



Put your data to work with a modern analytics approach

Make data accessible to your entire organization using the Lake House approach, which combines the best of both data lakes and purpose-built data stores

Adapt to market changes with agility

Organizations everywhere face unprecedented, unpredictable times—and these times are changing rapidly. Many organizations have already begun modernization efforts and strategies to become data-driven, putting them a step ahead of those that continue to resist change. Accessing insights from their analytics helps even these organizations who are moderately data-driven make more informed moves and rapidly respond to change.

Organizational leaders adapting successfully to changing times know that their *next* action matters more than their *last*. If they can increase agility and address some of the top concerns, such as growth and gaining efficiency, they can continue to drive results and reinvent their organization for a whole new wave of innovation.

Making data more connected, accessible, and available can help you achieve rapid growth and discover more efficient ways of working. That means every person in your organization who needs access to the data should have it—anytime, anywhere, and everywhere they need it.



Challenges stand in the way of transformation

- 1** | The first challenge that organizations face is the sheer size and scale of data that they handle every day. Data continues to grow exponentially every year. Over the next three years, there will be more data created than in the prior 30 years combined.¹ The old on-premises tools and the old data stores from the last 20 to 30 years are not going to meet today's demands. To handle the massive scale and tremendous growth in data volumes that we see today, organizations need new data stores that can scale and grow as needs change, whether that is from the gigabytes and terabytes handled today or the petabytes and exabytes expected in the future.
- 2** | Secondly, organizations need to easily access and analyze all types of data such as log files, clickstream data, voice, video, etc. This data comes from a variety of sources and is stored in silos across multiple data stores. To gain new insights from all this data, organizations need to be able to easily break down these data silos so that they can combine and analyze all their data regardless of where it lives.
- 3** | The third major challenge that organizations face is adapting with greater urgency to changing customer preferences and market dynamics. To make better and faster decisions, organizations need to empower their employees with secure access to data and the ability to perform analytics and machine learning on their data in an agile and cost-effective way. Organizations running legacy data infrastructure on-premises spend a great deal of time on hardware and software installation, configuring the infrastructure for performance and availability and spending unnecessary time on capacity planning to scale their systems. All of this reduces agility and slows down decision-making.
- 4** | Finally, in a world of increasingly important data security, privacy, and compliance regulations, organizations need to be able to carefully define, monitor, and manage who has access to specific pieces of data through tried and tested data governance and security controls. They need to do this not just for the data in their individual data silos but in a comprehensive and unified way across all their data, analytics, and machine learning solutions.

As a result, there is an acceleration in organizations looking to modernize their data and analytics infrastructure by moving to the cloud and adopting a Lake House approach. By doing this, organizations are able to transform data that was previously siloed into continuous data connectivity and accessibility.

