



Summary

Customer Rutherford Appleton Laboratory (RAL)

Location Chilton, Oxfordshire, England

Scope One of the United Kingdom's principal national research labs, RAL serves 10,000 researchers every year.

Industry Government and Academic Research

Challenge RAL's disparate legacy storage systems caused performance bottlenecks and administrative headaches that restricted the modeling efforts of the climate research team.

Solution The team consolidated their storage onto Panasas ActiveStor® and gained the scalability, performance, and manageability that their workflows demanded.

Results After the deployment, researchers no longer had to worry about managing data movement and were free to focus on their work instead.

Accelerating Global Climate Research at Rutherford Appleton Laboratory

BIG DATA FUELS BIG DISCOVERIES AT RAL

Rutherford Appleton Laboratory (RAL) is home to the Science and Technology Facilities Council (STFC), an organization that works to keep the United Kingdom at the forefront of international science. The council revolves around a singular vision: Understanding the universe from the largest astronomical scales to the tiniest specks of matter.

Following that vision, the work taking place at RAL spans a variety of disciplines, including particle physics, computational science, astronomy, energy, and climate and atmospheric science. Each year, about 10,000 scientists and engineers use the lab's high-performance computing (HPC) facilities to conduct research that seeks to answer some of the world's biggest questions.

The climate research team at RAL is one such group. Led by Dr. Bryan Lawrence, Professor of Weather and Climate Computing at the University of Reading and Director of Models

and Data at the National Centre for Atmospheric Science (NCAS), the team engages in cutting-edge climate and atmospheric science and supports both national and international communities in their climate data analyses.

One of the team's major applications is climate modeling. These computer models are made up of mathematical equations that use thousands of data points to simulate the transfer of energy that occurs within the Earth's atmosphere, land, and oceans. The more data collected, the more accurate the model. Researchers leverage the known climate patterns of the past



“A lot of the data we have from satellites and numerical models is absolutely vast, and you have to be able to keep that data long term. So, you need robust hardware, and you need volume.”

Dr. Sam Pepler

Data Curation Manager at the Science and Technology Facilities Council

to measure a model's validity before applying that model to future predictions. The results allow them to anticipate environmental impacts ranging from rises in sea levels to a specific region's drought risk.

These models are some of the best tools we have for addressing climate change threats today. They are also extremely data-intensive – one typical global model contains enough code to fill 18,000 pages of printed text. Sam Pepler, the British Atmospheric Data Centre Manager at NCAS and STFC Curation Manager, characterized the storage requirements of these workloads primarily in terms of capacity, scalability, and reliability: “A lot of the data we have from satellites and numerical models is absolutely vast, and you have to be able to keep that data long term. So, you need robust hardware, and you need volume.”

Upgrading Storage to Enhance Climate Modeling

While the climate scientists at RAL were eager to expand their modeling efforts, the lab's traditional NAS infrastructure left much to be desired. Anticipated data growth in this area meant accompanying surges in capacity and performance needs that their legacy storage systems simply could not support.

Dr. Pepler reported, “We had a difficult management job where we had an enormous number of NAS storage boxes and we had to control what was on each one individually. It was very difficult; you can't

suddenly grow your NAS box to a huge size.”

When the U.K.'s Department of Business, Innovation and Skills (BIS) supplied RAL with a £145 million e-infrastructure grant to invest in high-performance storage, the team seized the opportunity to eliminate their performance bottlenecks and administrative headaches. They could now consolidate their disparate storage systems onto a flexible and easy-to-manage scale-out NAS environment, and they could implement a parallel file system that would grant them the features they needed to expand their climate modeling workflows.

A narrow deployment window posed a challenge: The project's funding stipulated that all equipment must arrive within two months of the placed order, and the entire storage system needed to be online and fully operational by the end of the third month.

To meet the deadline and select the most suitable platform, the team enlisted the help of Dr. Peter Oliver at RAL's Scientific Computing Technology Group. Dr. Oliver advised them on the best HPC storage technologies available and sought out to find a low cost per terabyte solution that offered extreme scalability along with ease of integration and management.

The team evaluated Panasas ActiveStor® storage against competing hardware systems running GPFS

“The Panasas solution gives us powerful HPC capabilities to help leverage our massive datasets to advance essential scientific discovery.”

Dr. Bryan Lawrence

Director of Models and Data at the National Centre for Atmospheric Science

and Lustre. After a thorough analysis, they chose Panasas.

Panasas Delivers a Simple yet Sophisticated Storage Solution

Not only did the Panasas data platform provide seamless linear scalability in a fully integrated solution, but it also outshone the alternatives in cost, time to deploy, and ease of management. The fact that Panasas also delivered twice the required performance was an added bonus.

In total, RAL's ActiveStor installation represented one of the largest HPC storage deployments in Great Britain. The massive ActiveStor deployment comprised facilities distributed across four sites, introducing a new shared-resource approach to research for the participating organizations. Almost 8.5PB of ActiveStor storage was deployed across three NCAS research sites: 6.6PB at RAL, 720TB at the University of Reading, 180TB at the University of Leeds, and 900TB at a fourth site, the International Space and Innovation Centre (ISIC). In 2022, RAL purchased an additional 1.3PB of Panasas storage.

When asked about how consolidating onto Panasas will address their workload needs, Dr. Lawrence responded, “The big change in the last few years is that we’ve become a petascale environment. We’ve gone from lots of NAS and small devices, to consolidating onto a Panasas environment, which is going to cut our total cost of ownership and the number of people that have had to spend enormous amounts of time moving data. We can get our world-class data scientists back to working on the data, rather than moving the bits and bytes around.”

Capacity and performance planning, mount point management, and data load balancing across multiple pools of storage are all common administration problems that Panasas easily solves. Reflecting on the decision, Dr. Lawrence confirmed that “Panasas remains resilient even at scale, and the direct and parallel access to the storage pool means that we can work on our most complex simulations, unaffected by the system bottlenecks of our previous equipment. The Panasas solution gives us powerful HPC capabilities to help leverage our massive datasets to advance essential scientific discovery.”