Accelerating university AI research in the cloud

Introduction

When Emory University announced plans to dramatically expand its artificial intelligence (AI) research efforts through its AI.Humanity initiative, university leaders recognized the need to increase technological capacity to support researchers' needs.

The university had initiated a cloud-first strategy in 2016, leveraging Amazon Web Services (AWS) to give faculty, students, and researchers access to cloud services, including storage, compute, databases, and hosting. Today, around 1,000 faculty and staff have used the school's cloud platform, known as AWS at Emory, for everything from data analytics to web hosting.

But the coming influx of AI research demanded more flexible computing power and user-friendly access than the cloud platform or the university's on-premises infrastructure could provide. So Emory's IT staff again looked to the cloud for extra computing power. "We thought we could fulfill a lot of research activities on the cloud," says Circe Tsui, Emory's associate director of solutions architecture.

AI research needs

AI.Humanity aims to "bring together the full intellectual power of Emory University to shape the AI revolution to better human health, generate economic value, and promote social justice."[1] While Emory has a strong biomedical focus, the initiative involves a range of disciplines — from business and law to arts and sciences — as well as AI ethics.

To that end, Emory plans to hire up to 60 new faculty members over the next few years.[2] The university also launched its Center for Artificial Intelligence Learning in fall 2023, with the goal of promoting AI literacy across campus.[3]

AI research requires more computational power and flexibility than Emory could initially offer through centralized infrastructure.

"In partnership with AWS, we wondered if we could build a high performance computing (HPC) environment in the cloud that would be familiar to researchers and, ideally, provide a seamless interface whether the workloads were on premises or in the cloud," says John Ellis, interim enterprise CIO and senior vice provost for IT.

Compute in the cloud

Within six weeks, Emory's IT staff worked with AWS to create a proof-of-concept HPC cluster in the cloud. The IT team then recruited three research groups to test the cluster, which was designed with the same interface as the on-premises HPC.

"The test groups did some very impactful research, and in some ways they didn't even know they were using AWS," Tsui says.

The full shared cloud HPC cluster launched in September 2023. Migrating the existing on-premises interface to the cloud proved critical to adoption among researchers, according to Tsui.

"The researchers didn't have to learn something new," Tsui says. "On the back end, it did not matter if they were in the cloud or on premises. In fact, the

"The test groups did some very impactful research, and in some ways they didn't even know they were using Amazon Web Services."

Circe Tsui, Associate Director of Solutions Architecture, Emory University

researchers were able to complete some important projects that stalled due to the infrastructure limitations of existing on-premises servers."

The HPC cluster benefits from the cloud security and controls put in place by Emory engineers. AWS CloudFormation tools allow for a streamlined, DevOps-style method to deploy compute and create tailored instances on the cluster that can easily adapt to changing needs. CloudFormation templates can also serve as blueprints for launching additional clusters for specific research use cases.

Cloud-based research results

Emory researchers have used the power of cloud-based computing for some of their most complex projects. For example, one researcher, aiming to develop a program to enhance laryngoscopy images produced by low-quality medical equipment, trained a machine learning (ML) model with 21,000 images.

"Before the pilot, we attempted to scale the super resolution training on our on-premises servers with 16 GPUs and 48 GB of RAM; however, the training continued to be slow," says Judy Gichoya, associate professor in the Department of Radiology and Imaging Sciences at Emory School of Medicine. "With the A100s reserved on Emory AWS, we were able to train the models successfully across multiple architecture types. This project is serving as a foundation for edge deployment of mobile laryngoscopy in limited resource settings, and has informed us on how we can scale our work in radiology AI."

"With the HPC cluster in the cloud, we had eight high-end GPUs running to complete the machine learning in time to submit the findings at an imaging conference," Tsui says.

Other researchers have used ML to study biases in medical imaging. One team is developing a program that dynamically tailors treatments for Alzheimer's patients. Another is studying birdsong to understand how brain functions relate to motor activity.

Since much of Emory's research is closely linked with healthcare, the benefits of increased speed and computational capabilities are far from just academic.

"It used to take one researcher three days to sequence a genome," says Louis Leclerc, AWS enterprise account manager. "Now it takes two to three hours. That difference means being able to start someone's cancer treatment three days earlier."

"It used to take one researcher three days to sequence a genome. Now it takes two to three hours."

Louis Leclerc, Enterprise Account Manager, Amazon Web Services

Communicating cloud advantages

Emory is currently strengthening the data transfer pipeline between the university and the cloud. Tsui credits the university's cloud-first strategy as a key to success. But she also stresses the importance of listening to researchers.

"There are many good reasons to use the cloud — the elasticity, knowing the latest technologies are always available, the great algorithms," Tsui says. "But to drive adoption, you need to hear the researchers' needs. Learn about the computing power they require and their preferred timelines. Then match the technology to their needs. Our researchers recognize this value, and have repeatedly come back with expanded use cases like the recent Emory Health AI Bias Datathon that hosted four datasets to 130 researchers representing 69 institutions around the world."

This piece was written and produced by the Center for Digital Education Content Studio, with information and input from AWS.

Endnotes:
1. https://aihumanity.emory.edu/
2. https://news.emory.edu/stories/2022/11/er_ai.humanity_initiative_17-11-2022/story.html
3. https://news.emory.edu/stories/2023/03/er_center_ai_learning_01-03-2023/story.html

Produced by:

CENTER FOR DIGITAL EDUCATION

The Center for Digital Education is a national research and advisory institute specializing in K-12 and higher education technology trends, policy and funding. The Center provides education and industry leaders with decision support and actionable insight to help effectively incorporate new technologies in the 21st century.

www.centerdigitaled.com

Sponsored by:

aws

AWS is on a mission to accelerate the digital transformation of education through our highly scalable, secure, and cost-effective cloud solutions. We're committed to making education always available, personal, and lifelong for everyone. Contact us to learn how AWS can help you with your biggest IT challenges.

aws.amazon.com/government-education/contact

PAGE 1 AND 2 IMAGE: ADOBESTOCK.COM
PAGE 3 IMAGE: AILEARNING.EMORY.EDU