SUCCESS STORY

THE PAWSEY SUPERCOMPUTING CENTER IN PERTH, WESTERN AUSTRALIA, IS ONE OF TWO TIER 1 high-performance computing (HPC) centers in Australia and one of the most powerful facilities in the Southern Hemisphere. Pawsey supports scientific breakthroughs in radio astronomy, energy resources and engineering. At any given point, more than a thousand scientists rely on Pawsey's state-of-the-art facilities to conduct data-intensive research, complex simulations and advanced visualizations.

The Center also plays a pivotal role in the trailblazing Square Kilometer Array (SKA) project, which focuses on building a next-generation radio telescope that will be more sensitive and powerful than today's most advanced telescopes to survey the universe with incredible depth and speed.

According to Neil Stringfellow, executive director of the Pawsey Supercomputing Center, the SKA project will generate massive amounts of data from thousands of connected antennae, giving astronomers unprecedented insights into the formation of the universe. “The volume of data rates is growing rapidly from existing telescopes, which will expand even further with the SKA project in the future,” he explains. “To support this ground-breaking research, we must provide scientists around the world with easy access to high-end computing platforms and resilient, scalable storage.”

THE CHALLENGE

While Pawsey has a world-class reputation for driving radio astronomy research, the Center also supports a wide range of other projects that demand access to powerful HPC and storage resources. “The prevalence of diverse, data-intensive projects here is higher than at most other centers,” Stringfellow says. “The deep level of bioinformatics, data from observations and experiments, along with computational modelling research, produces extremely important data collections that need to be easily accessible for processing and analysis.”

An ongoing challenge for Pawsey's technology team is simultaneously keeping pace with approximately 50 large data collections while supporting up to 150 projects on the supercomputers. “Our storage needs continually evolve along with our science,” says Chris Schlipalius, senior storage infrastructure specialist at the Pawsey Supercomputing Center. “Researchers need to store massive media files, videos, images, text, and metadata. These all need to be stitched together, accessed, searched, shared, and archived.”

Scientific collaborations can also involve a half-dozen other research teams, which adds another layer of complexity. For instance, Pawsey works closely with various government departments throughout Australia and Western Australia. The organization also collaborates with its core members as well as industry participants to provide innovative solutions for the SKA pathfinders, the Australian Square Kilometer Array Pathfinder (ASKAP) and the Murchison Widefield Array (MWA). In addition, Pawsey has highly collaborative relationships with its Western Australian university partners, including Edith Cowan University, Curtin University, Murdoch University and The University of Western Australia as well as other key stakeholders on a range of research projects.

CHALLENGES

• Highly diversified, data-intensive research required massive compute and storage power to support up to 50 large data collections and more than 150 projects simultaneously
• Reliable and flexible file system was required for large parallel and smaller serial jobs
• Dynamic storage scalability was needed to address first-of-its-kind global scientific collaborations

SOLUTION

DDN's GRIDScaler® with 5 PBs of scalable storage as well as 2 PBs of additional DDN disk storage

RESULTS

• DDN storage delivers performance and stability
• Pawsey can keep pace with dynamic storage growth to support researchers and drive major radio astronomy projects
• DDN underpins the Center's ability to empower scientists in collecting, analysing and archiving a wide range of research data

THE CHALLENGE

While Pawsey has a world-class reputation for driving radio astronomy research, the Center also supports a wide range of other projects that demand access to powerful HPC and storage resources. “The prevalence of diverse, data-intensive projects here is higher than at most other centers,” Stringfellow says. “The deep level of bioinformatics, data from observations and experiments, along with computational modelling research, produces extremely important data collections that need to be easily accessible for processing and analysis.”

An ongoing challenge for Pawsey's technology team is simultaneously keeping pace with approximately 50 large data collections while supporting up to 150 projects on the supercomputers. “Our storage needs continually evolve along with our science,” says Chris Schlipalius, senior storage infrastructure specialist at the Pawsey Supercomputing Center. “Researchers need to store massive media files, videos, images, text, and metadata. These all need to be stitched together, accessed, searched, shared, and archived.”