Customer data in the real world

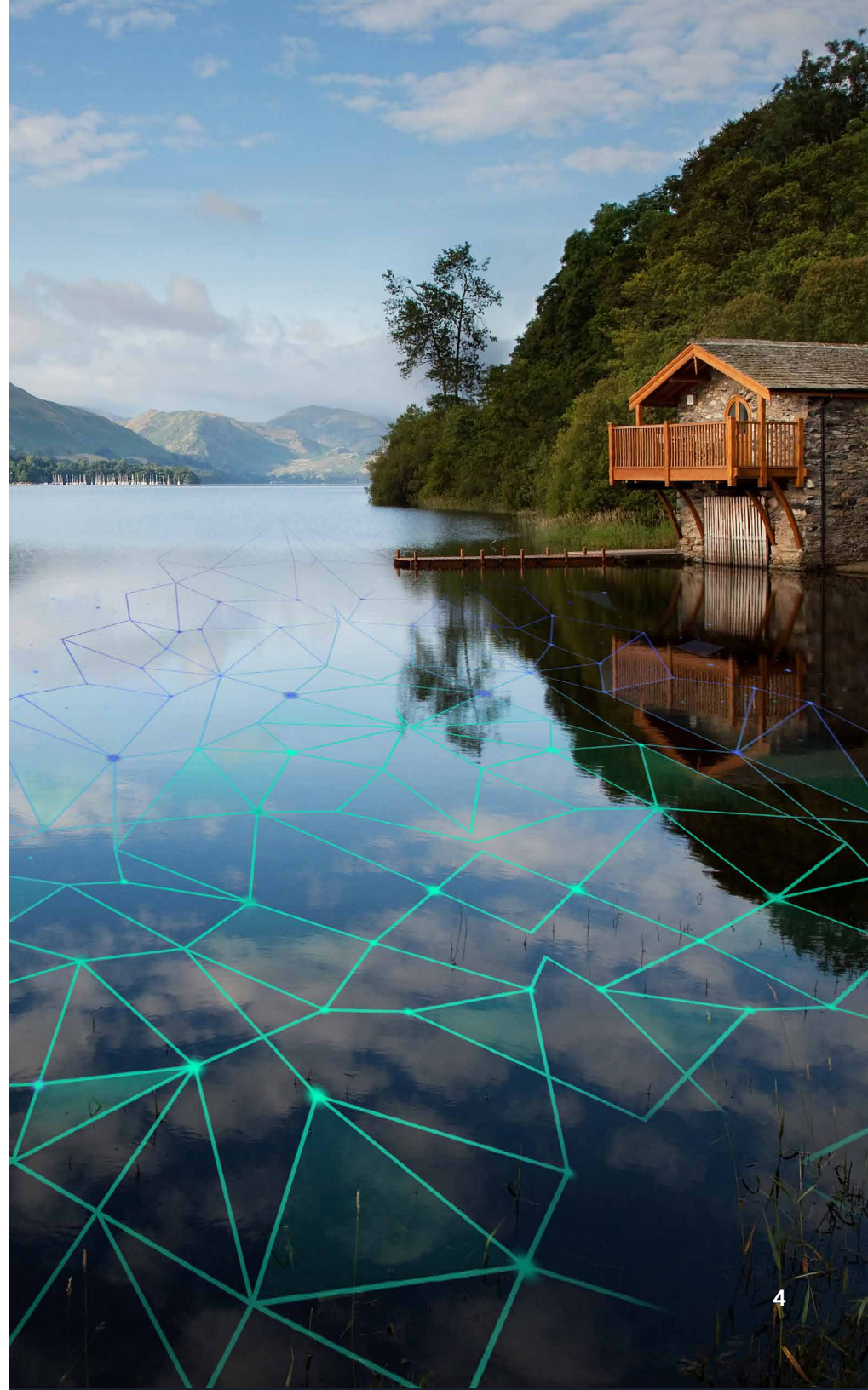
Creating a data strategy has become imperative for organizations that want to stay relevant in the future. To accelerate any major business initiative, from creating new customer experiences to building new revenue streams, leaders must be able to quickly gather insights, get to the truth, and leverage it to look around corners, predict what's next, and act on opportunities.

This requires organizations to:

- Store any amount of data in open formats
- Break down disconnected data silos
- Empower people to run analytics or machine learning using their preferred tool or technique
- Manage who has access to specific pieces of data with the proper security and data governance controls

AWS helps organizations do this through the Lake House approach, which brings together the best of both data lakes and purpose-built data stores.

With a Lake House approach, organizations can move any amount of data from various silos into an Amazon Simple Storage Service (Amazon S3) data lake. Unlike other cloud providers, organizations can store their data in S3 using standards-based open data formats to avoid being locked into any one proprietary data format or approach to analytics. Storing data in standards-based open formats makes it easy for any analytics or machine learning service to work on the data. It also eliminates the need to unnecessarily



move, transform, or reformat the data in order to get value from it. This is particularly useful when working with data at petabyte and exabyte scales. For example, Amazon Athena is an interactive query service that lets organizations instantly analyze data stored in S3 using standard SQL without having to set up and manage any servers.

In addition to using a data lake, organizations also use purpose-built data stores to get the best performance, scale, and cost advantages for their use cases. Amazon Aurora processes transactions at high speed, and Amazon Elasticsearch Service stores and analyzes large volumes of log data at low cost. In a world where we frequently need to work with terabytes, petabytes, and sometimes exabytes of data, purpose-built data stores are designed specifically to run a particular workload or use case extremely well. And that is why we have spent the last several years building the right tools for the right job. We provide 15 purpose-built databases, 12 purpose-built analytics services, and 30 machine learning services—more than you'll find anywhere else by a fair amount.

