How The New York Stock Exchange built its real-time market data platform on AWS

by Anthony Zawadzki, Vinil Bhandari, Alex Mirarchi, Diego Soares, and Vandana Premkumar | on 02 JUL 2024 | in Amazon EC2, Amazon VPC, AWS Direct Connect, AWS PrivateLink, Financial Services, Industries | Permalink | 💬 Comments | ↱ Share

This blog discusses how The New York Stock Exchange Group (NYSE) expanded its cloud-based market data product offerings by launching NYSE Cloud Streaming. NYSE Cloud Streaming is a real-time market data streaming solution that lands high-quality market data into AWS with sub-hundred milliseconds latency in a Kafka-compatible stream. This solution makes it easier for customers to consume and integrate real-time market data with cloud-native applications.

About The New York Stock Exchange

The New York Stock Exchange Group (NYSE) operates five equities exchanges and two options exchanges. These US-based markets are key global centers of liquidity that are purpose-built to meet the needs of corporate and Exchange Traded Fund (ETF) issuers and provide investors with greater choice in how they trade. The NYSE Group owns and operates the largest US equities exchange group by market share, including the New York Stock Exchange, a symbol of capitalism at its best and the belief that free markets offer every individual the chance to benefit from success. On an average day, The NYSE Group trades over 2.5 billion shares and processes over 500 billion messages across all of its markets.

In addition to owning and operating its seven primary exchanges, NYSE offers to customers a range of data solutions across different data types and levels of granularity, including real-time and historical tick data, corporate actions, and reference data. These data offerings play a critical role in how customers develop trading strategies, make investment decisions, manage risk, and streamline their post-trade processes.

Building NYSE Cloud Streaming to Solve Customer Challenges

NYSE has been on a multi-year cloud journey to expand access to their existing market data offerings and launch new services to meet emerging customer demands. In 2021, NYSE migrated all of their non-real-time data offerings to AWS and shared the key business drivers and results of the migration in an AWS blog the following year.

Driven by the initial success of NYSE's data access and distribution cloud journey, NYSE has worked with customers to identify and build new cloud-based solutions that deliver additional value and access. An area of strong interest from both existing customers and potential new customers is a way to access easily and efficiently high-quality real-time streaming market data natively in the cloud.

Previously, consuming high-performance real-time streaming data directly from NYSE required customers to build, manage, and maintain on-premises environments or work with a third-party provider to facilitate delivery of data. This process can be time-consuming and requires an up-front capital expenditure (CAPEX) investment. It also involves ongoing maintenance and tuning, as well as capacity expansions, hardware and software upgrades

Customers often cite the time-to-deploy, upfront CAPEX and complexity of managing these environments as challenges, and in some cases, as barriers to adoption. Another challenge customers voiced was the inability to receive streaming data where they want it and in the format they want it in. Customers running cloud-native applications cited difficulties with taking on-premises multicast feeds and integrating them into their products and applications running in the cloud.

NYSE solved these challenges by building NYSE Cloud Streaming, a data distribution platform which enables customers to access quickly and consume real-time streaming exchange data in a developer-friendly format. With NYSE Cloud Streaming, customers

gain access to market data with sub-hundreds of milliseconds latency in the cloud. They can also spin up an environment in AWS within minutes without the upfront cost and logistical hurdles of building out a physical infrastructure footprint. Besides easing challenges related to infrastructure, data is published via NYSE Cloud Streaming in a Kafka-compatible stream, allowing developers to integrate more easily NYSE data in a widely used format.

NYSE launched NYSE Cloud Streaming on AWS due to its widespread adoption within capital markets globally, industry-leading technology platform and expertise, and experience in helping data providers build and scale real-time data distribution platforms in the cloud. AWS was also a natural choice for NYSE as deploying this solution in AWS provides their customers with a streamlined way to access both real-time market data and their existing historical products such as NYSE Daily TAQ in a common infrastructure environment. Anthony Zawadzki, Head of NYSE Proprietary Data, noted that seamless connectivity to NYSE's feeds and between NYSE's different data products is important for customers. Zawadzki mentioned, "A key expectation of our clients is for NYSE to continue being at the forefront of innovation and to offer solutions that make it easier for them to run their businesses. With NYSE Cloud Streaming, we aim to continue that effort by accelerating time-to-value and lowering the total cost of data ownership. Now customers' cloud-native applications and workflows have a simple way to connect to and begin using NYSE data, while simplifying and streamlining operational overhead."

