# NorthBay

# How to Modernize
# A Data Platform
# for Customer Retention

Inside are insights into leveraging machine learning and **generative AI**, how to choose the right modernization approach for your financial organization, best practices for data sharing and governance, and other top-of-mind topics.

aws
**PARTNER**
Premier Tier Services

## Ten Frequently Asked Questions About Data Platform Modernization

Every financial services IT leader is feeling the pressure to deliver on the promise of optimizing customer retention with innovative technologies including machine learning and generative AI. But most are struggling with how to move forward – and how to deliver results fast.

Inside established organizations in particular, CTOs are wrestling with whether to modernize their existing on-premises data platform with a centralized data lake in the cloud, or if now is the time to pivot toward the decentralized data mesh strategy that's rapidly gaining traction. Either way, without a better way to solve the seemingly never-ending data quality problems that at best cause delays and at worst compromise success, no amount of modernization is going to move the needle.

Sound familiar? As an AWS Premier Consulting Partner and Managed Service Provider, NorthBay Solutions works closely with CIOs/CTOs and their technology teams to unlock the value of their data in the cloud for deeper insights, faster decisions, and agile innovation. More than 300 customers rely on NorthBay to help them reach their business goals and help their teams become highly cloud proficient.

| 10+ | 350+ | 550+ |
|:---:|:---:|:---:|
| Years experience with AWS | AWS certified staff | AWS Launches |

Based on 10+ years of experience implementing enterprise-wide data lakes and data mesh solutions for financial services firms, we've compiled answers to the 10 questions most frequently asked by CIOs/CTOs on the journey to data and analytics modernization.

1. How can I enable our line-of-business leaders to leverage AI/ML and other innovative technologies to optimize customer retention?
2. What are the best financial services datasets for solving business challenges using analytics and machine learning?
3. How can I support enterprise-wide data sharing while enforcing data governance?
4. How can my IT team that has limited cloud and AI skills build and run a data platform?
5. Is "modernize in-place" a short-term fix or a longer-term solution?
6. Which is the right approach for analytics modernization: A centralized data lake or decentralized data mesh?
7. Why are so many financial services CIOs/CTOs choosing the cloud, especially AWS Cloud, for their data platform?
8. Are there proven reference architectures to accelerate my modernization journey?
9. Have fintech innovators found new ways to solve the data quality conundrum?
10. What are the top use cases for generative AI in the financial services industry?

### ① How can I enable our line-of-business leaders to leverage AI/ML and other innovative technologies to optimize customer retention?

Financial services customers want their provider to understand their specific situation and to provide them with the right offers at the right time. Moreover, they are increasingly seeking providers who can automate their financial decisions and money management—such as identifying advanced signs of financial stress and responding with solutions and advice—in order to save them time and money.

As awareness and acceptance of AI-powered interactions increases, so too does demand that continue to deliver a more humanized experience. In fact, research conducted by Capgemini reveals that 71% of financial services customers want to know when companies are enabling interactions via AI, and when making major financial decisions, 45% prefer human-only interactions.

**66% of customers** expect financial institutions to understand their unique needs, and 52% expect personalized offers.
source: **Salesforce**

Whether your organization chooses to modernize its existing data platform or embrace data mesh, enabling line-of-business leaders to leverage AI/ML requires the following capabilities:

### Solution blueprints

Pre-defined, high-level designs or templates that serve as a guide or roadmap for developing AI/ML applications, helping to streamline the development process, ensure best practices, and accelerate time-to-market.

### Data catalog

A comprehensive data catalog that provides metadata about the data, including data lineage, data definitions, and data relationships helps data scientists and other users to discover and understand the data in the data lake.

### Collaboration and sharing

Collaboration and sharing of data sets, models, and code, as well as the ability to collaborate on projects and workflows, is critical.

### Machine learning capabilities

Machine learning capabilities such as data preparation, model development, and model deployment are mission critical. This includes integration with AI/ML frameworks and libraries, such as TensorFlow or PyTorch.

