DELL Technologies

intel

Accelerating Scientific Discovery

Researchers capitalize on the UK's largest academic supercomputer with the world's fastest HPC storage system to power data-driven science.

UNIVERSITY OF CAMBRIDGE

Research | United Kingdom

Business needs

The University of Cambridge needed a high performance computing and data storage system to help solve some of today's most demanding data-driven simulation and artificial intelligence challenges.

Solutions at a glance

- Dell EMC PowerEdge servers
- Intel® Xeon® processors
- Intel® Omni-Path Architecture and Mellanox® HDR100
- OpenStack and open source software

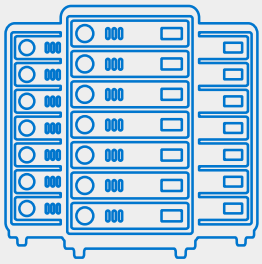
Business results

- Enabling data-centric AI research for academic and industrial users
- Fueling scientific discoveries with a unique data-acceleration architecture
- Making HPC resources accessible to researchers via a cloud interface
- Advancing the UK government's industrial strategy

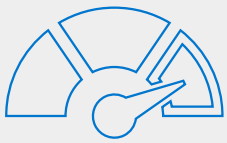
Cumulus supercomputer metrics

3.8 petaflops

74,000 cores



Storage system transfer speeds



500 gigabytes/sec

2 million IOPS

The pursuit of excellence

The University of Cambridge is one of the world's oldest universities and most highly regarded centers of academic scholarship. Its reputation for outstanding academic achievement is known worldwide and reflects the intellectual achievement of its students and faculty, as well as the world-class research carried out at the University and its Colleges. That's all part of the University's mission to contribute to society through the pursuit of education, learning and research at the highest international levels of excellence.

These principles guide the work of the University's Research Computing Service, which provides leading-edge research computing services across all academic disciplines and to the broader realm of the UK scientific and industrial community. Today, to help its constituents stay at the forefront of artificial intelligence and scientific research, the Research Computing Service operates the UK's largest academic supercomputer — the Cumulus–UK Science Cloud.

Accelerating scientific discoveries

The Cumulus system, delivered by Dell Technologies, helps researchers accelerate scientific discovery with more than 2 petaflops of compute performance and an innovative architectural feature called the Data Accelerator, or DAC. This feature proactively copies data from the cluster's disk storage subsystem and pre-stages it on fast NVMe storage devices that feed data to applications at a rate required for maximum performance.

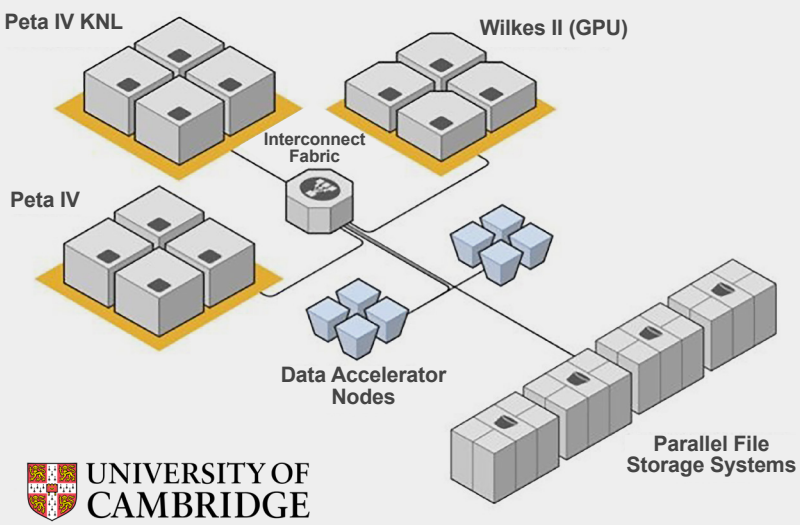
Dr. Paul Calleja, the University's Director of Research Computing Services, says the DAC is a response to the challenging data processing requirements that come with the size of the datasets used in today's scientific research projects.

"Data processing is becoming more and more of an essential element within research computing," Dr. Calleja says. "We are working with very large datasets, and the traditional spinning disk file systems can't keep up. We find applications being slowed down due to lack of data."

With the DAC, the Cumulus system leverages new solid state storage devices to avoid the drags on system performance that arise when compute nodes have to wait for data to be fed to them. The result is record-breaking storage performance.

"Currently, the Data Accelerator is the fastest HPC storage system in the world," Dr. Calleja says. "We are able to obtain 500 gigabytes per second read/write with IOPS up in the 20 million range."

The Data Accelerator that is embedded in the Cumulus cluster incorporates components from Dell Technologies, Intel and Cambridge University, along with an innovative orchestrator built by the University of Cambridge and StackHPC. This unique architecture allows the Research Computing Service to leave data on cost-effective disk storage until it is required by an application. At that point, the software in the system caches a copy of the data in the DAC nodes, which are based on fast solid-state storage devices that can move data at the speed of the cluster's parallel processing nodes.