One of NYSE's early decisions was what feeds to provide through NYSE Cloud Streaming. NYSE made NYSE Best Quote and Trades (BQT) the first feed available on NYSE Cloud Streaming as BQT offers a comprehensive view of Level 1 data, providing a unified view of the Best Bid & Offer and Trades for all five NYSE Group equity exchanges and the NYSE Trade Reporting Facility (TRF), and various critical summary information such as open, high, low, close prices, and consolidated volume. This provides subscribers a cost-efficient, consolidated market data feed that has lightweight network access requirements and is NYSE's most popular market data feed among FinTechs and cloud-native subscribers.

Designing and Architecting NYSE Cloud Streaming

To build NYSE Cloud Streaming, NYSE had to design a platform that offers high throughput and low latency while allowing for easy integration directly into customer applications. The first decision NYSE had to make was to deploy a multicast feed or build a publish-subscribe solution. While there is rising demand for multicast-based cloud distribution solutions, NYSE saw more customer interest for unicast streams delivered in a publish-subscribe distribution model that will most easily integrate with their display applications and tools.

This led NYSE to build a streaming data platform that is compatible with Kafka APIs, using Amazon Elastic Compute Cloud (Amazon EC2) and Redpanda's platform. This platform provides the benefits of Apache Kafka including high throughput, low latency, and most importantly, a multitude of client SDKs in virtually any programming language, making it easy for developers to get started in their language of choice.
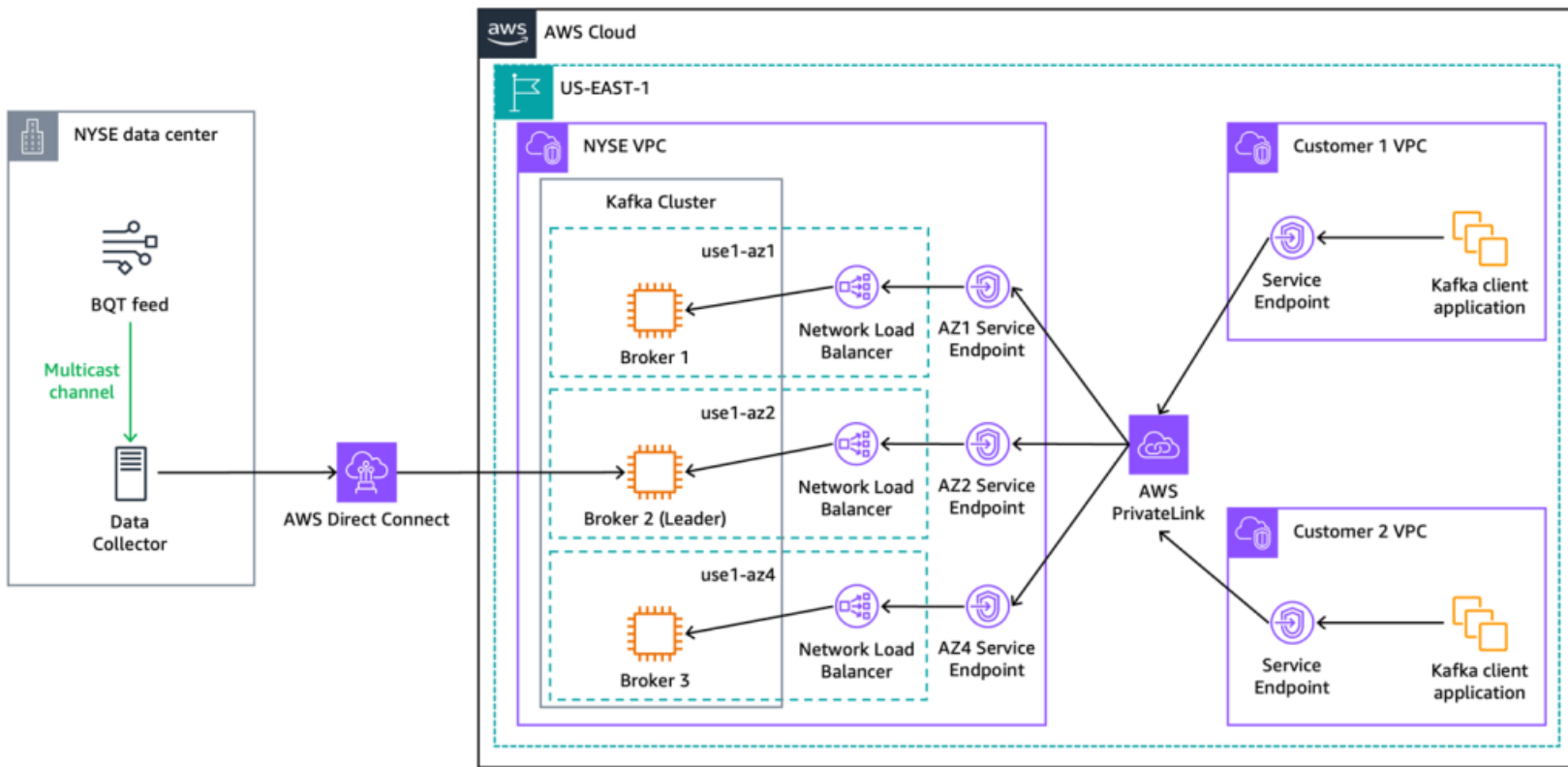
Figure 1: NYSE Cloud Streaming architecture