### Analytics tools

Integration with analytics tools such as Tableau, Power BI, or Jupyter Notebooks are required to enable data exploration, visualization, and reporting.

### Data integration

A scalable, reliable, and automated means of integrating structured and unstructured data from various sources to ensure that data is continuously updated and available for AI/ML applications.

### Data quality

High-quality data is essential for accurate AI/ML modeling. Processes must be in place to ensure data quality, such as data validation, cleansing, and transformation.

### Data governance

A robust data governance framework is necessary to ensure that data is secure, compliant, and well-managed. This includes policies and procedures for data access, data privacy, and data retention.

### Scalability and performance

The capacity to handle the large volumes of data required for AI/ML, to store and process both structured and unstructured data, as well as the ability to scale up or down based on demand.

## 2 What are the best financial services datasets for solving business challenges using analytics and machine learning?

Financial services organizations are leveraging analytics and machine learning to mitigate risks such as fraud and cyber threats, as well as to provide a modern customer experience. Following are five top financial and economic datasets and data sources for financial data for machine learning.

### ✓ AWS Data Exchange

Includes insurance claims, consumer spending patterns, personal income data, customer demographics, ESG (environmental, social, and governance) ratings, and Financial APIs.

### ✓ Data.gov

The United States government's open data website that provides access to datasets published by agencies across the federal government.

### ✓ Global Financial Data (GDF)

An extensive database of current and historical financial data on topics such as market indicators, exchange rates, commodities, incomes, and more.

### ✓ American Economic Association

Macroeconomic data from across the US covering employment, economic output, budget, economic trade, and more.

### ✓ Kaggle

An online community of data scientists and machine learning practitioners that is also a source for publishing and finding datasets, including 5,000 finance-related datasets.



THIRD-PARTY DATA FOR FINANCIAL SERVICES ORGANIZATIONS

Third-party data for Financial Services organizations | (1:12)

## 3 How can I support enterprise-wide data sharing while enforcing data governance?
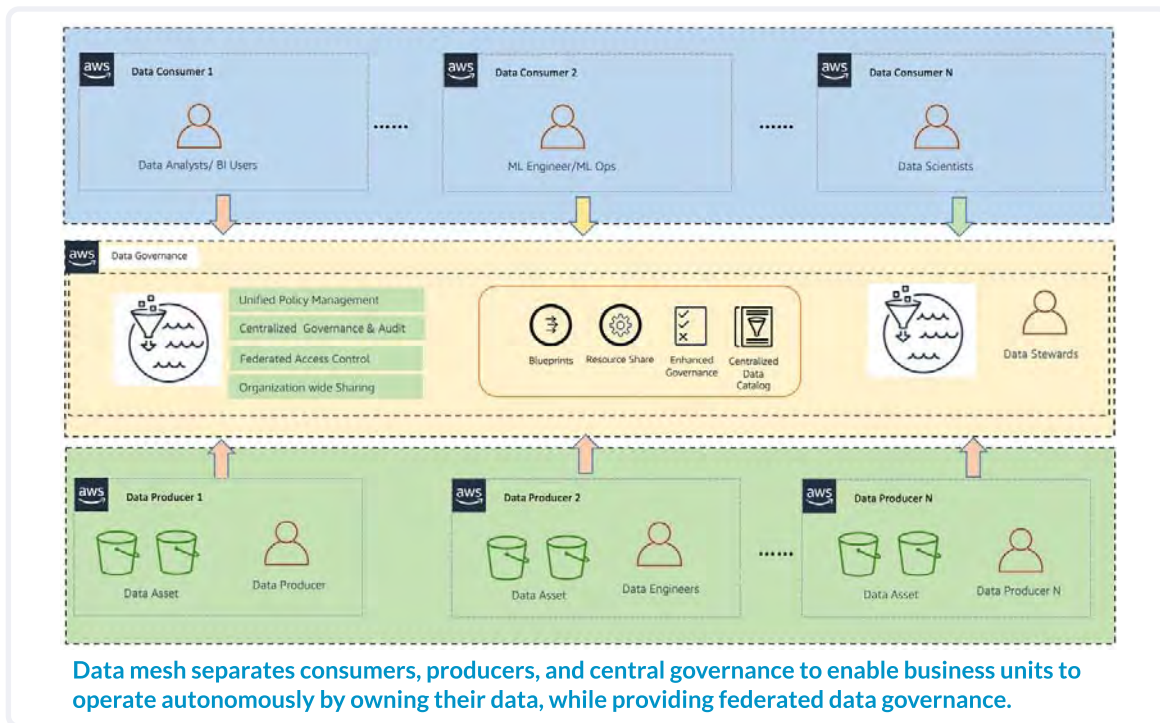
By integrating data from various source systems (CRM, Web, third party datasets, etc.), and sharing it across lines of business, financial services organizations focused on customer retention can:

