Hewlett Packard Enterprise | AMD

Accelerating life-changing research through open science

How the GP-ARGO project is speeding up research that's vital to ensuring food security

Tackling some of the world's most pressing problems often relies on complex calculations or the analysis of large volumes of data. But many underserved academic institutions in the U.S. don't have the computing resources to complete this work. The Great Plains Network's GP-ARGO project is changing this by giving 58 academic institutions access to distributed high-performance computing. This is empowering researchers to make important scientific discoveries, from methods for increasing crop production amid climate change to creating a better understanding of fundamental physics, helping address food security issues and other challenges.

Enabling computational and data-intensive research

Sharing resources and data can accelerate life-changing research, as the global response to the COVID-19 crisis has shown. For example, by crowdsourcing computing power from more than a million machines, the distributed computing project Folding@home sped up the development of COVID-19 antivirals.

Inspired by these successes, research and academic institutions across the U.S. are increasingly mobilizing and contributing computing power to large-scale projects through initiatives such as the Open Science Grid (OSG).

The OSG is a partnership between institutions that facilitates access to computing and storage resources to support data-intensive projects like Folding@home. To bring these resources closer to smaller universities and colleges across the Midwest and Great Plains, a consortium of institutions and states called the Great Plains Network has created a gateway to the OSG and built high-performance computing (HPC) systems that are shared across the region.

GPN
GREAT PLAINS NETWORK

Industry: Higher education

Country: United States

Vision

Enable researchers in the Midwest and Great Plains to conduct large-scale research and make important scientific discoveries

Strategy

Make distributed and powerful computing capacity accessible to underserved academic institutions in the region

Outcomes

- Enables 58 academic institutions to make important discoveries using HPC technology
- Expands research scope through machine learning technologies
- Supports the region's agriculture by helping crops and animals adapt to warming climate
- Helps address global food security problems by better predicting crop yields

The network has been providing connectivity to member universities for many years. Now its Great Plains Augmented Regional Gateway to the Open Science Grid (GP-ARGO) is powering research that's vital to the region. Studies in ecological genomics, for example, make sure plants and crops can adapt to the region's warming climate and the diseases that come with that change. Similar projects improve predictions of crop yields, helping address global food security problems.

By putting OSG nodes in underserved institutions, GP-ARGO has increased the computing power of local researchers and enabled them to collaborate and find answers to some of the problems affecting the Midwest and Great Plains.

"It has also given them the opportunity to train students to learn how to manage and use these computing resources, explain how the systems work, and help the science that's being researched," says GP-ARGO principal investigator Daniel Andresen.

Advancing research that's important to the region

GP-ARGO has supported 58 academic institutions and 90 OSG projects in fields as diverse as bioinformatics, biological sciences, astrophysics, and chemistry. It has had a particular focus on studies relating to plants, crops, and livestock, given the region's large agriculture industry. As climate change continues to affect how plants grow, these studies are becoming even more critical to the region and beyond.

"I'm seeing a lot of work in genetics and high-throughput phenotyping, especially relating to crops like wheat, milo, and corn," says Andresen, who is director of Kansas State University's Institute for Computational Research in Engineering and Science. "Researchers are trying to make sure plants are adaptive to climate change and drought and illnesses."

One of these researchers is Dr. Stephen Welch, professor of agronomy at Kansas State University. Using a computer model that combines genetics and field monitoring technology, Welch's research is creating more accurate methods for predicting the yield of wheat varieties and fields. His team has developed a radar that can measure wheat canopy height in a speedy and nondestructive way.

Measuring wheat height is vital to understanding crop yield, biomass, and resistance to disease. For the region's wheat-producing states of Kansas, Oklahoma, North Dakota, South Dakota, and Texas, it can increase crop production by enabling farmers to use new and better varieties of wheat. Globally, improved crop yields can help feed a growing population that's forecast to double by 2050.

"Dr. Welch's research requires significant amounts of data, so it benefits from our HPC systems' high bandwidth and support for parallel processing," says Andresen. "The project is a multistate, regional collaboration powered in part by our HPE ProLiant equipment because it really needs the amount of computational capacity that systems like this provide."

"It has very pragmatic outputs in terms of ensuring the supply of wheat and, ultimately, helping solve the world's food issues."

"Researchers are trying to make sure plants are adaptive to climate change and drought and illnesses. It has very pragmatic outputs in terms of ensuring the supply of wheat, and ultimately, helping solve the world's food issues."

– **Daniel Andresen,** Director, Institute for Computational Research in Engineering and Science, Kansas State University

Making large-scale analysis possible

Researchers are also investigating animal health as the region's climate continues to warm. "The hotter the temperature, the more prone to diseases animals tend to be, especially cows, pigs, and chickens," says Andresen. "Kansas cows outnumber humans by a very substantial margin, so trying to keep them healthy is a really big deal."

He expects more research related to agriculture and agronomy—or the science of soil management and crop production—to be rolled out now that specialized continuous integration (CI) resources are more accessible in the region.

"It's in those fields where you're going to see a lot of growth, particularly because they're being driven by large-scale data. The GP-ARGO nodes are big enough and the networks are fast enough that we can actually do this type of analysis."

