MIRIA CUSTOMER SUCCESS · ADVANCED RESEARCH/HPC

DiRAC

DiRAC Durham University deploys Atempo Miria Archiving

Thanks to Miria “we’re able to push data to tape and retrieve it very fast.”

“Miria for Archiving is incredibly powerful and feature-rich and should meet our future needs even though we’ve barely scratched the surface with the solution, we’re very impressed so far.”

— DR. ALASTAIR BASDEN, TECHNICAL LEAD FOR THE DIRAC MEMORY INTENSIVE SERVICE

THE ORGANIZATION

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 programme. Research scientists across the UK can apply for access to DiRAC’s powerful computing facilities.

DIRAC AND ATEMPO: SOME BACKGROUND

DiRAC’s Memory Intensive facility (COSMA) in Durham called on the services of Atempo, the Data Protection and Management specialists, together with their UK partner, OCF, to implement a multi-petabyte archiving project for their Lustre and Spectrum Scale (GPFS) data.

The DiRAC Memory Intensive service, the seventh increment in a series of HPC clusters at Durham, provides researchers with 452 nodes and 12,656 cores of computing power.

We spoke to Dr. Lydia Heck, DiRAC’s Technical Manager and Dr. Alastair Basden, Technical Lead for the DiRAC Memory Intensive service, both of whom are based in Durham University’s Institute for Computational Cosmology (ICC). Dr. Heck and Dr. Badsen gave us their feedback on Atempo’s Miria for Archiving solution which meets the requirements for research project archiving to tape. We also received some insight into how archiving and backup needs may evolve in the coming months.

DATA CONSIDERATIONS

Data considerations are central to DiRAC's missions. Petaflop computing and petabyte storage are essential requirements for research projects.

One of DiRAC@Durham's recent missions was to switch from Spectrum Scale (GPFS) storage to DDN's Lustre storage. The aim being, in the words of Dr. Heck, "to effectively enrich storage with a less expensive solution." The growing need for ever-higher memory intensive computing generates significant data volumes within HPC environments. Additional compute power is closely aligned with increased data storage requirements and greater tiering needs. By the time the latest funding phase was rolled out, the amount of generated data storage had surpassed 20 petabytes in 2022. DiRAC's Data Management Plan includes an archival components for the research databases and finished peer-reviewed scientific research documents. It is the research data which requires archiving.

THE FACILITY

Four UK universities, Cambridge, Durham, Edinburgh & Leicester, are responsible for delivering DiRAC's HPC services. These universities provide core HPC facilities along with expertise to enable multiple research, support, knowledge transfer and industrial partnership projects. The DiRAC Project Office is hosted by University College London.

THE SOLUTION IN PRACTICE: ATEMPO MIRIA FOR ARCHIVING

The role of Miria is to archive research data and relieve higher cost disk storage by offloading DiRAC research data from primary storage to four LTO tape destinations. Storage requirements grew from 6.5 PBs to 20 PBs with the latest funding phase, which represents a 10-fold increase in processing and data creation.

The existing archiving solution was slow and not scalable enough. In late 2018, Atempo offered a Proof of Concept (POC) on Spectrum Scale GPFS and Lustre. Miria proved to be the ideal solution for DiRAC's data archiving requirements, as rapidly demonstrated by Atempo. With a Miria Archiving Server and a dedicated Miria Data Mover, Lustre file systems are directly accessed and all data archive flows are efficiently handled from source to destination.

Miria Archiving User Interface empowers users to swiftly archive and restore their data. Researchers are free to organize the file structure of each research project as they see fit when archiving their data, reaching bandwidths of 20GBit/s between DiRAC's highly connected geographical sites.

All the physical data moving equipment along with the Miria server is installed and running and Dr. Badsen reports that:

> "The HTTPS protocol is used for all user and administrator actions, allowing for easy setup of an SSH tunnel and enabling global file archiving for users worldwide."

> "Archiving performance on Lustre file systems data flows is running at full tape speeds, which is perfect."

Archiving requires creating data copies on two tape locations, with the DiRAC service at Durham currently utilizing LTO7 and LTO8 technologies, both of which are fully supported by Miria.

In addition to user-performed file archiving, the technical teams at DiRAC Durham are utilizing Miria to back up users directories.

LOOKING BEYOND ARCHIVING

Beyond Miria's Archiving capabilities, Atempo also demonstrated FastScan capabilities for the backup of GPFS and Lustre. By optimizing the way new and modified files are recognized and stored, Miria's FastScan feature minimizes lengthy file system rescanning. This, in turn, enables DiRAC to efficiently perform high-performance backups of its critical file storage platforms.

DiRAC is keen to work alongside the DDN and Atempo teams for this integration and provide access to DiRAC's valuable beta testing on an HPC-scale file and storage environment. Dr. Heck concludes:

> "This is exactly what DiRAC expects from its active technology providers. To help justify expenses in significant resources, we need to be able to tell our stakeholders that we are also on the forefront of new and innovative solutions in many domains which include data movement and data protection."
>
> — DR. LYDIA HECK, DIRAC'S TECHNICAL MANAGER

CONCLUSION

To conclude, Miria Archiving solution is running with very positive results in terms of performance, efficiency and security. The Atempo and DiRAC teams will continue to work in a spirit of cooperation and mutualized resources to build on this initial success.

BUSINESS BENEFITS

- **Reduced storage costs**
- **Fast & high capacity archiving & backup**
- **Scalability adapted to customer needs**

WEBSITE
www.atempo.com