its data using Amazon Redshift, which enables users to query and combine exabytes of structured and semi-structured data across data warehouses, operational databases, and data lakes using standard structured query language (SQL).

Vinesh Kolpe, former vice president of information technology at Magellan Rx Management, says the company reduced operational costs by 20 percent on Amazon Redshift.



Infrastructure innovation

By moving to the cloud, you also get the benefits of AWS infrastructure innovation, which enables our data services to have greater price performance. For instance, many of our data services enable you to run workloads on <u>AWS Graviton</u> instances with little or no code change to get better price performance: <u>Amazon Relational Database Service</u> (Amazon RDS) with up to 50 percent better price performance, <u>Amazon EMR</u> with up to 30 percent lower cost, up to 15 percent improved performance for Apache Spark, and up to 20 percent performance improvement for Amazon Aurora.



Amazon RDS offers

50%

better price performance for Apache Spark on AWS Graviton Amazon EMR drives up to

20%

performance improvement for Amazon Aurora



Automate costly and timeconsuming processes

Managing your data infrastructure and getting value out of your data is time and people-intensive. Adopting technologies that help you automate time consuming and repetitive data management processes can save you resources that can be put to more innovative projects and speed up your time to value.





Zero ETL

Organizations want to break down data silos to effectively use data for faster, better decision making. But before that data can move into a data warehouse or data lakes and be put to use for analytics and ML, it has to be cleaned and organized through a process called extract, transform, and load (ETL).

ETL is critical to getting value from your data, but it's also complicated, time consuming, and difficult to do. The time spent on ETL can range from a few hours for simple data integration tasks to several weeks or even months for more complex projects involving large volumes of data and intricate data transformations. It often requires custom code that has to be regularly updated and offers plenty of opportunities for error.

To alleviate the significant cost and resource burdens of ETL, AWS is investing in a zero ETL future so you can quickly and easily connect to and act on all your data. For common use cases where ETL is regularly repeated with little value-add, we are integrating our services to decrease and, in some instances, eliminate the need for ETL. This includes features like the zero ETL integration between **Amazon Aurora** and **Amazon Redshift** for real-time analytics. We have also built integrations between our services to make it easier to do analytics and machine learning without the need for individuals to delve into the complexities of ETL. This includes a direct integration between Amazon S3 and Amazon Redshift for real-time data streams; and **Amazon SageMaker** integrations with Amazon Redshift, Aurora, **Amazon Athena**, and **Amazon Neptune** to perform ML with no data movement or ML expertise. And with federated querying with Amazon Athena and Amazon Redshift, you can run queries across data stored in your operational databases, data warehouses, and data lakes to provide insights across multiple data sources with no data movement.





Intelligent data services

ML can help automate typical tasks associated with maintaining data infrastructure in the cloud. We recommend looking for data services with built-in intelligence to eliminate time-consuming manual efforts.

For example, data engineers and developers spend many hours trying to uncover and remediate performance issues with databases—a costly yet prime-for-automation use case. <u>Amazon_PevOps Guru</u> uses ML to help developers easily detect, diagnose, and resolve performance and operational issues in their Amazon RDS engines.

Another example is <u>Amazon Macie</u>, which uses ML to automatically discover sensitive data stored on Amazon S3—providing visibility into data security risks and enabling you to automate protection against those risks. <u>Oportun</u>, a fintech lender and neobank with over 1.9 million members, needed a better way to quickly identify and remediate potential security risks to its members' personally identifiable information (PII). With Amazon Macie, Oportun was able to increase the accuracy of sensitive data discovery by 95 percent while reducing the cost by 99 percent.



95%

increase in the accuracy of sensitive data discovery²

99%

reduction in the cost of discovering sensitive data²



Amazo Macie



Leverage serverless architecture

Serverless architecture—building without managing servers—provides a more cost-effective, flexible, and efficient way to manage your data, build and run applications, and run analytics without the need for significant upfront investment in infrastructure or ongoing operational costs. AWS offers serverless options for several data services to help you innovate faster by reducing the amount of time and effort required for infrastructure management.

