

DevOps

AWS DevOps for automotive data lakes

Automotive Customer

AWS DevOps for Data lakes

CUSTOMER — International automotive OEM

The client is a leading international automotive OEM with a strong online presence tailored to their global customer base. They aim to provide comprehensive and accurate information about their vehicles, enabling customers to make informed decisions regarding purchases, financing, and aftermarket services. Their platform integrates various offerings, including car sales and purchases, automotive services, accessories, and financing options, creating a one-stop shop for current and potential car owners.

CHALLENGE — Ensuring Flexibility and Scalability

The client faced several challenges in consolidating and analyzing data from a wide range of structured and unstructured sources. They needed to develop a comprehensive data lake solution capable of handling data from various legacy on-premises systems, as well as other external data sources with different API technologies and response times. The goal was to create a unified, scalable, and cost-effective platform for data storage, processing, and analysis.

Addressing these challenges required a flexible and adaptable data lake solution that could efficiently integrate and manage disparate data types from multiple sources. The solution also needed to be scalable to accommodate the growing volume and variety of data, as well as support potential expansion into new markets. Furthermore, the client sought to ensure that the data lake solution would align with their strategic objectives and security requirements.

MHP was engaged to design and implement a data lake architecture and operational model tailored to the client's specific needs. Leveraging their extensive experience with cloud technologies and their understanding of the client's business, MHP proposed a serverless data lake solution built on AWS services to address these challenges and effectively manage the diverse data sources.

SOLUTION

Rapid and Automated Deployment through Modern Methodologies

MHP assembled a team of developers and an architect to design, implement, and operate a serverless data lake solution in close collaboration with the client's business team. MHP proposed a serverless approach for its low cost and high flexibility, utilizing AWS services for ingesting, processing, and analyzing data from various on-premises and other data sources, as well as implementing monitoring, logging, and a Continuous Integration/Continuous Deployment (CI/CD) pipeline.

Data Lake Architecture: AWS Serverless Best Practices

The data lake architecture follows AWS serverless best practices, leveraging serverless services for optimal performance. The architecture features AWS Glue, AWS Lambda, Amazon S3, Amazon Athena, Amazon Redshift Spectrum, AWS Lake Formation, and AWS CloudWatch, offering scalability, adaptability, and quick response to requirements changes.

Key AWS Services:

- **AWS Glue:** A fully managed extract, transform, and load (ETL) service used to categorize, clean, and enrich data from various data sources and store it in the data lake. Data quality checks can be implemented using Glue's DataBrew profiling feature.
- **AWS Lambda:** Serverless compute service that runs the code in response to events, such as data ingestion or processing.
- **AWS Database Migration Service:** AWS DMS allows near realtime data replication into the data lake and the managed service allows simplified operations.
- **AWS Step Functions:** A visual workflow service that helps developers to orchestrate data processing.
- **Amazon S3:** Used as the storage foundation for the data lake, offering cost-effective, scalable, and durable storage for structured and unstructured data.
- **Amazon Athena:** An interactive query service that allows users to analyze data in Amazon S3 using standard SQL, enabling quick data exploration and analysis.
- **AWS Lake Formation:** Simplifies and automates the process of setting up, securing, and managing the data lake, ensuring data governance and access control.

- **AWS CloudWatch:** Used for logging, monitoring metrics, and setting up alarms, ensuring data lake performance monitoring and visibility.

Data Ingestion and Integration Process

The data lake ingests and integrates data from various on-premises and other data sources through AWS Glue and AWS Lambda. AWS Glue's ETL capabilities enable the extraction, transformation, and loading of data from disparate sources into the data lake, while AWS Lambda can be utilized for processing events or custom data processing tasks.

Key Performance Indicators (KPIs) for Data Ingestion and Integration:

- Efficiency of data ingestion from multiple sources
- Accuracy and consistency of data transformation and loading
- Scalability of the data ingestion process
- Integration with on-premises and other data sources

Data Processing and Analysis

With data ingested and integrated into the data lake, processing and analysis are facilitated through a combination of Amazon Athena, AWS Glue and AWS Lambda. Amazon Athena enables users to run ad-hoc queries on the data in Amazon S3 using standard SQL. AWS Glue and AWS Lambda can be employed for custom data processing tasks or running code in response to specific events.

Key Performance Indicators (KPIs) for Data Processing and Analysis:

- Speed and efficiency of data querying and analysis
- Scalability of data processing and analysis capabilities
- Flexibility in accommodating custom data processing tasks
- Seamless integration with the data lake storage foundation (Amazon S3)

Data Security, Governance, and Access Control

Ensuring data security, governance, and access control is crucial for any data lake solution. AWS Lake Formation streamlines and automates these processes, enabling the creation of a secure and well-managed data lake. By defining policies and permissions within AWS Lake

Formation, access to specific data can be controlled, ensuring compliance with privacy regulations and protecting sensitive information.

