AWS FOR DATA

A builder's guide to AWS analytics services

The right analytics tools to make data and insights available to everyone in your organization

Adapt to market changes with agility

Organizations everywhere are facing unpredictable times. While you can't predict the future, you can put a solid data architecture in place that allows you to adjust rapidly in times of change. A modern data architecture with the right analytics tools can make that possible. A data-driven organization built upon a modern data architecture will have the capacity to gain insights, make decisions quickly, and build for the future. If your organization can use data to increase agility, save costs, and leverage new technologies like artificial intelligence (AI) and machine learning (ML), you'll be in a position to drive results and reinvent your organization for a new wave of innovation.

Data-driven organizations treat data like an organizational asset instead of the property of individual departments. They make data available and accessible to the people and systems that need it to fuel innovation and uncover new opportunities. With the right governance controls in place, data becomes an important asset that allows the organization to remain relevant now and in the future.

Challenges stand in the way of transformation

1. The first challenge is a technology challenge. Data volumes are large and ever-growing. The data your organization needs to analyze exists in every format. Navigating decisions around the best way to store and manage your data can ultimately affect the agility of your business. Understanding the best tools to derive insight from the intersection of data from across your organization – whether sales and supply chain data, or Amazon product reviews, or TikTok mentions – is difficult. And finally, ensuring the use of customer data is in compliance with your organization's policies and global regulatory requirements.is complex Organizations need to be able to adeptly respond to all of these technology challenges to be data-driven.

2. Next, there are challenges that confront the people who need to access your data. Often, the people challenges are the hardest to overcome. Different types of users have different tooling needs and skillsets. Business users want a visual experience to analyze data. Data teams want notebooks or integrated development environments. Business Intelligence users want a data warehouse, and the data science team only works with data lakes. Enabling them with the right and preferred tools to complete their work efficiently, accurately, and insightfully is essential.

3. Finally, there are business challenges to manage. More users desire access to more data to accelerate their decision-making. And the number of data-led projects are growing faster than budget. It's important to consider the direct and hidden costs of building on existing systems, while optimizing for the future and managing cost.

4. Underlying all of these challenges is the complexity of managing data governance and security. In a world where data security and privacy, as well as regulatory compliance, have become integral to business growth, organizations need to be able to carefully define, monitor, and manage who has access to specific data. Access control needs to be implemented not in a siloed, piecemeal fashion but in a comprehensive and unified way across all their data, analytics, and machine learning solutions.

Business leaders recognize these needs. They are taking action to modernize their data and analytics infrastructure by moving to the cloud and adopting a modern data architecture approach. With a cloud-based approach, organizations can break data silos and have seamlessly connected data stores that lead to new use cases and insights not previously possible.

Turn insight into action with a modern data architecture approach

Amazon Web Services (AWS) helps organizations turn insight into action through the modern data architecture approach, which provides a comprehensive plan to manage, access, analyze, and act on data. There are three key pillars of an end-to-end data strategy:

1. **Break down your data silos,** by connecting your data lakes and your data warehouses, and all data sources wherever they are.

2. **Break down your people silos,** by connecting all your data with everyone in your organization. Implement end-to-end governance that puts data securely in the hands of the right people in your organization.

3. **Break down your business silos,** delivering the best price performance, making it easier to balance cost and modernize the legacy systems, while setting the pace of innovation.

CUSTOMER DATA IN THE REAL WORLD

Customers across all industries are becoming truly data-driven using a modern data architecture built on AWS

Epic Games, creator of the blockbuster cross-platform game *Fortnite*, uses AWS to build new games and enhance the in-game experiences for its millions of online players around the globe. The AWS fault-tolerant, highly performant infrastructure enables Epic Games to reliably support more than 125 million players worldwide who flock to *Fortnite* to win a "Victory Royale."

"We decided to go all-in using AWS because they enable us to offer a quality gaming experience to millions of gamers around the world, simultaneously. We are excited to work with AWS to expand our use of analytics, machine learning, and containerized applications using Kubernetes to make our sizable infrastructure even easier to maintain."

Chris Dyl, VP, GM of Online Services, Epic Games

Nasdaq

"We were able to easily support the jump from 30 billion records to 70 billion records a day because of the flexibility and scalability of Amazon S3 and Amazon Redshift."