- Create a 360-degree view of customers.
- Identify friction points and create more personalized experiences.
- Trigger early interventions based on anomalies.

Data governance plays a crucial role in ensuring that the shared data is accurate, trustworthy, and secure. With a centralized data platform, a robust data governance process is needed to ensure that data that is ingested and shared is secure, compliant, and well-managed. This includes policies and procedures for data access, data privacy, and data retention.



**Representative centralized data lake platform.**
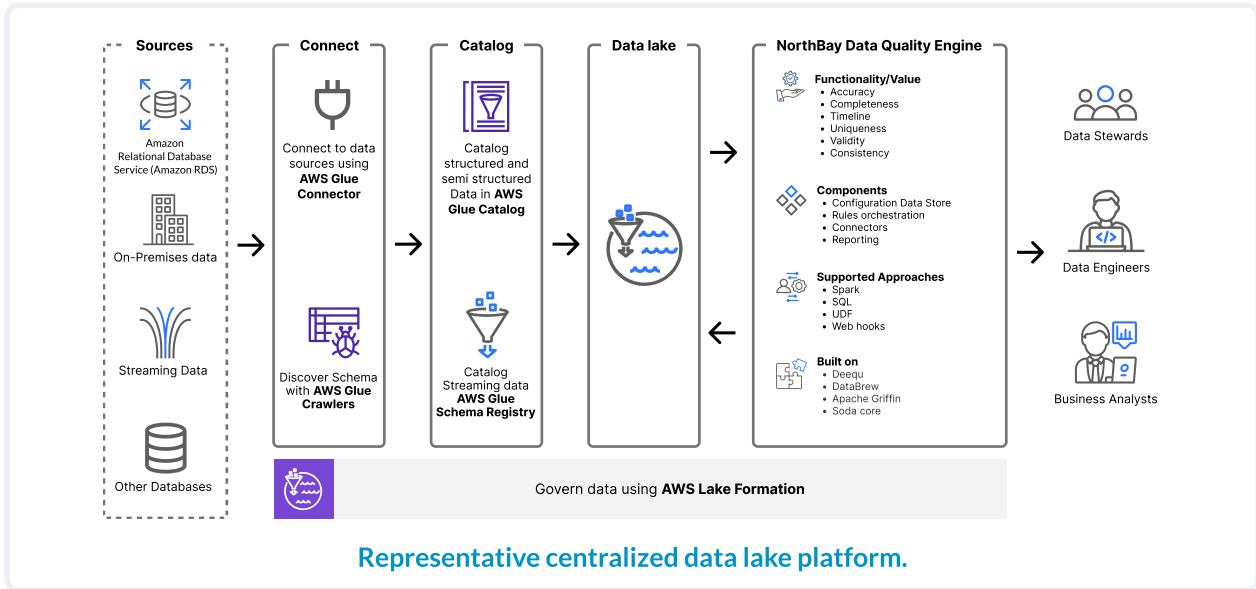
With a decentralized data mesh approach, a data governance framework is essential for effective data sharing. Typically, a centralized data team provides a data catalog of available datasets and builds the workflow to enable users to access specified datasets for specified purpose and timeframe, and publishes guidelines and standards that ensure that the shared data is accurate, trustworthy, and secure, while also ensuring compliance with legal and regulatory requirements.



