

PAWSEY LEVERAGES VERSITY FOR MASS STORAGE

Industry

Scientific Research

Challenges

- Exponentially increasing storage needs
- Legacy solution unable to meet performance and scale requirements, costly to upgrade and maintain, and unfit to work with modern technologies
- Inability to access and leverage PBs of scientific data
- Need to integrate pool of warm object with offline cold storage

Solution Focus

- High-performance tape management solution
- S3 protocol support between online and offline storage pools

Key Results

- Took over 150PBs of cold data from a legacy platform in a week, without a data migration
- Improved management efficiency with Intuitive GUI and automation workflows
- Improved IT spend: eliminated the need to upgrade expensive legacy system and re-used enterprise tape libraries, drives, and cartridges
- Able to leverage cutting-edge technologies like Kafka, Grafana, and Keycloak, full 100 GBE integration

The Pawsey Supercomputing Research Centre houses some of the most powerful supercomputers in the Southern Hemisphere and supports data analysis in scientific domains such as astronomy, chemistry, life science, engineering, physics, earth science, and more.

Hitting the Limits with Legacy Systems

Pawsey needed to adopt a data management platform that would keep pace with exponentially increasing storage requirements for some of the world's most data-intensive, big-science projects. One such project is the Murchison Widefield Array (MWA), a radio telescope located in outback Western Australia at *Inyarrimanha Ilgari Bundara*, the CSIRO Murchison Radio-astronomy Observatory. Made from thousands of small antennas, the telescope collects immense amounts of radio data, which are used to examine the Universe in more detail than previously possible. MWA data are stored at Pawsey and accessed by astronomers through an online platform, the All-Sky Virtual Observatory. Through Pawsey, MWA researchers can download their observations, process data, and run simulations on the world-class supercomputing systems.

The raw data rate into the MWA correlator is 136 Gbit/s (17 GB/s) which is the equivalent of more than 4 HD movies every second. The data ultimately ends up being transported and archived at Pawsey, where, as of September 2022, the MWA has 43 PB of data – the equivalent storage of 1.3 million 32GB iPads!

The work and cost of maintaining the current system combined with the sheer volume of the data produced by the MWA and other projects requiring high-performance, long-term storage motivated Pawsey to look for a solution to deal with the data deluge. Pawsey sought a new scalable, performant, and economically viable solution to manage the vast amounts of data gathered and produced by research projects.

Solving for today

Research data was previously stored on a legacy hierarchical storage management (HSM) solution. Over 100 PB of legacy data was stored in a proprietary format on the system, creating a potential long term vendor lock-in situation for Pawsey. A traditional data migration requiring Pawsey to read back all the existing data and write it to a new system would have taken years and consumed considerable

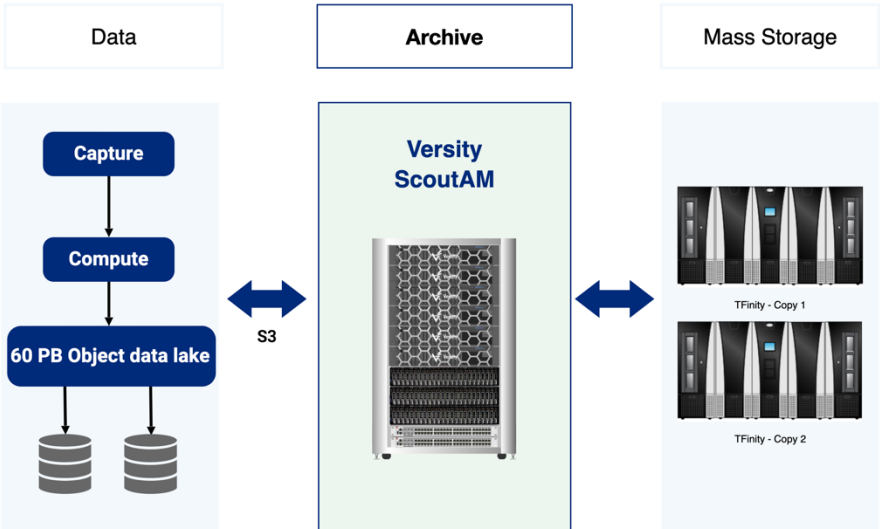
management and financial resources. A fast and reliable method for switching a large existing data collection to a new solution was required.

Pawsey needed to increase the performance and scale of their cold storage system to meet future long-term data storage requirements. They already understood that large-scale tape storage is the absolute lowest-cost data storage solution, which has other benefits, such as low power consumption and an airgap for security. They also wanted to incorporate a large object storage data lake between their scratch storage and their long-term cold storage. This required a platform capable of providing scalable S3 access to their tape storage resources, which would enable a cloud-scale cold storage service for scientists and researchers.

Pawsey Solution Architecture

Preparing for the future

Using Versity's Zero Data Migration feature, Pawsey was able to avoid a time-consuming and costly data migration and instead had a smooth and quick transition. Versity enabled Pawsey to maintain access to their existing data on tape without changing the format, while still making that data transparently available to researchers, scientists, and third-party tools. Versity delivered a high-performance S3 interface to tape, so Pawsey was able to still use tape as their large-scale on-



premises cloud storage pool, yet using the Versity platform, they could quickly locate and access the data they needed. Versity's mass data management software was easy to install and operate, further reducing the administrative burden for the Pawsey IT team. The solution's seamless fit into centralized monitoring, reporting, and authentication services, along with effortless integrations into the existing 100GBE network, aligned with Pawsey's desire to utilize standard procedures and methods.

Results

Using Versity's high performance and powerful capabilities to manage 150PBs of unstructured data, Pawsey was able to manage existing data and ingest huge amounts of new data while easily managing the massive Banksia data archive. Through Versity's full-featured, rich APIs, users and applications were able to interact with, manipulate, and capitalize on the data in the system. Data accessibility was finally possible through regular, standard interfaces and was available to plug into Pawsey's data-driven workflows.

"We are very happy with Versity's solution. We were impressed with the Zero Data Migration capability, 150PBs of data in one week! The solution was very easy to install, configure, and operate - we were up and running in no time. The modern S3 and REST APIs with broad coverage for external tooling and applications to support our data-driven workflows was a home run because we can now easily utilize Kafka, Grafana, and Keycloak to further boost our scientific mission." said Chris Schlipalius, Team Lead & Senior Systems Administrator. "We also recognized Versity's continued commitment to mass data management through their vibrant roadmap and a rapid feature release cadence. Coupled with Versity's scalable, modular, and simple upgrade approach, we know we have a future-proof solution."



Versity gives you clarity and control over your archival storage, so you can allocate more resources to your core mission.

Contact
sales@versity.com
1-844-726-8826