



CASE STUDY

DiRAC Memory Intensive Service at Durham University preserves complex cosmological simulation data

“Collaboration is key at DiRAC sites and we expect it from our technology providers. We have seen Spectra step up to the mark more than once since the deployment of our Spectra T950 Tape Library. We’ve received very good support and advice from the Spectra team at every step of the way.”

Dr. Alastair Basden, Technical Lead for the DiRAC Memory Intensive Service, Durham University



*Spectra
T950
Tape
Library*



Durham University is home to the DiRAC Memory Intensive Service, based in Durham’s Institute for Computational Cosmology (ICC). DiRAC (Distributed Research Utilizing Advanced Computing) is the integrated supercomputing facility for theoretical modeling and HPC-based research in particle physics, astronomy and cosmology and nuclear physics. It is a key part of the infrastructure supporting the UK’s Science and Technology Facilities Council (STFC) Frontier Science program. Four UK universities – Cambridge, Durham, Edinburgh and Leicester – are responsible for delivering DiRAC’s HPC services.

Learn more at dirac.ac.uk.

DiRAC

The Challenge

DiRAC provides research scientists across the UK with a variety of computer architectures, matching machine architecture to the algorithm design and requirements of the research problems to be solved. The DiRAC Memory Intensive Service, the seventh increment in a series of HPC clusters at Durham University, provides researchers with 452 nodes each with 512GB RAM, totalling 12,656 cores of computing power. Researchers at Durham are primarily focused on cosmology and astronomy. They leverage the facilities to understand more about physics and the universe by generating cosmological simulations of galaxy formation and evolution, mapping initial star conditions and tuning them to achieve a match over time with what is presently seen in the sky using telescopes. These simulations require large amounts of memory and RAM, and the Durham facilities are unique in providing about 230TB of RAM spread throughout their HPC cluster. This research data gets stored to disk, currently amassing around 10PB in primary storage across four generations of GPFS and Lustre file systems.

The growing need for ever-higher memory-intensive computing generates significant data volumes. Petaflop compute and petabyte storage requirements are integral to DiRAC-supported projects. DiRAC’s future Data Management plans include an archival component for both research database and finished



The Ogden Centre for Fundamental Physics is home to Durham’s ICC and the DiRAC Memory Intensive Service.

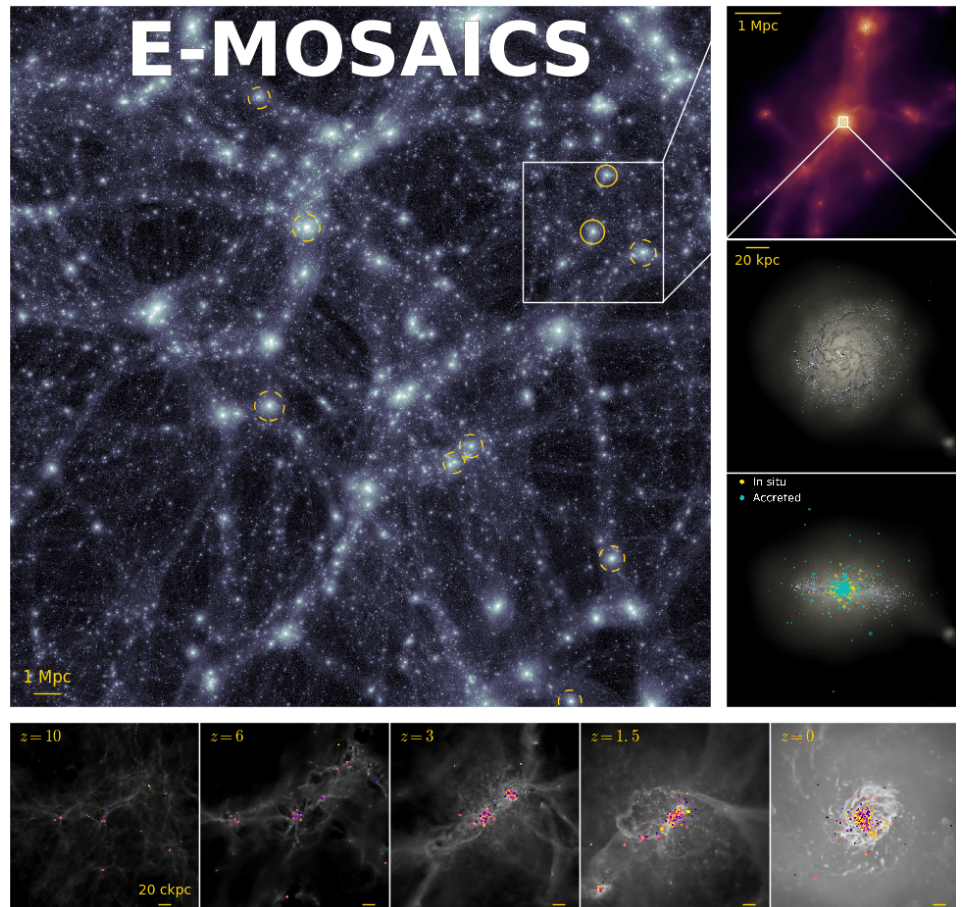
CASE STUDY: DiRAC Memory Intensive Service at Durham University