NYSE Cloud Streaming sources data from the original BQT multicast feed in NYSE data centers, transforms it into Protocol Buffers (Protobuf) format and publishes it into its streaming data platform over dedicated AWS Direct Connect links. NYSE chose Protobuf due to the wide programming language support, its performance in serializing structured data (think JSON, but smaller, faster, and simpler) and its encoding results in messages that are optimized for network transport to reduce data transfer costs.

Once the data lands in AWS, it is stored on the streaming data platform, which spans three Availability Zones (AZs) to guarantee data integrity and high availability. To provide customers with the lowest end-to-end latency, NYSE publishes the AZ IDs in which the platform is hosted. This allows customers to co-locate their application in these AZs to reduce total latency and data transfer costs.

NYSE built its own Kafka cluster for easier management and cost-effectiveness. Vinil Bhandari, Head of Cloud and Full Stack Engineering at NYSE, noted that, "NYSE's Cloud Streaming real-time offering will be a best-in-class Kafka implementation that is battle-tested for high throughput and low latency. We are quite excited to usher in a new era of data delivery and look forward to pushing the envelope here with more state-of-the-art technological offerings."

To achieve cross-account connectivity NYSE used AWS PrivateLink, allowing customers to consume market data privately in their AWS accounts. To connect to NYSE Cloud Streaming, customers must provide their AWS account information and receive a list of bootstrap brokers, authentication details, and the Amazon Virtual Private Cloud (Amazon VPC) endpoint service information that the customer will use to request connectivity to NYSE Cloud Streaming.

After customers establish AWS PrivateLink connectivity to NYSE Cloud Streaming, the next step is to set up their DNS infrastructure for naming resolution of the bootstrap brokers. NYSE Cloud Streaming uses TLS to encrypt the communication from its streaming platform to the receiving clients. By default, most clients, if not all, will validate the certificate during the connection handshake and will check the hostname of the server they are connecting to against the provided broker certificates.

In order for the bootstrap broker names to resolve to Amazon VPC endpoint IPs, customers will require an Amazon Route 53 Private Hosted Zone (Route53 PHZ) to resolve the broker's names to the right IP addresses in the customer subnets. Next, the

customer has to create an 'A record' in its Route53 PHZ, aliasing the entry to the VPC interface endpoint DNS name previously created. Now customers can start consuming the feed via Kafka compliant clients.

How NYSE Cloud Streaming Benefits Customers

With NYSE Cloud Streaming, customers from anywhere in the world can access real-time NYSE feeds in a matter of minutes without having to procure or manage on-premise infrastructure. In addition to accelerating time-to-market, NYSE Cloud Streaming simplifies and streamlines the data consumption process. With an AWS account, customers can access NYSE's real-time market data and take advantage of a full range of AWS services instantly to accelerate time to insight and reduce complexity.

NYSE also listened to customer feedback and decided to host NYSE Cloud Streaming initially in AWS US-EAST-1 region, which is where they host historical, corporate actions, and reference data products. Besides meeting many customers where they are currently deployed, this simplifies the process of consuming both NYSE's history and real-time data solutions.

As NYSE continues working closely with customers, they are evaluating ways to build on this first step such as potentially enabling cloud-based access to additional real-time datasets, like NYSE Pillar Depth, expanding into more AWS regions, and providing value-added offerings, such as additional access methods and analytics.

NYSE Cloud Streaming customers will benefit from:

- A faster time-to-market with no infrastructure procurement or reliance on third party infra providers, and faster times to connect to NYSE feeds.

- A scalable platform that provides data to customers located across the globe and access to potentially all NYSE Group feeds (dependent on customer demand).

- A cloud-based real-time data feed for market data delivery leveraging Kafka that uses industry standard security protocols.

- Easy integration with current applications and ability to choose your programming language of choice.

All these benefits lead to a better customer experience and a lower total cost of data ownership. Customers that want to get started with NYSE Cloud Streaming will be able to use a github repository that contains sample code to get started quickly. The code deploys all required infrastructure and a Kafka client ready to consume the BQT feed.

Comments