Robert Hunt, VP of Software Engineering, Nasdaq

Build your Modern Data Architecture foundation

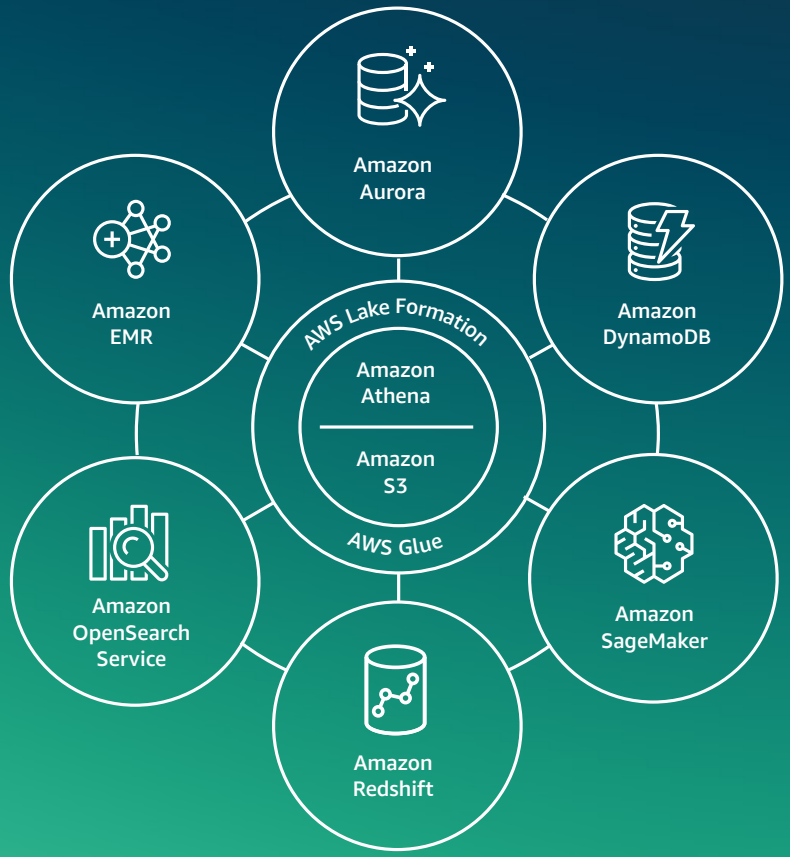
A Modern Data Architecture requires a solid foundation. As a first step toward a Modern Data Architecture, you need the best data warehouse and the best data lake solutions that work seamlessly together and provide analysts and data scientists a 360-degree view of the business.

Amazon Redshift

Amazon Redshift is the most price-performant cloud data warehouse. It lets you run and scale analytics in seconds on all your data without having to manage your data warehouse infrastructure.

Amazon S3

Data lakes, enabled by **Amazon S3**, let you store and retrieve any type of data at any scale. Amazon S3 is the best place to build a data lake with unmatched durability, availability, and scalability. It provides the best security, compliance, and audit capabilities and the fastest performance at the lowest cost. Amazon S3 also offers the most ways to bring data in and the most partner integrations.



Amazon Athena
Serverless, interactive analytics

Amazon EMR
Cloud big data processing at half the cost of on premises[1]

Amazon OpenSearch Service
Log and search analytics

Amazon Kinesis
Real-time streaming analytics

Amazon Redshift
Data warehousing with up to seven times better price performance[2]

Amazon SageMaker
High performance, scalable machine learning

Analytics services for impactful outcomes

Building a data lake or data warehouse is just one piece of a modern data architecture approach. You also need analytics services that can be used on top of that data, and are optimized for your use case.

AWS offers the broadest and deepest portfolio of analytics services. These services are all built so you never have to compromise on performance, scale, or cost when using them. For example, Amazon Redshift delivers up to 7 times better price performance than other cloud data warehouses, and Apache Spark on **Amazon EMR** runs 3.9 times faster than standard Apache Spark 3.0, which means petabyte-scale analysis can be run at less than half of the cost of traditional on-premises solutions.

[1] Petabyte-scale analysis can be run at less than half of the cost of traditional on-premises solutions
[2] Versus other cloud data warehouses

Break down data silos by unifying your data

You need to connect to all your data sources, handle system failures and data changes, and combine, clean, and normalize data to make it ready for analysis and machine learning.

Amazon Redshift and Amazon Athena both support federated queries, the ability to run queries across data stored in operational databases, data warehouses, and data lakes. This provides insights across multiple data sources with no data movement and no need to set up and maintain complex extract, transform, load (ETL) pipelines. Amazon Redshift data lake export allows organizations to unload data from their data warehouse to their data lake in open formats, ready for analytics.