peer-reviewed scientific research documents. Over time, Durham expects to see a 10-fold increase in processing and a corresponding augmentation in data creation and storage requirements. They estimate that they will generate upwards of 20PB of data to store by 2022. Furthermore, researchers may need to revisit the data for up to 15 years, meaning Durham must ensure the data remains uncorrupted for future use.

The Solution

Given their multi-petabyte requirements for research project archiving, Durham University was looking to implement a new system that would effectively enrich their data storage infrastructure with a less expensive solution. Durham sought a solution that would archive in an open file format and handle incremental and full backups, enabling them to implement a comprehensive data protection strategy that would ensure long-term storage and retention.

After exploring other options, Durham University deployed a Spectra® T950 Tape Library with LTO tape drives because of Spectra's reputation for outstanding support and long-term commitment to customer success. They initially deployed the solution using LTO-7 Type M media, and then upgraded to LTO-8 in 2019. By archiving research data to the Spectra tape library, Durham offloads primary storage, relieving higher cost primary storage capacity. The Spectra tape library seamlessly integrates with Durham's data mover of choice, Atempo Miria, which is used to manage both backup and archiving of critical data. Multiple copies of data are written to the Spectra T950 using the open Linear Tape File System (LTFS) file format, a common industry magnetic tape format that presents a standard file system view of the data. Archiving performance of Lustre file systems is running at full tape speeds. With the Spectra solution, researchers at Durham now have the security of knowing that their data is backed up, available and protected from failure.



Created by the Virgo Consortium utilizing the Memory Intensive Service, these "E-MOSAIC" simulations are the first simulations that self-consistently follow the formation and evolution of star clusters and their host galaxy in a full cosmological setting.

Looking to the future, Durham begun testing the archiving of data from other DiRAC sites to its data storage architecture. Spectra will continue to work in a spirit of collaboration to build enduring success for technology and storage requirements at Durham.

Environment Snapshot

- Spectra T950 Tape Library
- Four LTO-8 tape drives
- Spectra Certified Media
- BlueScale® Vision Camera
- BlueScale® Standard Encryption
- Atempo Miria Archiving

Why Spectra?

- User accessibility and performance
- Easy-to-use
- Scalability
- Affordability
- Excellent Support
- Recommended in industry

Solution Recap

Spectra T950 Tape Library – The T950 library is designed and built to meet the stringent requirements of the enterprise for data integrity, data security and high reliability. The T950 reduces staff involvement significantly, scales in capacity and throughput, and supports multiple generations of current and future tape formats.