To get the most from their data lakes and these purpose-built stores, organizations need to move data between these systems easily. For instance, clickstream data from web applications can be collected directly in a data lake, and a portion of that data can be moved out to a data warehouse for daily reporting. We think of this concept as *inside-out data movement*.

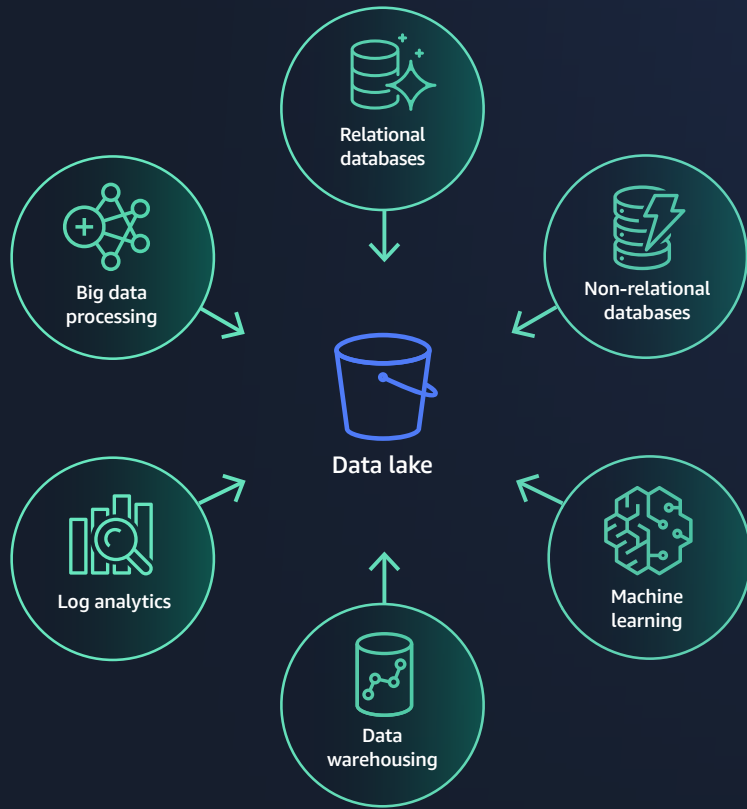
Inside-out data movement

There are many scenarios in which you'll need to perform inside-out data movement easily. One example of inside-out data movement is generating reports from clickstream data collected via a web application. This clickstream data gets stored directly in a data lake, so before you can generate your report, you have to move the data into a data warehouse.



Outside-in data movement

Additionally, you'll find yourself in scenarios where you need to move data in the opposite direction: from your data warehouse into your data lake. For example, suppose you want to run product recommendation algorithms against a larger dataset stored in your data lake using machine learning. In that case, you'll need to move data, such as sales product information by region, from your warehouse into that data lake.



Around-the-perimeter movement

You may find that you need to move data from one purpose-built store to another and never involve your data lake at all. Let's say you are trying to make your product catalog easier to search by offloading queries to a search service. You would need to move a copy of that catalog data from the warehouse where you store it to the search service.

When silos become barriers

As data in these data lakes and purpose-built stores continue to grow, it becomes harder to move all this data around. We call this data gravity. This force makes it harder to move data in, out, and around lakes and purpose-built stores.

But there's a way to tackle the weight of data gravity. You can construct your data systems around a central data lake and surround it with a ring of purpose-built services and allow data to be moved easily between these data stores in a secure and governed way. This agile, secure, and governable way of constructing your data system will also help liberate your data to give everyone access across your organization, unlocking the insights that lead to informed decision-making and organizational reinvention. This modern approach to analytics is called the Lake House.



Liberate your architecture

Taking a one-size-fits-all approach to your data architecture leads to compromises. So, the Lake House approach is not simply about integrating a data lake with a data warehouse but rather about integrating a data lake, a data warehouse, and purpose-built stores in such a way that enables unified governance and easy data movement.