You also want to make gaining insights from data fast and easy for all users. Amazon Athena is a serverless interactive query service that makes it easy to analyze data directly in Amazon S3 using standard SQL. Built on open-source frameworks, supporting open-table and file formats. Athena provides a simplified, flexible way to analyze petabytes of data where it lives.

You can also seamlessly move and integrate all your data at any scale with a critical piece of the data puzzle—**AWS Glue**.

AWS Glue is a serverless, scalable data integration and ETL service that makes it easier to discover, prepare, move, and integrate data from multiple sources for analytics and machine learning. AWS Glue provides you with all the capabilities you need for data integration so that you can gain insights in minutes instead of months.

The most impactful data-driven insights come from getting a full picture of your business and customers. This can only be achieved when you connect the dots between your different data sources across multiple departments, services, on-premises tools, and third-party applications. AWS is building a zero ETL future so you can quickly and easily connect to and act on all your data. With direct integrations between AWS services, we are eliminating ETL for common use cases so teams can move faster.

AWS Lake Formation
Centrally manage and secure data lakes

Amazon DataZone
Share, search, and discover data at scale

AWS Glue Catalog
Simple, serverless, and scalabe data integration

Break down people silos with end-to-end data access and governance

One of the most critical pieces of a modern data architecture is the ability to centrally authorize, manage, and audit access to data—making it available for all users to gain insights and put that data to work. It can be challenging to achieve this because managing security, access control, and audit trails across all the data stores in an organization is complex and time-consuming. It's also error-prone because it requires manually maintaining access control lists and audit policies across all storage systems, each with different security, data access, and audit mechanisms.

To empower developers, business analysts, and data scientists to break down silos and discover, collect, and analyze data in a secure and governed way, the AWS modern data architecture approach provides organizations with capabilities like AWS Lake Formation. Lake Formation uses the AWS Glue Data Catalog to automatically discover, tag, and catalog data. Lake Formation then lets user centrally define and manage permissions, governance, and auditing policies—all in one place. Finally, **AWS Lake Formation** simplifies cross-account data sharing, making it easier to securely share data across the organization.

It's also important to have a 360-degree view of data regardless of where it is stored. **Amazon DataZone** is a data management service that lets you publish data and make it available to a business data catalog through a personalized web application. You can access your data more securely regardless of where it is stored—on AWS, on premises, or in SaaS applications.

Organizations can now provide fine-grained access to data to the right user, at the right time and effectively meet their regulatory, governance, and compliance requirements.

Break down your business silos delivering the best price performance

AWS is committed to providing the best performance at the lowest cost across all analytics services, and we continue to innovate to improve the price performance of our services. In addition to industry-leading price performance for services like Amazon Redshift (up to 7 times better price performance than other cloud data warehouses) and Amazon EMR (3.9 times faster than standard Apache Spark 3.0), AWS customers have saved more than $750 million in storage costs by using Amazon S3 Intelligent-Tiering.[3]

AWS eliminates the need for manual infrastructure and server management. AWS managed and serverless architectures enable you to shift more of your operational responsibilities to AWS increasing your agility and innovation.[4] Serverless allows you to build and run applications and services without thinking about servers, eliminating infrastructure management tasks. With serverless, customers only pay for resources they use without worrying about capacity planning. It also allows them to get the high availability and scalability because the services scale up when more capacity is needed and scale down when not.[5]

[3] Amazon Redshift, Amazon EMR, and Amazon S3 intelligent tiering saves organizations up to 40 percent on storage costs for data stored in a data lake
[4] With flexible pricing model AWS Savings Plan
[5] With AQUA for Amazon Redshift vs. other cloud data warehouses

CONCLUSION

Modernize your data infrastructure with AWS

The next wave of reinvention is here with data and analytics

With a modern data architecture enabled by a portfolio of AWS analytics services, you and your organization can build for the future, lead in your industry, and unlock limitless possibilities for reinvention. You'll be able to handle massive amounts of data with an architecture that scales and evolves as needs change while breaking down silos to combine, analyze, and gain insights from data of all types, regardless of where it lives. You can rapidly gain insights from all your data, improving agility and accelerating decision making to stay ahead of changing customer needs and market dynamics. And you'll be able to do it all with proven governance and security controls designed to meet today's strict data security, privacy, and compliance regulations.

AWS provides the broadest and deepest portfolio of analytics services. Our portfolio includes the most scalable data lakes, the broadest set of analytics services, and seamless data movement and unified governance—all delivered with the best performance at the lowest cost. AWS makes it easier than ever to get started. If you're ready to make your data work harder and smarter, we're ready to help.

Learn more ›