**Data mesh separates consumers, producers, and central governance to enable business units to operate autonomously by owning their data, while providing federated data governance.**

**4**   **How can my IT team that has limited cloud and AI skills build and run a data platform?**

The cloud has fundamentally changed the financial services industry — and how IT teams engage with technology partners. Consulting firms such as NorthBay, who is an AWS Premier Consulting Partner and an AWS Managed Service Provider, have a deep understanding of AWS technologies and best practices and how to best apply them to develop data lakes and analytics solutions.

Following are three ways we help financial services organizations modernize their data platforms to optimize customer retention:

1 Through the **NorthBay OneTeam process**, we work in close collaboration and coordination with our customers. We share cloud and AWS best practices so IT teams gain the skills to maintain and optimize the environment moving forward. In many cases, we take on the heavy lifting so that internal teams can stay focused on day-to-day-operations.

Or, we partner closely with internal teams to execute, while fully equipping internal teams via collaboration, knowledge transfer, and upskilling.



2 **NorthBay JAM Sessions for Financial Services** are 10-day, no-cost or low-cost immersive workshops where customer teams deep dive with our AWS experts. We work hand-in-hand with customer teams to plan, architect, and develop a rapid prototype solution, enabling them to fast-forward data platform modernization initiatives.

Use-case specific JAM Sessions for Financial Services include:

- Optimizing conversion rates.

- Increasing customer retention rates.

- Increasing customer lifetime value.

- Developing new products.



Get experience with AWS technologies in immersive JAM session.

3 **NorthBay Data Platform Modernization Accelerators** are a unique and proven set of processes and best practices that are purpose-built to speed up time-to-value for financial services organizations.

With NorthBay Data Platform Accelerators, IT teams can:

- Build a data-sharing culture.

- Transform operations.

- Modernize infrastructure.

- Remove risk every step of the way.



Watch this 3-minute video to learn more about NorthBay Data Platform Accelerators.

**5** **Is "modernize in-place" a short-term fix or a longer-term solution?**

Simply said, it depends on the capabilities of your current environment, your timeframe, and your goals.

Every organization has some type of data platform in place, whether it be an excel spreadsheet, a database, or a data warehouse. But in order to take full advantage of data analytics and innovative technologies such as machine learning or generative AI, these platforms need to be modernized so they work with both structured and semi-structured data.

Rather than replacing legacy systems up front, some CIOs/CTOs opt to start with a "modernization in place" strategy, which involves augmenting existing systems with a a centralized data platform in the cloud. Some start the modernize in place process by deploying highly sensitive workloads in a private cloud behind the firewall, where they can be kept indefinitely. Other applications can be replatformed or replaced with more modern, cloud-native versions on AWS.

**6** **Which is the right approach for analytics modernization: A centralized data lake or decentralized data mesh?**

A centralized data lake is a single, centralized repository that stores all of a financial services organization's data in its original, raw format. This approach offers several benefits, including a single source of truth for all data, easier management of data quality and governance, and centralized data security and access control.

**Data mesh** is a relatively new approach to data management and analytics that is gaining popularity among financial services organizations that deal with large and complex data environments and a siloed business model. Data mesh advocates for a distributed and decentralized architecture for data management by treating data as a product and creating a mesh of self-organizing, cross-functional teams that take ownership of their own data domains and are responsible for the entire lifecycle of their data products.

By treating data as a product and creating cross-functional teams responsible for specific data domains, financial services organizations can gain a deeper understanding of customers and create more personalized experiences to increase retention.