Key Performance Indicators (KPIs) for Data Security, Governance, and Access Control:

- Security and privacy of sensitive data
- Compliance with relevant data protection regulations
- Ease of managing data access policies and permissions
- Automation of data lake setup, security, and management tasks

Monitoring and Logging

AWS CloudWatch is employed for monitoring, logging, and setting alarms, providing visibility into the data lake's performance and ensuring prompt issue resolution. By gathering and analyzing logs and metrics, stakeholders can continuously optimize the data lake solution and identify areas for improvement.

Key Performance Indicators (KPIs) for Monitoring and Logging:

- Efficiency of issue detection and resolution
- Visibility into data lake performance and operations
- Integration with other AWS services for monitoring and logging
- Customizability of monitoring metrics and alarms

SOLUTION

DevOps Transformation for Streamlined Data Lake Management

To ensure the successful implementation and ongoing management of the serverless data lake solution, MHP incorporated a DevOps transformation as part of their strategy. This transformation focused on fostering collaboration between development and operations teams, automating processes, and continuously improving the data lake infrastructure.

Collaboration and Communication

A key aspect of the DevOps transformation was establishing effective communication and collaboration channels between developers, data engineers, operations teams, and business stakeholders. Regular meetings and synchronization sessions were scheduled to discuss progress, share feedback, and address any challenges or concerns. This approach facilitated the rapid development and deployment of new features and improvements for the data lake solution.

Automation and Continuous Integration/Continuous Deployment (CI/CD)

MHP implemented a CI/CD pipeline to automate the build, test, and deployment processes for the data lake solution. This included the use of AWS CodePipeline and AWS CodeBuild, which enabled seamless integration with automatic deployment of infrastructure updates. The CI/CD pipeline not only reduced manual intervention but also accelerated the release of new features and enhancements.

Monitoring, Feedback, and Continuous Improvement

As part of the DevOps transformation, MHP integrated monitoring and logging solutions, such as AWS CloudWatch, to provide visibility into the performance and operation of the data lake infrastructure. This enabled the team to identify areas for improvement, optimize the system, and promptly address any issues. By continuously gathering feedback from stakeholders and analyzing system performance data, MHP ensured that the data lake solution evolved in response to changing requirements and business needs.

OUTCOMES – Enhanced Flexibility and Responsiveness

Through a serverless architecture and DevOps approach, MHP developed a cost-effective, highly flexible data lake that allows for quick deployment of new features and critical fixes. The application can be easily released to new markets or regions, and with monitoring in place and well-established communication channels, all stakeholders receive essential information promptly.

Weekly meetings between business, IT, architects, and developers ensure rapid development cycles for new features, and the combination of monitoring, Infrastructure as Code (IaC), and CI/CD enables real-time feedback on the application environment. The use of AWS CodeCommit in combination with AWS CodeBuild for the CI/CD pipeline ensures seamless collaboration and automation of testing and deployment processes.

By employing a DevOps approach and leveraging AWS serverless technologies, MHP was able to deliver a flexible, cost-effective, and scalable data lake solution that integrates multiple on-premises and other data sources. Key outcomes include rapid data ingestion and integration, efficient processing and analysis, enhanced security and governance, and seamless collaboration among stakeholders.

ABOUT THE PARTNER - “ENABLING YOU TO SHAPE A BETTER TOMORROW”

Functioning as a technology and business partner, MHP digitalizes its customers’ processes and products, and guides them through IT transformations along their entire value-creation chain. MHP is a digitalization pioneer for the mobility and manufacturing sectors with expertise that can be transferred to a wide range of industries. MHP is a distinguished AWS partner, currently holding the Advanced tier status and offering Consulting as well as Software services to its customers. Additionally, MHP is a member of the APN Immersion Day program as well as AWS Well-Architected Partner and to date obtained 10 AWS Service Delivery Program Validations (SDPs), two of which MHP achieved as a launch partner.

MHP serves over 300 customers worldwide, including large corporations and innovative SMEs. MHP advises on both operational and strategic issues, offering proven IT and technology expertise as well as specific industry know-how. MHP operates internationally as OneTeam with headquarters in Germany and subsidiaries in the USA (since 2011), UK (since 2016), Romania (since 2014), and China (since 2013).

The MHP Group has been shaping the future alongside its customers for over 25 years. The MHP team of over 3,300 employees is united by the company’s promise of excellence and sustainable success. This promise continues to drive MHP – today, tomorrow, and in the future.

“MHP: DRIVEN BY EXCELLENCE.”