Scientific collaborations can also involve a half-dozen other research teams, which adds another layer of complexity. For instance, Pawsey works closely with various government departments throughout Australia and Western Australia. The organization also collaborates with its core members as well as industry participants to provide innovative solutions for the SKA pathfinders, the Australian Square Kilometer Array Pathfinder (ASKAP) and the Murchison Widefield Array (MWA). In addition, Pawsey has highly collaborative relationships with its Western Australian university partners, including Edith Cowan University, Curtin University, Murdoch University and The University of Western Australia as well as other key stakeholders on a range of research projects.
**BUSINESS BENEFITS**

- Pawsey is supporting researchers making ground-breaking scientific discoveries, including the Square Kilometer Array project, which is building the most advanced telescope in the world to study the formation of the universe.
- With DDN storage, Pawsey can empower researchers to speed scientific discoveries in radio astronomy, renewable energy, and geosciences.
- Pawsey is positioned as an international center for scientific exploration, collaboration and excellence.

> The volume of data rates is growing rapidly from existing telescopes, which will expand even further with the SKA project in the future. To support this ground-breaking research, we must provide scientists around the world with easy access to high-end computing platforms and resilient, scalable storage.

**Neil Stringfellow**

Executive Director of the Pawsey Supercomputing Center

> Stability was a significant component in our evaluation. Storage resources need to be available and accessible so that scientists can perform ground-breaking research without disruption and without silent data loss.

**Chris Schlipalius**

Senior Storage Infrastructure Specialist at the Pawsey Supercomputing Center

> For example, running oceanographic drift models on Pawsey resources have enabled scientists at The University of Western Australia to determine debris locations of the missing Malaysian Airlines MH370 flight while potentially bringing researchers closer to finding the plane's crash site. Similarly, 3D and time-lapse 4D data-hungry seismic wave modelling is central to industrial exploration and decision-making in gas extraction and CO2 sequestration projects.

In supporting diversified research, Pawsey must accommodate a mix of extremely large and small processing workloads. “Today's researchers are writing small yet demanding simulation codes, therefore supporting mixed I/O processing is paramount,” Schlipalius says. “They're also using rapid application development, which requires a lot of metadata access that can strain storage and file systems.”

**THE SOLUTION**

The Pawsey Supercomputing Center sought reliable high-performance, high-density disk infrastructure that would provide the flexibility to scale storage both out and up as needed. “There's often a trade-off between performance and stability,” adds Schlipalius. “Our researchers need both, along with tiering options, so that we can put data where it makes the most sense at different project phases.”

The ability to plug seamlessly into other existing tape and cloud storage as well as back-end systems was important, as was the ability to address the Center's diverse research requirements. “Handling mixed I/O workloads is critical while protecting against data corruption,” Schlipalius notes. “Equally important is ensuring that scientists can access their data regardless of location.” A proviso to this ease of data access is ensuring secure access.

“Pawsey recognized the requirement for a system that supports a wide variety of data storage needs,” adds Schlipalius. With these requirements in mind, Pawsey surveyed the storage solution landscape. The team recognized DDN's scalable GRIDScaler® file storage system capabilities, which integrated storage with various features, including IBM's Spectrum Scale parallel file system, tiering choices, replication, data protection and data management. “We have a couple of front-end data access services that require high-speed storage connections over Ethernet,” adds Schlipalius. “These services are provided with high availability, stability and the reliability we demanded to support researchers.”

The introduction of IBM's Spectrum Scale parallel file system was novel for Pawsey as only a small number of centers currently use this technology in Australia. To ensure Pawsey staff could upskill quickly and improve their knowledge to better support researchers and knowledge sharing with colleagues, DDN established an IBM® Spectrum Scale user group in Australia. The technical group has been meeting regularly, ably assisted by DDN's sponsorship and support.

Adding DDN storage with IBM's Spectrum Scale gave Pawsey confidence in keeping up with dynamic storage growth, with the security of knowing additional systems can be added non-disruptively. Another decision factor was DDN's reputation in hardware resiliency. “Stability was a significant component in our evaluation,” Schlipalius says. “Storage resources need to be available and accessible so that scientists can perform ground-breaking research without disruption and without silent data loss.”