With data mesh, product or business unit teams that take ownership of their own data domains and are responsible for the entire lifecycle of their data products.
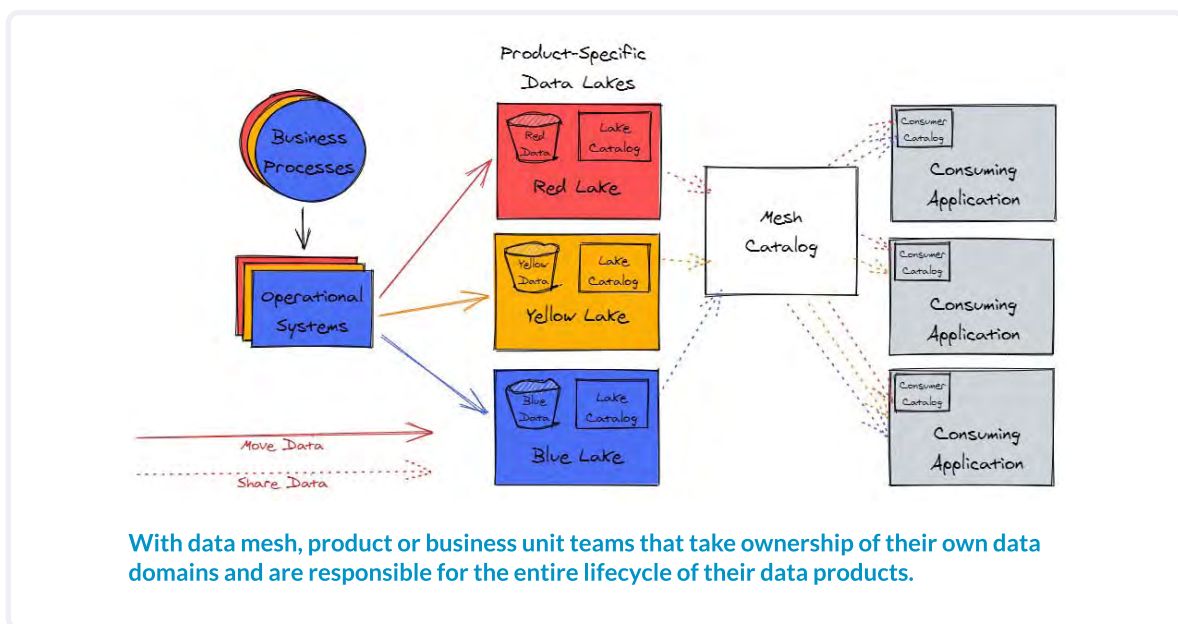
Deciding between a centralized data lake and data mesh requires careful consideration of your organization's data environment, analytics needs, governance requirements, and overall business goals. Both approaches have their own benefits, and the choice ultimately depends on what works best for your organization.

| CENTRALIZED DATA PLATFORM (physical centralization) | DECENTRALIZED DATA MESH (logical centralization) |
|---|---|
| **Data Ownership** | |
| The centralized data team owns all pipelines. | Data as a product: Each organizational domain owns their data end to end. |
| **Data Governance** | |
| • IT creates, owns, and self-governs data strategy, standards, access, quality, and security for data provided by line-of-business leaders/data producers.<br><br>• Provides a more controlled and regulated environment for data management and analytics. | • IT creates strategy, standards, governance and security measures in a data catalog. Line-of-business leaders/ data producers own and self-govern access, security, data quality.<br><br>• Leverages federated data governance to ensure data is secure, accurate, and not misused. |

| Data Access and Usage | |
|---|---|
| Centralized data team conducts analysis on behalf of stakeholders/ data consumers. | Self-service data usage by data analysts and data scientists/ data consumers, and purpose-built analytics and machine learning services. |

### 7  Why are so many financial services CTOs choosing the cloud, especially AWS , for their data platform?

AWS is a pioneer at the intersection of financial services and technology. Thousands of organizations, from established enterprises to fast-growing fintechs, are redefining their future on AWS.

