Accelerating Biological Research with Multicloud Kubernetes

Biologists mine massive data sets in their quest for discovery. And with help from Cisco Container Platform they're solving nature's mysteries faster than ever.

"Bioinformatics" is exactly what it sounds like—a discipline that applies today's most advanced computer technologies and data science expertise to some of the world's most daunting biological riddles in an effort to accelerate research.

The team that specializes in bioinformatics at Clemson University is a perfect example. They're using their expertise to dive into autism research, plant genomics, and even cancer research. They're making tremendous strides and contributing to solutions that will ultimately help stabilize our food supplies, help children with autism thrive, and, yes—even push toward the cures for some kinds of cancer.

Clemson University

Industry
Education

Location
Clemson, SC

Number of employees
2,681

Website
Clemson.edu

Data: The answer and the obstacle

The big challenge—regardless of which of these projects they're working on—is the data. "The data sets we use are huge. We analyze DNA sequence data sets, which could be gigabytes to terabytes in size individually, so we have to process those on supercomputers. We use a campus cluster at Clemson, which is a Top 500 Supercomputer, but we also have to go into the public cloud now to be able to process our data," said Professor of Genetics and Biochemistry, Alex Feltus.

In order to expand their processing to other platforms using a common compute infrastructure, they have containerized their workflows. "It's really hard to reproduce these huge experiments because the configuration needed is very exact," explained graduate student Cole McKnight. "Using containers allows you to contain something, and then reproduce it and move it to other labs and platforms." The team needed a solution that would allow them to deploy container clusters to multiple public clouds and their on-premises cloud quickly and easily so they could analyze these massive data sets and move forward with their research.

And that is where Cisco Container Platform (CCP) came in.

Simpler = faster

The team began using CCP to deploy Kubernetes clusters and quickly realized that it had the potential to accelerate their research by simplifying the cluster deployment and management process. "The value Cisco Container Platform gives me is it allows me to manage multiple clusters from different resource providers all under one platform, so instead of having to learn multiple different platforms and interact with each of those, I can just interact with Cisco Container Platform," said McKnight.

Feltus adds, "As a researcher I don't want to spend six months gearing up to learn how to use the system. This is a way for us to learn one system, and be able to run jobs across one or more clouds."

Real-world impacts

When asked for an example of how better technology can have a concrete impact in their world, Feltus and his team cite the case of Bill Paseman, a kidney cancer patient and founder of rarekidneycancer.org.

Paseman was diagnosed with papillary renal cell carcinoma, type 2, five years ago, which is rare enough that there were no known drug treatments for it at the time.

A successful Silicon Valley entrepreneur, Paseman responded as you might expect someone from the innovation capital of the world to respond: he had his tumor and his DNA sequenced so that researchers such as Alex and his team could see how his genes were interacting with one another and ultimately try to pick out the genes that make a cell cancerous versus noncancerous.

In fact, they traveled to Silicon Valley last year to work with other companies and universities at a hackathon where they collectively worked to understand how his tumor had mutated during its progression and ultimately find new drug targets for him.

The teams were able to derive valuable insights from this process that can help with treatment options for this particular kind of cancer.

The future of research

As scientists dive even deeper into the secrets our DNA holds, the volume of data they generate in the search for cures will continue to grow. “Every biolab will soon generate terabytes of data and maybe 10 to 20 years from now exabytes of data,” said Feltus.

“People can barely process things at the giga scale right now, or tera scale. It’s really important that the computer systems match what the biologist needs so they can do their work. So we’re really excited working with Cisco to be able to pull in massive amounts of data, stuff that doesn’t even fit on our campus cluster, put it out into the cloud, and do all the data-intensive experiments.”

Graduate student Reed Bender concurred, adding “As an end user primarily interested in biology and playing more catch-up with the computer science side of it, Cisco’s platform has been very easy and very intuitive to work with, so I can focus more on biology and less on computer science.”

Featured technology: Cisco Container Platform

Turn-key container management software for multicloud, consistent production-grade Kubernetes

- Runs on any infrastructure* as a lightweight self-hosted software (optimized for Cisco HyperFlex™ and Cisco UCS®)
- Automates the installation and deployment of self-service, 100% upstream Kubernetes clusters
- Includes all the necessary networking, storage, logging/monitoring, load balancing and registry tooling
- Integrates natively with Elastic Kubernetes Service (EKS), Azure Kubernetes Service (AKS), and Google Kubernetes Engine (GKE) (coming soon!)
- Optimized for AI/ML workloads with multi-GPU support
- Built for the enterprise with hardened security and enhanced availability features like multi-master and self-healing

* Deployed on top of VMware vSphere, OpenStack/CVIM, bare metal (coming soon)

C36-743109-00 11/19