A better understanding of how the universe works

In more traditional fields of research, OSG computing resources have been powering diverse projects in physics, including those of researcher Tinghua Chen from Wichita State University. Chen's research further explores the properties of the Higgs boson, a particle that's a building block of the universe. In particular, he has done a comparative study of the electroweak Higgs boson at future colliders, using the OSG for his Monte Carlo simulations.

These simulations are pivotal to testing and refining the current understanding of physics by generating predictions and comparing them with data from experiments. By studying the Higgs boson, Chen and others are gaining a deeper understanding of how fundamental particles and forces interact in the universe.

Probing the COVID-19 family tree

At Oral Roberts University in Tulsa, researchers led by computer science professor Stephen Wheat are investigating the evolution of the virus behind COVID-19 and its genetic family tree. They are analyzing half a million SARS-CoV-2 RNA samples from patients around the world, potentially contributing to the development of therapeutic strategies and the prevention of infectious diseases in the future.

"We're looking at the micro-mutations of SARS-CoV-2 that happened over a period of a year and a half, and to look at what mutations were regionally oriented and which ones were replicated in other spaces, and what's the dimensionality of the family tree," says Wheat, a GP-ARGO co-principal investigator.

The project will run calculations on OSG computing facilities via GP-ARGO. These types of applications require vast amounts of processing power, so high core count CPUs coupled with GPUs help run these calculations in parallel.

"It's interesting in the sense that we will compare each sample to all of them," adds Wheat. "So that's a 400,000 squared calculation. And that involves a lot of small jobs and a large number of small files that we can gather and run on the OSG."

> These fields are driven by large-scale data. The GP-ARGO nodes are big enough and the networks are fast enough that we can do this type of analysis."
>
> – **Daniel Andresen,** Director, Institute for Computational Research in Engineering and Science, Kansas State University

Expanding researchers' computing resources

Through GP-ARGO, the OSG provides a gateway for researchers and institutions to take advantage of the 30,000 cores that are currently sitting behind firewalls across the U.S. This and GP-ARGO's own 2000 or so cores are supporting computations for large-scale research that involves multiple disciplines and academic institutions. This is enabling researchers across the Midwest and Great Plains to collaborate and find answers to some of the problems affecting the region.

"Maximizing the number of cores involved perceptibly improves response times for users," says Andresen. "It's going to dramatically expand the resources available to our local researchers."

Reducing backlogs, better supporting scientists

GP-ARGO has installed 18 distributed HPC systems across the Midwest and Great Plains. Each system comprises an HPE ProLiant DL385 Gen10 Plus v2 server with AMD EPYC™ 7713 processors and an NVIDIA® A100 Tensor Core GPU. GP-ARGO chose the AMD processors for providing the needed compute power without requiring additional cooling systems, helping ensure high energy efficiency.

"We have run EPYC processors on our local cluster for several years and have been impressed by the value and overall performance, both on benchmarks and user applications. EPYC provided more cores per node, which allows us to service more OSG jobs per unit time than the competition," says Andresen.

With GP-ARGO's current HPC systems almost always running a significant backlog, "adding computational resources from EPYC has allowed us to reduce backlogs and better support scientists," he adds.

Having HPC systems in place is more important than ever as science becomes increasingly digital and driven by data analysis using machine learning. The compute power provides the bandwidth to gain insights faster, freeing up resources for other analyses.

"A lot of smaller institutions are relatively rural," notes Andresen. "Many of them have labs where they do experiments with plants, animals, or things like that. But that's just the beginning rather than the end, where you say, 'I have my experiments with plants, I write up the data, and I'm done.' Now, it's more like, 'Great, I'm submitting this to machine learning and going on from there to do a deeper level of analysis.'"

> We're looking at the micro-mutations of SARS-CoV-2 that happened over a period of a year and a half, and to look at what mutations were regionally oriented and which ones were replicated in other spaces, and what's the dimensionality of the [virus's] family tree."

– **Stephen Wheat**, Computer Science Professor, Oral Roberts University

Supporting complex calculations and discoveries

The combination of numerous cores, powerful GPUs, and the large memory nodes of GP-ARGO's computing systems has made diverse types of applications available to researchers around the region. This is enabling them to perform more complex calculations. More researchers are also now running machine learning workloads, enabling them to rapidly make sense of massive amounts of data.

"Each of our nodes has an A100 GPU in it to support much, much faster AI operations," says Andresen. "Previously, the number of researchers with an A100 sitting on their desktop was near zero because these are expensive resources. And so being able to get out and use these really expands the scope of what they can do."

Combined with the exceptional performance of two AMD EPYC 7713 CPUs with 64 cores, 128 threads, and 256 MB L3 cache, GP-ARGO's computing systems are gaining faster time to results.

"Our cores from AMD really support CPU-intensive applications. And with HPE ProLiant servers, we've made sure our machines can manage that CPU-GPU combination and are powerful enough with large capacity to meet the current and future needs of our researchers."

Solution

Hardware

- HPE ProLiant DL385 Gen10 Plus v2 servers
- AMD EPYC CPUs
- NVIDIA A100 Tensor Core GPUs

Key partners

- AMD

Explore more

→ **Discover** more Digital Game Changer stories

→ **Learn** about HPE Compute

hpe.com/partners/amd

Chat Email Call Updates

Hewlett Packard Enterprise

a50007232ENW