Amazon Aurora serverless saves customers up to 90 percent. All AWS analytics services including Amazon Redshift and Amazon EMR are available as serverless options, which enable additional economic and operational advantages. The serverless architecture makes it easier to analyze data at any scale without having to configure, scale, or manage the underlying infrastructure for data services. This helps reduce infrastructure costs, increase scalability, reduce operational overhead, and speeds up access to insights.

All AWS analytics services are available as serverless options



Roche, the largest pharmaceutical company in the world and the leading provider of cancer treatments globally, leverages Amazon Redshift Serverless to complete its data management without having to manage clusters and optimizes cost by provisioning just the right amount of capacity to meet demand. "Amazon Redshift Serverless is reducing the operational burden, lowering costs, and enabling scale for the Roche Go-to-Market domain. This simplification is a game changer, helping us rapidly onboard and support a variety of analytics-heavy use cases without friction," says Dr. Yannick Misteli, head of engineering, global product strategy at Roche.



Standardize your machine learning

Many organizations see ML as an essential part of their data strategy, but it can be time consuming and expensive, especially if the organization lacks the resources to scale ML effectively. Standardizing ML can help organizations achieve more efficient, consistent, and scalable results from their ML projects. For example, standardized ML practices can be easily replicated and applied across different use cases or departments, which improves consistency and increases collaboration among data scientists and other stakeholders. Standardizing ML can also help organizations streamline their workflows and avoid duplication of efforts. This can improve speed, increase efficiency, and help reduce the costs of innovating with ML.

Organizations can standardize their ML development with <u>Amazon SageMaker</u>. SageMaker makes using ML faster, cheaper, and easier and enables more people to innovate with ML through a choice of tools—an integrated development environment (IDE) for data scientists and a no-code interface for business analysts. It also allows customers to access, label, and process large amounts of structured and unstructured data for ML, reduce training time from hours to minutes with optimized infrastructure, and boost team productivity up to 10 times with purpose-built tools. SageMaker also helps you automate and standardize MLOps practices and governance across your organization to support transparency and auditability.

Another benefit of Amazon SageMaker is that it offers a number of flexible pricing options, including pay-per-use and annual subscriptions, that can help businesses optimize their ML costs based on their usage patterns. With On-Demand pricing, customers can pay only for what they use, no long-term commitments or upfront payments needed. <u>Amazon SageMaker Savings Plans</u> allow customers to save up to 72 percent compared to On-Demand pricing with a one-or three-year hourly spend commitment.



bazaarvoice

Bazaarvoice, a leading provider of product reviews and user-generated content solutions, migrated its ML workloads to Amazon SageMaker. SageMaker enabled the company to accelerate ML model deployment, reduce costs, and allow its engineers to deliver new features faster to clients. Not only can the company now deploy models faster, it has reduced costs by 82 percent—and it's reinvesting those cost savings to improve its service further.

Bazaarvoice reduced its ML inference costs by

82%

by using Amazon SageMaker Serverless Inference



Silicon innovation reduces machine learning training costs

AWS continuously works to create new ways of increasing performance while lowering the cost of storing, accessing, and getting value from data. To help customers improve performance while lowering costs, AWS has developed a portfolio of custom chips, including <u>AWS Trainium</u>, which reduces the cost to train ML models by up to 50 percent compared to similar GPU-based instances, and <u>AWS Inferentia</u>, which delivers high performance at the lowest cost for ML inference applications.



50% cost reduction to train ML models





Empower business users to selfserve their data needs

To scale data-driven decisions and innovations, business users need tools to access and understand data and insights quickly and easily. One major area of cost reduction here is to reduce the degree of separation between data and insights. You can achieve this by empowering business users to make data-driven decisions by themselves without engaging data engineering or data science teams or by learning a new skill.