Here are three reasons why:

1  **Core legacy system modernization**

Monolithic, expensive legacy systems make it difficult to gain insights and take actions quickly. With core banking system modernization solutions on AWS, banks can augment or replace these systems with applications based on open APIs. The result? Reduced costs, improved agility, and more opportunities to innovate and better serve customers.

2  **Data analytics**

By leveraging these services, organizations can store, process, analyze, and visualize large volumes of data in a secure and cost-effective manner, and use machine learning to gain insights and improve decision-making.

Data analytics services such as Amazon Redshift, Amazon Athena, Amazon QuickSight, and Amazon Kinesis can be used to analyze data in real-time, create dashboards and visualizations, and perform ad-hoc queries on large datasets.

3  **Machine learning**

Top companies such as Intuit, Vanguard, NuData, and more use Amazon SageMaker for machine learning to improve security, privacy, compliance, and governance, as well as to detect fraud, assess credit risk, and automate operational processes. SageMaker brings together a broad set of capabilities purpose-built for ML, helping financial services organizations prepare, build, train, and deploy high quality machine learning models to support regulatory and compliance mandates and exceed the highest customer expectations.

**Growcredit**

**Modernizing and Managing Grow Credit's Applications and Operations on AWS**

Learn how NorthBay helped Grow Credit develop a data lake and build pipelines to enhance data science and data analytics.

**Read the Customer Story**

**8** **Are there proven reference architectures to accelerate my modernization journey?**

With a centralized data lake, along with data services purpose-built for the data lake, organizations can act with speed and agility. Data movement has to be seamless but can't compromise on security and governance.

Data lakes leverage an architecture approach where all structured, semi-structured, and unstructured data are stored at any scale. Querying and data processing over a data lake is typically done using advanced analytics tools, languages, and techniques designed for use by data scientists and data engineers.

A data lake house is an architecture that enables querying of data across a data warehouse, data lake, and operational databases to gain faster and deeper insights that would not be possible otherwise. A data lake house on AWS uses various AWS technologies that adheres to the following five pillars:

**1** **Purpose-built analytics services**

Use the right tool for the right job to not only facilitate ease of use, but also to achieve best performance across the board.

**2** **Unified data access**

Enabling seamless data movement into the lake, out from the lake or around the perimeter. Democratizing access to not only data but advanced capabilities including AI/ML.

**3** **Unified governance**

A single place where access control and security policies are defined and managed.

**4** **Performant and cost-effective**

Right sizing compute and storage, and using the right tool and technique that suits the scenario.

The reference architecture below visualizes how to organize a lake house implementation into 5 logical layers, where each layer is composed of multiple purpose-built components that address specific requirements.

