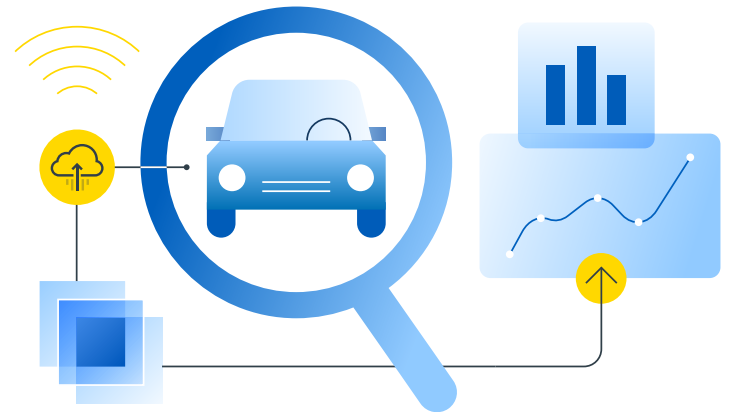# Overhauling the Cloud for a Major Automobile Retailer

Faced with several large-scale challenges, an industry-leading automotive retailer looked to CapTech to provide needed technical expertise and support. CapTech was engaged to overhaul the cloud infrastructure supporting operations of the data science and machine learning department, improve Artificial Intelligence (AI) capabilities, and implement machine learning operations (MLOps) processes to manage a growing number of AI models and capabilities.

Objective

In an effort to stay ahead of the constantly evolving retail industry, our client wanted to leverage advanced AI computer vision capabilities. These capabilities included enhancing client-facing images for increased engagement, as well as developing computer vision-based internal process control for the retailer equipment and facilities. As the company's IT initiatives grew, it became increasingly difficult for its data scientists to develop and deploy models quickly and efficiently. With various engineering teams using different templates and methods, it was difficult to manage the lifecycle of these models, leading to growing technical debt. Additionally, there were challenges in deploying reproducible models into production, which impacted maintenance and upgrades. Finally, the high cost of cloud computing was also a major factor, which affected the retailer's ability to scale AI and ML initiatives.

CapTech®

Solution

A new serverless architecture was designed for delivering computer vision and machine learning services. Reusable MLOps templates were designed and created for deploying microservices infrastructure as code and establishing common patterns to manage ML resources across several departments. Azure ML resources and services were used to handle data management, automated training, deployment, and monitoring of AI models. Finally, custom deep-learning computer vision models were trained using proprietary data to develop new capabilities and applications.

Results

At the conclusion of the engagement, the retailer saw significant benefits. The newly designed serverless Azure architecture led to an estimated monthly savings of $50,000 compared to the previous infrastructure. CapTech established an MLOps framework that greatly accelerated the data scientists' and engineers' productivity, shortening the length of deploying feature updates from weeks to days as well as improving reproducibility, performance, and cost monitoring across the various models and capabilities. The custom models were deployed in production, leading to an approximately 1.9% increase in click-through rate for items displayed on the retailer's website, as well as a 50% decrease in lost business thanks to the increased speed of detection of faulty equipment and processes.

Tools

- Azure cloud services (function apps, servicebus, event grids, cosmosDB, managed endpoints, Bicep templates)
- Python and common data science, ML and computer vision libraries
- Azure Devops
- Azure ML
- Locust
- Pytorch, Tensorflow

About CapTech

CapTech is an award-winning technology consulting firm. We're passionate about the work we do and the results we achieve. Since we believe next is now, we guide our partners to reimagine the way they do business through the power of technology. But to influence organizations—and to impact the most lives—we think as boldly as possible, whether we're completing the task at hand or keeping an eye on the horizon for opportunities that have true transformative potential. Ultimately, we view every client engagement as an opportunity to start a new relationship—or to deepen an existing one.

Let's do next together.

in ◎ **@captech_consulting** ✖ **@CapTech_Tweets**

CapTech®