RED OAK strategic

Commodities Trading Company:
Amazon Web Services Elastic Map Reduce & Apache Spark Deployment

Project Overview:

A Commodities Trading Firm engaged Red Oak Strategic to assist with their daily pricing forecasts and historical risk analyses. The company executes commodities trades daily and sought to manage their portfolio's exposure to historical risk. The company maintains their data warehouse, in AWS Redshift, where they store historical pricing data and ingests new data from a variety of outside sources daily. The company typically executes spread trades between various classes of commodities, so having accurate pricing forecasts and dependable risk-governance is crucial to their operations and profitability.

Apache Spark Workloads on AWS Elastic Map

As an AWS Partner, Red Oak Strategic was brought into the project by Commodities Trading Firm leadership to assist in establishing a reliable and cost-effective computing infrastructure capable of processing large data workloads at speeds that aligned with their daily trading timeline. The company consumes forward-looking data round-the-clock but it has a window each day when trade price estimates must be made, and then evaluated against historical observations so that their algorithms can calibrate a set of transactions for the day and execute them.

After a discussion of their team's requirements it was determined that AWS EMR would give them the most flexible and cost-efficient compute power for their data workloads. First, AWS EMR allows for total control over EC2 instance types for cluster deployment. The company could start with small clusters and scale up as necessary. Second, AWS EMR makes it easy to set up Apache Spark on their clusters to integrate with their existing opensource code base and churn through very large datasets. There is an enormous number of trading combinations available to this firm because they can trade across the many classes of commodities and these trades must be evaluated over the most recent five-year time horizon (billions of rows). Apache Spark not only grinds through these data aggregating it into risk metrics, but it also has a fully featured machine learning platform built-in: MLlib. The company uses this ML library to make its pricing forecasts. Finally, AWS Lambda makes the scheduling of this analytics job a breeze: initiating a cluster, running the analytics workload, and then shutting down the cluster once finished. The whole system is budget conscious as the company only pays for a few hours of runtime per day.

Project Results & Client Success:

- Stabilized server infrastructure with increased uptime in line with AWS availability
- Increased data processing speeds with over 50% accuracy improvement in predictive models