The Lake House approach on AWS tackles data gravity concerns as well as compliance, security, and governance requirements. It provides an architecture for a broad and deep collection of purpose-built data services to support your organization in its reinvention, innovation, and analytics goals. It also delivers high-performing, cost-effective, and scalable data lake storage solutions.

A Lake House approach supports five things:

- 1 Allows you to build a scalable data lake rapidly
- 2 Supports purpose-built data and machine learning services that deliver performance for your organization's use cases
- 3 Ensures you can move data seamlessly in, out, and around the data lake and purpose-built data services
- 4 Maintains compliance with security, monitoring, and management of data access
- 5 Delivers cost-effective performance and scalability



Reinvent with scalable data lakes at the center

Data lakes, available with the Amazon S3 solution, let you store and retrieve any type of data at any scale. It has unmatched durability, availability, and scalability, the best security, compliance, and audit capabilities, the fastest performance at the lowest cost, and the most ways to bring data in.

It's possible to build your data lake manually with Amazon S3. However, AWS Lake Formation is a simple, straightforward solution developed to eliminate the need for the manual tasks associated with building your own. It can help you spin up secure data lakes in the cloud in days instead of months.

All manual and time-consuming tasks are automated and optimized with Lake Formation. It collects and catalogs data from databases and object storage, moves the data into an Amazon S3 data lake, cleans and classifies data using machine learning algorithms, and secures access to sensitive data.

Amazon S3 and AWS Lake Formation are ideal solutions for establishing data lakes in your Lake House approach because of their ease of use, durability, availability, and scalability.



Nu Skin Enterprises is a global direct selling company that distributes more than 200 premium-quality anti-aging products in both the personal care and nutritional supplements categories.

"We were challenged with expanding capability and scaling throughput of our existing analytics systems. Our data was distributed amongst various disconnected databases and SaaS solutions, making it difficult to analyze data at scale while restricting access to sensitive data. To overcome this challenge, we built a data lake solution on AWS. This allowed us to aggregate data from various data silos into Amazon S3 where we cataloged and secured all data using AWS Lake Formation."

Joe Sueper

VP Enterprise Architecture, Global Technology, Nu Skin Enterprises

Integrate purpose-built analytics services for impactful outcomes

Building a data lake is just one piece of a modern analytics approach. Organizations are using purpose-built analytics and machine learning services on top of their data lake in Amazon S3 to solve various analytics use cases better by storing or processing the data in a way that is optimized for that particular use case.

AWS offers the broadest and deepest portfolio of purpose-built analytics and machine learning services. These services are all built to be best-of-breed, which means you never have to compromise on performance, scale, or cost when using them. For example, Amazon Redshift delivers up to 3X better price performance than other cloud data warehouses, and Apache Spark on Amazon EMR (Elastic MapReduce) runs 1.7X 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.



² Petabyte-scale analysis can be run at less than half of the cost of traditional on-premises solutions
³ Vs. other cloud data warehouses



Amazon Athena
Interactive query



Amazon EMR
Big data processing
½ the cost of on-premises²



Amazon Elasticsearch Service
Log and search analytics



Amazon Kinesis
Real-time analytics



Amazon Redshift
Data warehousing
3X better price performance³



Amazon SageMaker
Machine learning

Nasdaq uses AWS services for flexibility, scalability, and performance

Challenge:

With the mission-critical nature of its business, Nasdaq is always looking for new cloud technologies to help it scale to meet the sometimes unpredictable nature of volatility across the capital markets ecosystem. To strengthen the ability to scale with market volumes and improve reporting and customer response times, Nasdaq moved from an on-premises data warehouse to AWS for its U.S. market's transactional data.

Solution used:

Migrated from on-premises data warehouse to Amazon S3 and Amazon Redshift

Results:

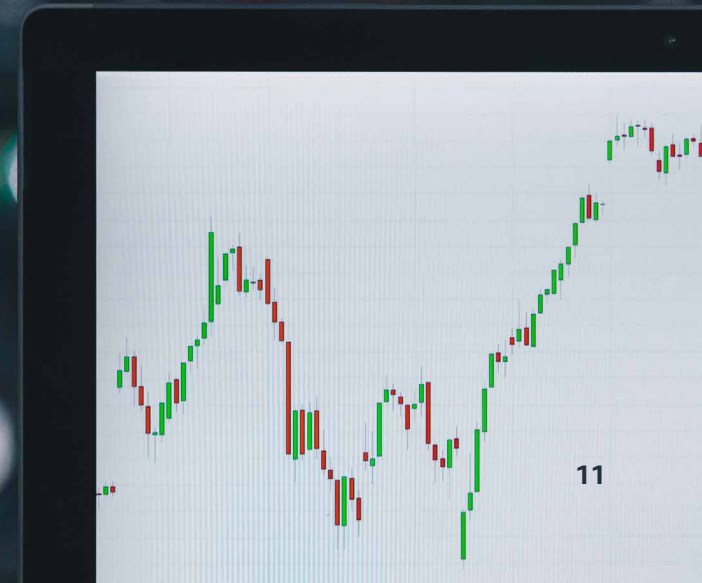
- Jumped from 30 billion to 70 billion records a day with no disruption
- Scaled compute and storage data independently to better support high volumes of transactions
- Improved market data load times leading to up to five hours faster time-to-insight



"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

Vice President of Software Engineering, Nasdaq



Let your data flow freely and securely

With data stored in several different systems, you need the freedom and flexibility to move that data between all of your services and data stores: inside out, outside in, and around the perimeter.

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 to provide insights across multiple data sources with no data movement and no need to set up and maintain complex extract, transform, and 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 can also seamlessly move your data between all the services and stores in your Lake House with a critical piece of the data puzzle—AWS Glue. AWS Glue is a serverless data integration service that allows you to easily prepare data for analytics, machine learning, and application development. AWS Glue provides you with all the capabilities you need for data integration so that you can gain insights in minutes instead of months.

To make it even easier to combine and replicate data across multiple data stores, AWS Glue Elastic Views enables developers to create materialized views across a wide variety of databases and data stores using familiar SQL, reducing the time it takes to combine and replicate data across data stores from months to minutes. AWS Glue Elastic Views handles copying and combining data from source to target data stores, continuously monitors for changes in source data stores, and automatically updates the materialized views to ensure that user-accessed data is always up to date.

Unify and govern data access for all members of your organization

One of the most necessary pieces of a modern analytics approach is authorizing, managing, and auditing access to data—liberating it for all 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 Lake House approach provides organizations with capabilities like AWS Lake Formation, which includes a data catalog that automatically discovers, tags, and catalogs data and provides an easy way to centrally define and manage security, governance, and auditing policies—all in one place. This enables organizations to provide fine-grained access to data to the right user at the right time and effectively meet their regulatory governance and compliance requirements.



Improve performance and reduce costs with a modern analytics approach

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 (3X better price performance than other cloud data warehouses) and Amazon EMR (1.7X faster than standard Apache Spark 3.0), Amazon S3 intelligent tiering saves customers up to 40 percent on storage costs for data stored in a data lake.

To set the foundation for the new scale of data, AWS also released AQUA (Advanced Query Accelerator) for Amazon Redshift. AQUA is a new distributed and hardware-accelerated cache that brings compute to the storage layer and delivers up to 10X faster query performance than other cloud data warehouses. AQUA is available on Amazon Redshift RA3 instances at no additional cost, and organizations can take advantage of the AQUA performance improvements without any code changes.

40%reduced
storage costs⁴**400
Gbps**ethernet
networking**350+**More than 350
instance types**72%**Up to 72% savings
on compute usage⁵**10X**faster query
performance⁶

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



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. AWS's 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

Director of Platform, Epic Games



Liberate and modernize your data with AWS

The next wave of reinvention is here with data and analytics

With a Lake House approach built on a portfolio of purpose-built services from AWS, you and your organization can build for the future, lead in your industry, and come out on the other side with limitless possibilities for reinvention. You can rapidly gain insights from all of your data and deliver access to all of your users.

AWS provides the broadest and deepest portfolio of purpose-built analytics services to realize a Lake House. Our portfolio includes the most scalable data lakes, the broadest set of purpose-built analytics and machine learning services, seamless data movement, and unified governance—all delivered with the best performance at the lowest cost. Tens of thousands of organizations have already built data lakes using AWS, and we make it easier than ever to get started.

**Learn more about reinventing
your business with data »**