After weighing all its selection criteria, Pawsey deployed two DDN GS12K storage appliances with 5 PBs of storage capacity, supporting an array of game-changing scientific discoveries, including the Desert Fireball Network (DFN) project. One of the first projects to take advantage of the new DDN storage, DFN uses cameras to track fireballs as they shoot across the Australian desert night sky, aiding in the discovery and retrieval of newly fallen meteorites.

After ingestion, this network of images then resides on the DDN GS12Ks. Researchers work with metadata to triangulate events that facilitate rapid meteorite recovery. “This is an example of a complicated, multi-step science project; encompassing data collection, interrogation, computational processing and analysis,” explains Schlipalius. “Storage plays a critical role in the..."
Stability is an advantage to DDN’s hardware and GRIDScaler® file system. We’ve had no significant issues securing, updating and maintaining this system.

Chris Schlipalius
Senior Storage Infrastructure Specialist at the Pawsey Supercomputing Center

THE BENEFITS

With its growing infrastructure, Pawsey is becoming well positioned to enable geoscientists to unlock the secrets of the solar system. In addition to high-profile projects, including DFN, ASKAP, and MWA, Pawsey made headlines for supporting a research team in surveying the entire Southern sky as part of the Galactic and Extragalactic All-Sky Murchison Widefield Array survey (GLEAM). GLEAM Researchers have stored and processed more than 600 TBs of radio astronomy data to produce the world’s first radio-color panoramic view of the universe and catalog more than 300,000 radio galaxies from the extensive sky survey. “Pawsey strives to enable researchers to focus on their science with the knowledge that their unique project data is readily available to drive game-changing discoveries,” notes Stringfellow. “Having a storage system with high reliability and resilience makes this job easier.”

The Pawsey team has worked hard to make storage access reliable and transparent, ensuring that researchers don’t need in-depth knowledge of writing scripts or dealing with file systems. “Stability is an advantage to DDN’s hardware and GRIDScaler® file system,” Schlipalius notes. “We’ve had no significant issues securing, updating and maintaining this system.”

Pawsey continues to make significant inroads in helping researchers push boundaries even further. The team recently deployed another 2 PBs of DDN scratch storage to support ongoing data analysis in radio astronomy research. “Scientists come to us with lots of speculative technologies with different protocols and access methods, so flexibility is key,” adds Stringfellow. “We need to adapt to changing research requirements, whether we need to support Big Data analytics, HPC processing and global data sharing as well as connect with some new service or technology.” For the Pawsey Supercomputing Center, the future holds the promise of additional scientific breakthroughs and constant spikes in HPC and storage needs. In terms of computing power and storage requirements, the SKA project alone is predicted to surpass the most demanding research to date, including the Large Hadron Collider.

“We’re predicting an increase of up to 15 PBs of data annually for the SKA precursor telescopes alone, but it’s difficult to forecast the amount and rate of growth going forward,” concludes Stringfellow. “We need to enable both ASKAP and MWA scientists to carry out their missions to collect ground-breaking data on the evolution and formation of galaxies. The recent 2 PB of DDN storage will support these data collection missions.”

Stability is an advantage to DDN’s hardware and GRIDScaler® file system. We’ve had no significant issues securing, updating and maintaining this system.

Neil Stringfellow
Executive Director of the Pawsey Supercomputing Center

Pawsey strives to enable researchers to focus on their science with the knowledge that their unique project data is readily available to drive game-changing discoveries. Having a storage system with high reliability and resilience makes this job easier.

Pawsey Supercomputing Center

Pawsey strives to enable researchers to focus on their science with the knowledge that their unique project data is readily available to drive game-changing discoveries.

Pawsey Supercomputing Center
ABOUT DDN®

DataDirect Networks (DDN) is the world’s leading big data storage supplier to data-intensive, global organizations. DDN has designed, developed, deployed, and optimized systems, software, and solutions that enable enterprises, service providers, research facilities, and government agencies to generate more value and to accelerate time to insight from their data and information, on premise and in the cloud. For more information, visit our website www.ddn.com or call 1-800-837-2298.