<u>Amazon QuickSight</u> is our serverless, ML-powered business intelligence (BI) tool that enables business analysts to easily create, publish, and embed interactive data visualizations and dashboards to get insights from data. With <u>Amazon QuickSight Q</u>, users can query their data in natural language without writing a single line of code. And with the unique pay-per-session pricing, you pay only when your users access the dashboards or reports, which makes it cost-effective for deployments with many users.

<u>Amazon SageMaker Canvas</u>, a visual point-and-click interface, enables business analysts to generate accurate ML predictions—such as predicting customer churn from product consumption and purchase history data or predicting unplanned maintenance using historical event and operating data—without prior ML experience.



Best Western Hotels & Resorts, a global organization with eight brands and varied ownership structures across geographies, was able to consolidate analytics and BI projects.

Best Western Hotels
& Resorts ran Amazon
QuickSight at less than 30%
of the cost of its previous BI
solutions with Oracle.



2 Use insights and intelligence to drive cost reductions across your business

AWS makes it easier for you to build an end-to-end data strategy with databases, analytics, and AI and ML services that are the most cost-effective and deliver the best price performance at scale. But that's just one part of how AWS can help you reduce costs. For most companies, technology makes up a relatively small part of their cost structure, and there is an enormous opportunity to use AWS to develop new ways to streamline operations and reduce costs across the organization. We have helped a wide variety of customers identify areas of cost-savings using insights derived from their own operational data.

Let's explore some practical use cases where AWS has enabled organizations to discover cost-savings opportunities



Streamline operations

<u>Carrier</u>, a global provider of sustainable home, building, and refrigeration solutions, wanted to improve the "cold chain"—the process of moving perishable products from the point of origin to the final consumer. One-third of the world's food produced every day is never consumed due to waste, loss, or damage, primarily due to cold chain inefficiencies. Together with AWS, Carrier has developed a digital platform, Lynx, to apply advanced IoT, ML, and analytics technology to the cold chain process.

Real-time visibility throughout a cargo's journey now allows Carrier to deliver faster, data-driven decisions to customers across the supply chain. If a refrigerated truck's doors are open too long or a delivery is running late, customers informed of these potential problems in real time can manage issues as they arise. By sharing data across its own platforms, Carrier has been able to improve equipment uptime and fleet optimization and predict and manage costs, schedules, and resources. As a result, Carrier and its customers have realized decreases in energy consumption, delivery delays, cargo loss, and spoilage in transit—all adding up to more food for people to eat and cost savings across the board.

Lynx will help optimize cold chain operations, decrease energy use, and enhance outcomes with reduced costs, fewer delays, and less cargo loss and spoilage in transit. With AWS and Carrier's Lynx platform, 475 million tons of lost food could be saved with refrigeration each year.







Unify customer profiles to drive marketing efficiency

With 9,000 hotels globally, <u>Wyndham Hotels & Resorts</u> needed easy access to its customer data to expand its reach, grow its business, and optimize ad spend. Wyndham wanted to expand its loyalty program and drive direct bookings but lacked a single source of truth for customer records that would allow it to send more targeted communications. Customer profiles were scattered across disparate systems, and any attempt to match and merge that data was made through manual and inaccurate methods.

Wyndham worked with AWS Partner Amperity to build unified customer profiles that it needed to create segmented audiences and target with the right messaging. By bringing its data into Amperity built on AWS, unified customer profiles for both Wyndham's loyalty and non-loyalty members were created, giving its teams a single source of truth for customer insights and the ability to segment audiences and send targeted messaging to increase bookings and loyalty sign-ups. Information that previously took weeks to access can now be gained in just minutes, freeing up time to focus on other key initiatives. As a result, Wyndham saw \$2.5 million in annual incremental revenue from newly enrolled members and a 35 percent decrease in paid media costs.







Improve manufacturing quality and efficiency

Dutch brewing company <u>Heineken</u> retrofitted its breweries with digital capabilities, enabling employees to connect with equipment through purpose-built applications. With the help of AWS ML technology, automatic adjustments to the brewing process decreased extract losses by 11.2 percent. And as more breweries connect to Heineken's digital platform, optimizing the brewing process could save the company up to \$14 million.