UNIVERSITY OF CAMBRIDGE

The Data Accelerator

The Data Accelerator that is embedded in the Cumulus cluster incorporates components from Dell Technologies, Intel and Cambridge University, along with an innovative orchestrator built by the University of Cambridge and StackHPC. This unique architecture allows the Research Computing Service to leave data on cost-effective disk storage until it is required by an application. At that point, the software in the system caches a copy of the data in the DAC nodes, which are based on fast solid-state storage devices that can move data at the speed of the cluster's parallel processing nodes.

DELL Technologies

The Cumulus system: a look under the hood

The Cumulus system is based on Dell EMC PowerEdge servers, Intel® Xeon® processors and the Intel Omni-Path Architecture (Intel OPA) and Mellanox HDR InfiniBand. It also incorporates OpenStack software to control pools of compute, storage and networking resources and make them accessible to users via a cloud interface. OpenStack provisions the x86 and GPU bare-metal and virtualized hosts in the cluster.

The DAC component of the system was built using Dell EMC PowerEdge R740xd 2U servers with Intel Xeon processors. The DAC servers have 16-core Intel Xeon Gold CPUs running at 2.5 Ghz. Each of these servers contains a PCIe-x switch to connect six P4610 Series NVMe solid state drives (SSDs) to each CPU, for a total of 12 SSDs for each server. The Cumulus installation uses 24 of these servers with Intel Optane™ DC persistent memory, with 12 NVMe drives per server. To balance the system (based on assessing the peak performance of NVMe and network bandwidth), each of the 24 servers has two Intel OPA adapters.

In addition to the optimization of the individual DAC server configurations, the Cambridge Research Computing Service worked closely with Dell Technologies and Intel to optimize the network topology to best exploit the DAC configuration. And on the OpenStack side, the system makes use of the iDRAC capabilities embedded in Dell EMC servers for provisioning and management of the cluster, in particular with respect to infrastructure as code driven by the Ansible IT automation platform.

The Cumulus system, which is open to academic and industrial users across the UK, was delivered in partnership with Dell Technologies and StackHPC, a UK company specializing in the convergence of HPC and cloud. It was funded with investments totaling more than £18 million from the UK's Engineering and Physical Sciences Research Council (EPSRC) Tier 2 HPC activity, the UK's Science and Technology Facilities Council (STFC) via the DiRAC and IRIS HPC programs, and the University of Cambridge.

The Cumulus–UK Science Cloud

is an outgrowth of the University of Cambridge's Peta4 supercomputer, launched in 2017. At 1.6 petaflops, Peta4 entered the Top500 supercomputer rankings as the fastest academic supercomputer in the UK.[1]

1 University of Cambridge Research Computing Services, https://www.hpc.cam.ac.uk/.

Empowering AI research teams

With its performance-driven architecture, the Cumulus–UK Science Cloud has what it takes to help users solve extremely difficult data-driven, simulation and AI challenges.

"For people who need to do analytics or machine learning and process lots of data, we are bringing together on one system high levels of compute and high levels of I/O, combined with Hadoop and machine learning frameworks delivered within an OpenStack software environment, which allows both customizability and security for the tenants," Dr. Calleja says. "With all those things together, this machine can be used to deliver data-centric research to new and emerging communities."

Dr. Calleja notes that AI researchers require supercomputing capacity that is capable of processing huge amounts of data at very high speeds — which is exactly what Cumulus is designed to do.

"You cannot feed these people enough compute," he says. "They will eat whatever you give them. Cambridge's supercomputer provides researchers with the fast and affordable supercomputing power they need for AI work."

Many AI projects involving Cambridge researchers are already under way, according to Dr. Calleja.

"In the life sciences, we are working on medical imaging analysis and genomics, and in astronomy, scientists are using AI research to map exoplanets," he says.

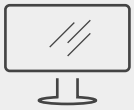
One long-running initiative, the Square Kilometre Array project (SKA), is using OpenStack and the Data Accelerator to prototype next-generation HPC systems. The SKA is a new global radio telescope project, currently in the design study phase. The telescope is planned to be a "world array" that will be located partly in South Africa and partly in Australia. It will have sensitivity 100 times greater than the most sensitive radio telescopes of the present generation, and the ability to survey the sky up to 1 million times faster.[2]

2 University of Cambridge Department of Physics, Cavendish Laboratory, Square Kilometre Array (SKA) research.

“We also have some fairly groundbreaking work around genomics with the Genome 10K Project, where researchers are using the Hadoop data analytics engine to drive groundbreaking results around very large-population genome studies,” Dr. Calleja says. Under that project, established in 2009, a consortium of biologists and genome scientists are working to facilitate the sequencing and analysis of the complete genomes of 10,000 vertebrate species.

Those are the kinds of challenges that are best conquered with the processing power of the world’s fastest supercomputers — like the Cumulus–UK Science Cloud.

“The cloud is open to all UK academic and industrial users,” Dr. Calleja says, “and it is ready and able to solve the UK’s most challenging data-driven simulation and AI problems. Cumulus has what it takes to drive both pure research and industry product development — faster, further, bigger.”



Learn more about Dell EMC advanced computing

Unlock the value of data with artificial intelligence

Share this story

DELL Technologies

intel