Helping Marathon Oil Create a Next-Generation Cloud Native Data Platform

Based in Houston, Texas, Marathon Oil Corporation (NYSE: MRO) is an independent exploration and production (E&P) company focused on four of the most competitive resource plays in the U.S., complemented by a world-class integrated gas business in Equatorial Guinea. In 2020, Marathon Oil set out to build a next-generation data platform with the overall goal to make data more accessible and enable more data-driven business decisions.

THE CHALLENGE

To create a next-generation data platform to advance its data and analytics vision, Marathon Oil needed to start with building a solid foundation for data ingestion, transformation and consumption. The platform had to support ingesting data varying in volume, variety and velocity from many sources, provide advanced data transformation and refinement capabilities, and integrate with various business intelligence and analytics tools. To support rapid, iterative development and a low operational overhead, Marathon Oil wanted the platform to leverage serverless offerings in the cloud, infrastructure as code and automated deployments across environments. With a new data foundation in place, over 20 data pipelines and data sets needed to be migrated from an on-premise Hadoop platform to the new cloud data platform.

THE SOLUTION

Building Marathon Oil's Next-Gen Data Platform
Marathon Oil turned to EPAM, an Amazon Web Services (AWS) Advanced Consulting Partner with deep expertise in cloud advisory and application modernization, to help them design and build a next-gen cloud-based data platform on AWS and Snowflake and migrate existing data pipelines to the new platform. EPAM experts collaborated with AWS solution architects and Marathon Oil's internal teams to create a data platform with the following features:

- A serverless, consumption-based model
- Near infinite compute and storage services that can be scaled independently
- Integration with existing AD-based security model
- Integration with existing BI and analytics tools
- Infrastructure as a code and automated CI/CD pipelines

By implementing this functionality, EPAM and AWS helped Marathon Oil create a cloud native data platform built for scalability, reliability and effective data management.

Migrating Applications to the New Cloud Native Data Solution
After creating a next-gen data platform, EPAM helped Marathon Oil migrate over 20 data pipelines and associated data sets from an on-premise Hadoop-based platform to the new data platform on AWS. This process required refactoring and re-architecting existing data pipelines to be more cloud-native, as well as extensive testing that leveraged automated data quality pipelines to verify migrated data. With the migration complete, the company was able to remove recurring software licenses, reduce operational overhead and improve performance and reliability of the existing data pipelines.

THE RESULT

The new data platform that EPAM helped Marathon Oil develop delivers the performance, scalability and reliability required. With a repeatable process for adding new data ingestion pipelines, the company continues to add additional data sources. The cloud data platform is transforming Marathon Oil's business by:

- Centralizing data in a single platform and making it more accessible to business users
- Providing the flexibility and scalability to support existing and future use cases
- Removing recurring software licenses and reducing operational overhead from legacy on-premise platforms
- Providing a repeatable process for building and deploying new data pipelines

FAST FACTS

Client: Marathon Oil Corporation

Location: Houston, Texas

Industry: Oil and Gas

TECH STACK

Amazon CloudWatch

Amazon EMR

GitHub

AWS Glue

AWS Lambda

Python

Amazon Simple Storage Service (Amazon S3)

Apache Spark

AWS Step Functions

Snowflake

HEAR FROM THE CUSTOMER

"With EPAM's help, we were able to build a data platform that is truly next gen. EPAM provided the necessary expertise and personnel to get the job done, with the balance of cost, time and quality we were looking for. The platform is enabling us to make data more accessible and extract additional value from it. It is an important piece of the puzzle to becoming a more data-driven organization."

ABOUT EPAM

Since 1993, EPAM Systems, Inc. (NYSE: EPAM) has leveraged its software engineering expertise to become a leading global product development, digital platform engineering, and top digital and product design agency. As an AWS Advanced Consulting Partner, EPAM works with its global customers to design, migrate, build and support sophisticated cloud applications on AWS with increased flexibility, scalability and reliability.

CONTACT US AT SALES@EPAM.COM OR LEARN MORE AT WWW.EPAM.COM.