For more insights, read the blog post Taking a Look at AWS' Lake House Architecture.
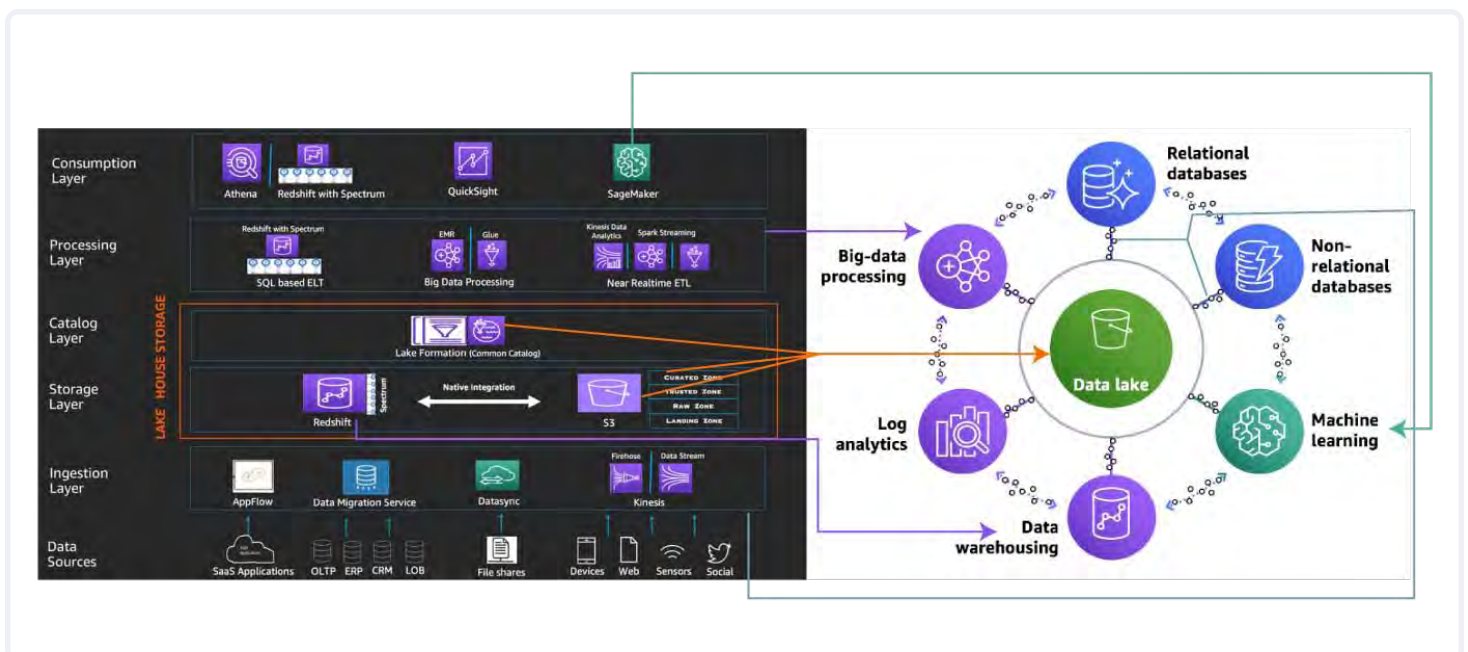
## 9  Have fintech innovators found new ways to solve the data quality conundrum?

In today's data-driven business world, ensuring data quality is critical for making accurate and informed decisions. Poor data quality can lead to incorrect decisions, wasted resources, and lost revenue. To address this issue, many financial services organizations are turning to advanced solutions such as the **NorthBay Data Quality Engine.** The Engine provides a range of benefits, including:

- Tracking and notifying schema evolution, checking and validating the number of columns and their data types, and checking the business logic of dimension table columns.

- Calculating data quality metrics such as size, completeness, compliance, mean, and uniqueness with rules configurations that are based on the latest trends in data lake technology, and enables no-code configuration of new rules.

- Providing a wide range of data cleansing and transformation tools to ensure that data is clean, formatted correctly, and of high quality.

**Data Quality Checks**

Data quality rules replace manual scripts to ensure accuracy, consistency, completeness, and validity. New rules can be configured without coding. A wide range of data quality checks help organizations to quickly find and fix data quality issues before they impact business operations. The engine also allows businesses to perform data quality checks on new datasets quickly and easily, reducing the risk of data quality issues arising.

**Data Cleansing and Transformation**

Comprehensive cleansing and transformation tools and processes ensure that all data is clean, formatted correctly, and of high quality.

- Cleansing tools include capitalizing and lowercasing text, formatting dates and phone numbers, adding suffixes, and removing characters.

- Transformation operations include changing data types, removing columns, duplicating columns, and renaming columns.

---

NorthBay Data Quality Engine Solution Brief

**Download now.**



NorthBay

How to Solve
**The Data Quality Conundrum in Financial Services**

---
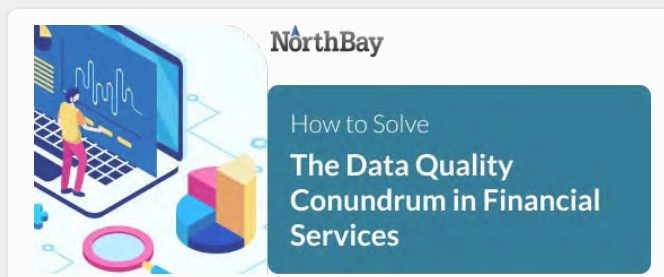
## 10  What are the top use cases for generative AI in the financial services industry?

