OKĒRA

LEADING US SECURITIES REGULATOR BUILDS AND MAINTAINS ONE OF THE WORLD'S LARGEST FINANCIAL DATA LAKES

Background

This securities regulator monitors 445 billion market events per day to detect and prevent fraud, abuse, and insider trading in the US financial markets. In the mid-2000s, they decided to migrate their data analytics infrastructure to the cloud when their on-prem solutions failed to keep up with the volume and velocity of data they were collecting. They also anticipated significant cost savings because they would not need to license, manage, and maintain a hefty on-prem infrastructure. Now they have a massive data lake of over 200 petabytes on AWS.

Because deep archiving comes at such a major drop in cost, there is a strong economic incentive to move less frequently used data there. After testing large restores and bringing back several hundred terabytes reliably – "much, much more reliably than tape" — the company now transitions many petabytes of historical data in deep storage, resulting in significant costs-saving.

Since moving to the cloud, the organization now boots up ephemeral clusters to achieve an astonishing degree of elasticity and variability adjusted for daily data volume. For certain workloads they take advantage of spot pricing, which can be 30% – and sometimes 50% – of the on-demand pricing.

Challenge

The regulator encouraged the use of best-of-breed and emerging technologies in order to perform advanced analytics. This means they often had to shuffle data between different platforms, which created significant challenges around data governance. It wasn't easy to keep track of data location or determine which data could be moved to a different storage appliance to allow the necessary performance to complete the jobs at hand.

The organization knew they wanted to separate storage and compute, which would give them the ability to dynamically scale their workloads up and down using public cloud resources. However, this also added to the complexity of managing data access entitlements. Their data was simply too big to segment into datasets curated for different user populations. IAM roles were too complex and too coarse-grained to accommodate their business requirements.

Why Okera

The regulator needed a centralized approach to managing data entitlements that would give them fine-grained data access controls across a variety of data science and analytics tools querying single source-of-truth S3 buckets.

Among the many data science and analytics approaches they use are Spark jobs running on Amazon EMR. They needed to lock down access to the underlying files, but still allow authorized users to query the data using SparkSQL. Okera nScale™ isolates access to S3 data so higher level compute frameworks (such as Spark, Hive, or Presto) running on EMR analyze data after it's been redacted, masked, tokenized, etc. And as the name suggests, nScale delivers perfectly elastic scalability, as it is co-located on the EMR nodes.

Objectives Achieved

The organization is now in the final stages of a multi-year project called Unified Data Access, with a pilot team using Spark on Amazon EMR and Domino DataLabs. Once the pilot is complete, enabling universal data authorization for 3,500 employees.

Okera's universal policies are used to provision fine-grained access control to the underlying data. Each query is dynamically authorized, such that the same S3 buckets can be queried by multiple users, using a variety of tools. With Okera nScale, they enforce data access privileges on up to 150,000 nodes a day of massively scalable, elastic cloud compute.

Okera's universal data authorization approach significantly improves the regulator's data security and data privacy posture while at the same time allowing them to deliver the analytics and insights that are fundamental to achieving their mission.

ABOUT OKERA

Okera, the Universal Data Authorization company, helps modern, data-driven enterprises accelerate innovation, minimize data security risks, and demonstrate regulatory compliance. This allows employees, customers, and partners to use data responsibly, while protecting them from inappropriately accessing data that is confidential, personally identifiable, or regulated. Okera began development in 2016 and now dynamically authorizes access to hundreds of petabytes of sensitive data for the world's most demanding F100 companies and regulatory agencies. The company is headquartered in San Francisco and is backed by Bessemer Venture Partners, ClearSky Security, and Felicis Ventures. For more information, visit www.okera.com or contact info@okera.com.