USCIS maximizes cost-effective use of the cloud with robotics cloud automation

Operation Cloud Control helps USCIS make the most of the cloud

U.S. Citizenship and Immigration Services (USCIS) oversees lawful immigration to the United States and became an early cloud adopter among federal agencies. In 2014, the agency began moving some workloads out of two legacy, on-premises data centers and into the cloud with Amazon Web Services (AWS) to overcome delays in infrastructure delivery and speed application development.

Early on, much of the agency's cloud movement was lift and shift—simply moving workloads as-is to the cloud. Over time, IT staff trained themselves on cloud operations and began to take greater advantage of the flexibility and scalability that cloud computing offers.

Three years later, USCIS began adoption of microservices, an emerging architecture in which each application function is built and deployed independently. With microservices, development teams can rapidly build new components of apps to meet changing mission requirements.

However, once the agency deployed more than 20 microservices, IT leaders realized they could not efficiently orchestrate the environment. So, they began migrating microservices that powered the agency's Electronic Immigration System (ELIS) for case management onto a Kubernetes platform within AWS. The migration enabled USCIS to automate many of the manual processes involved in deploying, managing, and scaling the system.

Soon after, agency leadership tasked the Office of Information Technology (OIT) to digitize all of the agency's immigration benefits processes. Around that time, IT leaders realized that they needed stronger governance and standards around cloud. They moved to domain-driven design, which allowed IT to work across the agency's directorates.

"Domain-driven design enabled us to work across the IT architecture, teams, and communications channels," said Rob Brown, USCIS chief technology officer. "We began to understand what development teams were doing across the agency. As a result, we began to consolidate teams, platforms, and toolsets, taking advantage of opportunities to reuse capabilities instead of buying more."

Cloud sprawl becomes a management and fiscal challenge

USCIS's early and open embrace of cloud computing brought great benefits to IT operations but also a sizeable management challenge, Brown observed.

"Our wide-open aperture for cloud services provided so much learning, adoption, and innovation," he said. "It also brought cloud sprawl: a proliferation of cloud resources, some of which were underutilized or abandoned over time. Now, we are working toward a balance, providing autonomy for teams as long as they operate within our governance guardrails."

The challenges associated with cloud sprawl compounded in spring 2020, when the agency curtailed immigration services because of the COVID-19 pandemic. Unlike most other federal agencies, USCIS self-funds through fees collected for the provision of immigration and citizenship benefits, not through congressional appropriations. With USCIS offices closed because of the pandemic, fee collection dropped precipitously, and agency units worked to employ cost-saving measures.

At OIT, leaders developed a cost-management strategy, as well as a 90-day cost control project to realize immediate savings and establish processes to lock in those savings moving forward.

The project, named Operation Cloud Control (OCC), began with a baseline assessment of cloud operations. OIT educated staff about the need to right size cloud instances, and it created dashboards to measure progress, which IT teams as well as executive management could access.

"This was not a back-room project," Brown noted. "This was first light. Everybody was looking at our reporting on a weekly basis."

OIT began to redesign applications to operate more effectively in the cloud, and it rolled out design-cost principles and policies. Robotic Cloud Automation (RCA), a library of serverless cloud automation solutions developed by Simple Technology Solutions (STS) that leverages AWS's native tagging capabilities and AWS Lambda scripts, enforces these policies. The library is a one-time cost to USCIS, rather than a recurring expense.

RCA automatically identifies cloud sprawl using those tags and governance as code. Then, in contrast to other solutions, it remediates cloud instances, environments, and resources that are over-provisioned, over-scheduled, or not compliant with the agency's usage standards.

"We wanted to make sure that the autonomous bots reinforce the policy, which ensures good cloud hygiene," said Aaron Kilinski, principal and chief technologist at STS.

USCIS uses RCA to:

- **Delete orphaned resources**. If the owner does not delete unused cloud resources within a specified period, they automatically shut down.
- **Schedule services**. If the owner does not put cloud resources on an allow list so they can run during non-business hours, they automatically turn off at night and on weekends.
- **Manage the data lifecycle**. RCA applies lifecycle policies to AWS Simple Storage Service (Amazon S3) buckets and Amazon Elastic Block Store (Amazon EBS) Snapshots to reduce data storage costs
- **Move to modern infrastructure**. Workloads running on outdated servers are moved to current AWS technology, giving the agency greater bang for its buck
- **Eliminate overprovisioning**. RCA rightsizes workloads and employs autoscaling to accommodate spikes in demand
- **Reduce on-demand cloud use**. Wherever possible, the agency moves away from higher-cost, on-demand cloud use to reserved instances, which can save more than 70 percent

Moving to reserved instances alone saved more than $2.5 million during 2020-2021. In total, the OCC project saved nearly $4 million during the same time frame.

> "When we had one or two Amazon accounts for the ELIS program in 2014, management was feasible with a few people," Kilinski noted. "But today, with more than 100 AWS accounts and hundreds of services in every account, it's impossible to do without automation. Recognizing that is part of enterprise cloud maturation."