

Modernize Data and Analytics by building a Lake House Architecture on AWS



Harness the power of your data with AWS Analytics

Financial institutions are collecting more types of data than ever before, to better understand their customers, assess risk, comply with regulations, and drive innovation. This data is often stored in silos that cannot scale to meet enterprise needs. End results may include business risks like trading errors and compliance breaches, or operational inefficiencies such as delays in onboarding clients, manual processes for account maintenance functions, and inability to handle business growth without a added headcount. This offering explores a modern and comprehensive approach to mitigating these problems through an AWS Lake House architecture.

Prior to recent advancements in data lake philosophies, traditional data warehouse strategies were the design of choice, often requiring frequent operational overhead while creating an additional layer of considerable complexity between data lake producers and consumers. Over time, this often results in a high-cost and/or unreliable working-model that consistently produces more cost, and more operational overhead, as the size of your warehouse data increases. Lake House architectures can reduce or eliminate the need for a data warehouse by leveraging object-storage querying mechanisms, eliminating the need for data transfers between a data lake and a data warehouse.



Comprehensive cross-account data queries

As the emphasis on gaining deeper and richer insights when bringing together relevant data increases, so do the exponentially growing data volumes. As a modern data lake allows you to have a single place to run data, a Lake House effectively utilizes a set of Catalog resources across AWS accounts. This Lake House approach not only integrates your data lake and data warehouses, AWS accounts can also be transformed into an account with new permission expectations. This Lake House architecture is the utmost importance of account structuring when one plans out this 'technical line-of-business'. This requires a strong understanding of account structuring and organizational units. If businesses are embarking on this journey for the first time, it's critical that explicit time is taken to plan out how this will be setup, orchestrated, monitored, managed, and then even further enabled for development with the organization, moving forward.

Features

Lake House architectures build on Data Lake concepts to incorporate components that facilitate features such as:



Ingestion and Curation Engine

Ingestion & curation is a critical part of building an effective Lake House. It is the process that facilitates getting data into a data lake. The process of sourcing, cleaning, and preparing data into a reputable, consumption-ready state.



Data Lakes in a Lake House Architecture

Data Lakes are commonly used to collect, organize, and leverage data for any number of reasons. They consist of large or extremely large sets of data. The process and components involved in the creation of a shareable Catalog reference of consumable data within a data lake.



Data Mesh Scalability and Compliance Enablement

The data mesh concept is important to understand when working with Data Lake Governance in Lake House Architecture. This covers how a data lake becomes a scalable, managed, shared, compliant, and cost-effective Lake House.

Vertical Relevance on AWS

AWS provides scalable, resilient and secure services that enable customers to build efficient solutions to meet their market needs. Vertical Relevance brings significant Financial Services and AWS technology experience to evaluate current and targeted capabilities against business goals and identify solution architectures and management processes to achieve those business goals. We partner with our customers to evaluate, prioritize and execute on programs to evolve their existing applications into cloud native solutions that are secure, reliable, perform efficiently and are optimized for cost, while managed using best practices.

Break out of database deadlock and into a world of innovation

By implementing a Lakehouse, an organization can avoid creating a traditional data warehouse. Organizations are enabled to perform cross-account data queries directly against a Lake Formation Data Lake through Redshift Spectrum External Tables and/or Athena. Table and Column-Level access granularity achieved through Lake Formation Permissions. Data Lake Governance enabled through Lake Formation Resource Shares. Multi-regional, parameterized, infrastructure-as-code deployments. Full data flow and processing pipeline with Glue Jobs, orchestrated by a single Step Function. Lake House architectures can reduce or eliminate the need for a data warehouse by leveraging object-storage querying mechanisms, eliminating the need for data transfers between a data lake and a data warehouse. Lake House architectures build on Data Lake provide benefits such as:



Improved Visibility & Continuous Reliability

No need to move data into and out of a data warehouse to perform analytics on a Data Lake



Improve operational efficiency

Reduction in operational overhead regarding Data Lake resource access permissions and delegation.



Centralized data access audit logging and compliance

Lake Formation and the use of a Lake House architecture ease the burden of realizing these complex compliance regulations using the Data Mesh concept.



Automation of data ingestion and exposure process

Step functions can be triggered on a regular schedule (daily, weekly, etc.) to facilitate automated scheduled curation of data.

How one of the world's largest investment companies is building a Retail Data Lake House



Challenges

Slowing down the delivery of data and analytics projects but also putting the organization at risk of fines for failing to meet security regulations.



Solution

Redesigned the structure of the governance approach, created a data consumer platform, and leveraged the existing compute solutions.



Results

Transformed data and analytics management strategy from several siloes into an efficient, centralized Lake House that will allow teams to reap all the benefits of data and analytics in the

Get started with Vertical Relevance solutions on AWS

If you are looking to provide automation, consistency, predictability, and visibility to your software release process [contact us today](#).