Financial services leaders are finding new ways to leverage generative AI to overcome long-standing challenges as well as unlocking new opportunities to delight customers.

Here are four use cases where the power of generative AI shines strong:

1  **Working with unstructured data.**

   Traditional methods struggle to analyze and extract insights from unstructured data due to its complexity and lack of organization. However, generative AI models and Natural Language Processing (NLP) interact with data to extract structured information from unstructured text, images, audio, and video and identify specific pieces, such as extracting product names and their corresponding prices from customer reviews. With generative AI, financial services organizations now have a means of ingesting unstructured data into the data platform.

**2**  **Speeding up data migration.**

Generative AI techniques help to accelerate:

**Data schema**
Automatically infer the schema for unstructured or semi-structured data, reducing the manual effort required for this critical step in the data lake implementation process.

**Data cleansing**
Cleanse data by identifying and removing duplicate records or correcting invalid values, ensuring that the data lake is populated with high-quality data.

**Data normalization**
Normalize data, ensuring that it is consistently formatted and structured across all sources. This is particularly important when data is coming from multiple sources or is in a variety of formats.

**Data categorization**
Categorize data based on its content, such as by identifying the topic or domain of the data. This can help to organize data in the data lake and make it easier to find and analyze.

**Data enrichment**
Identify relationships between data points or generate additional data based on existing data. This can help to provide additional context and insights for analysis.

---

**How to leverage generative AI for a data platform on AWS**

- **Identify and collect data sources:** Gather data from various sources, such as customer data, transaction data, and website analytics data.

- **Ingest data into the data lake:** Use AWS Glue to ingest and store data in the data lake, making sure to follow best practices for data lake implementation.

- **Apply generative AI:** Use generative AI tools such as the forthcoming Amazon Bedrock and Amazon SageMaker to automatically cleanse, normalize, categorize, and enrich the data in the data lake.

- **Analyze data:** Use AWS tools such as Amazon Redshift, Amazon Athena, or Amazon EMR to analyze the data in the data lake, generating insights and making data-driven decisions.

---

**3**  **Creating Q&A systems and chatbots.**

NLP interacts with data to build Q&A systems. These systems process questions and search for relevant information in large knowledge bases to provide accurate answers. The models are trained on annotated data where questions and corresponding answers are provided.

Chatbots leverage NLP models to extract relevant information from user inputs, such as intent classification, entity recognition, and sentiment analysis. This enables the chatbot to accurately understand user requests and provide appropriate responses.

**4**  **Building and productizing large language models.**

Generative AI can be used to build large language models such as ChatGPT that are specific for financial services use cases, including customer retention. The models are specifically trained on a wide range of financial data.

For example, Bloomberg recently announced BloombergGPT, which the company built from scratch and claims outperforms similarly-sized open models on financial NLP tasks by significant margins. The company is using BloombergGPT to improve existing NLP workflows, while also imagining new ways to put this model to work to delight customers.

## About NorthBay Solutions

NorthBay Solutions is a U.S.-based organization focused on delivering technology driven business solutions. With over a decade of successfully delivering solutions around AI/ML, big data and data lakes-based analytics, DevOps, applications modernization and database migrations, NorthBay is geared to help customers solve their business challenges.

**APPLICATION & DB MIGRATION**

**APPLICATION & DATA MODERNIZATION**

**DATA LAKE AND ANALYTICS**

**MANAGED APPLICATION SERVICES**

**DEVOPS ENABLEMENT**

**CLOUD APPLICATION DEVELOPMENT**

**MACHINE LEARNING AND AI**

**10+**
YEARS AS A TRUSTED AWS PARTNER

**350+**
AWS CERTIFIED PROFESSIONALS