Maintaining equipment and optimizing downtime is another step where ML assists. AWS ML technology makes intelligent predictions about when the next cleaning of a heat exchanger should take place. The technology could save Heineken up to \$21 million once scaled across all its breweries. And in the packaging process, ML helps optimize the speed of machines to ensure a steady stream of product continues down the line. The models have generated an output capacity increase of up to 10 percent, with potential savings of between \$14 million and \$36 million for the company. This digital transformation is setting Heineken up to thrive, now and in the future.

★ Heineken®





Proactively detect and prevent online fraud

Delivery Hero is a delivery service available in 74 countries that processes millions of orders a day. Its rapid growth in the first half of the pandemic was due to acquiring customers through promotions in the form of acquisition vouchers. When Delivery Hero faced a problem with customers abusing the acquisition vouchers, it needed a solution to stem financial losses.

Delivery Hero wanted to build a solution called Account Association to solve the problem of voucher abuse. Its initial design included data stored in a relational database. But as the solution evolved, the team understood the need to map affinities between data and bad actors. The team migrated the data to Amazon Neptune to generate account association graphs. With the insight that voucher abuse is a graph problem, the Delivery Hero team built an architecture based on Amazon Neptune and Amazon DynamoDB and was able to bring value to the business by increasing the amount of blocked fraudulent purchases by 32 percent.

increase in the amount of fraudulent purchases blocked by using Amazon
Neptune and Amazon Dynamo DR **Neptune and Amazon DynamoDB**







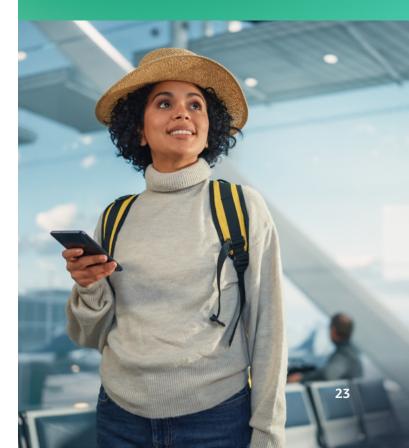
Use insights to find savings opportunities

<u>United Airlines</u> is partnering with AWS to transform every aspect of its operations and customer experience. This includes using <u>AWS IoT Core</u> and other AWS services to create an intelligent airport and process data from over 20,000 sensors, producing data at a rate of 28 gigabytes per minute. With this intelligent system, United can produce real-time insights, helping it optimize its ground equipment capabilities across more than 36,000 assets.

With less than \$10 million invested in its intelligent airport initiative, United Airlines has achieved \$120 million in savings for equipment it no longer requires, reduced emissions up to 7 percent, and saved 1.3 million personnel hours.

"This is an example where we're using the data and unleashing the data in real time—we're not storing it in a database. We're directing it to where it needs to go like a hose, trying to make sure that we use this data in motion to drive real-time decisions," says Jason Birnbaum, chief information officer at United Airlines. "The ultimate goal is that we create a great experience for our customers."







Conclusion

Whether it's due to the increasing volumes of data or ongoing economic uncertainty—optimizing the costs of your data strategy is critical to remain competitive in today's complicated and ever-changing business landscape.

AWS offers everything you need to build an effective end-to-end data strategy and pricing and services options that help you optimize costs at every step of the way—from ingesting, storing, and querying data to analyzing, visualizing, and generating predictions using ML models. Only AWS provides the most comprehensive set of data capabilities to give you optimal price performance for any data workload or use case. From databases for applications to storage for data lakes and from analytics and ML to end-user tools and solutions, AWS provides the right capabilities so you don't have to compromise on performance, cost, or results.

AWS enables you to take control of costs and continuously optimize your data spend while also using advanced analytics and ML to achieve cost reductions across your business. With AWS Data services, you can unleash the potential of your data and innovate for the future.

Learn more about how AWS can help you innovate with data while reducing costs >

