MIRIA CUSTOMER SUCCESS · ADVANCED RESEARCH/HPC

THE UNIVERSITY of EDINBURGH

Streamlining HPC Data for Discoveries

EMPOWERING USERS WITH EFFICIENT ON-DEMAND DATA TRANSFER

Miria: A Flexible Data Management Solution Empowering Scientific Discovery at the HPC Facility of the University of Edinburgh

"Atempo's Miria offers user-friendly and automated data management through its REST API and web interface, making it a perfect fit for HPC centers facing quick changes in their requirements and potential extreme needs."

— ANTONIN PORTELLI, PROFESSOR OF HIGH ENERGY PHYSICS, UNIVERSITY OF EDINBURGH, TECHNICAL DIRECTOR OF THE DIRAC HPC FACILITY

THE ORGANIZATION

We had the opportunity to speak with Professor Antonin Portelli, professor of High Energy Physics at the University of Edinburgh, to collect his feedback and experience on using Miria Data Management in their HPC installation. Professor Portelli is also one of the Technical Directors of the DiRAC HPC facility, one of the main UK national supercomputing facilities for fundamental science.

PROJECT BACKGROUND

DiRAC state-of-the-art supercomputing sites distributed throughout the country faced the challenge of managing the massive volumes of data required for particle physics and cosmology research.

Edinburgh's HPC team recognized the unique data storage and management challenges faced by DiRAC's users in their research, and sought innovative solutions to address them as computing demands continue to grow. Their goals included:

- Finding a scalable data protection solution to back up scientific data, which is expensive to generate,
- Providing user-centric data services to empower scientists with reliable data movement and data archiving solution, ultimately saving time and energy and enabling further scientific breakthroughs.

DIRAC AND ATEMPO: SOME BACKGROUND

DiRAC HPC facility is the primary provider of HPC resources to the STFC (Science and Technology Facilities Concil) Particle Physics, Astroparticle physics, Astrophysics, Cosmology, Solar System and Planetary Science, and Nuclear physics theory community.

DiRAC services are distributed over four different supercomputing sites: University of Cambridge, University of Leicester, Durham University and University of Edinburgh. The Edinburgh facility is actually the largest system and is mostly used for particle physics research.

The HPC services offered include modeling, simulation, data analysis and data storage. Edinburgh hosts DiRAC's Tursa, an Extreme Scaling GPU-based system made of 724 NVIDIA A100 GPUs connected to a 5PB DDN Lustre backed by a 16PB Spectra tape library.

> "Atempo's strategic alliances, its understanding of HPC needs and the openness of its solution Miria, made it simple for Atempo to build a flexible data management solution that is 100% adapted to our needs."

ATEMPO MIRIA: A SUCCESSFUL IMPLEMENTATION

Atempo's Miria Data Management solution carried by the GSI ATOS/EVIDEN was selected by Edinburgh's team as meeting all the requirements for scalability and flexibility.

The implementation began with the initial backup of the Lustre environment (4PB of Lustre to tape, with Miria in High-Availability and FastScan), ensuring that existing data was safely stored and protected. Next, the Miria Archiving component was added, enabling efficient long-term data storage and retrieval (8PB).

Atempo and ATOS/EVIDEN partnered to deliver a solution tailored to DiRAC's needs offering the following benefits:

- **Enhanced Automation Capabilities:** The integrated REST API streamlines integration with custom data management commands for scientists to use in command line with existing workflows, saving time for both scientists and IT.
- **Scalability and High-Performance Tape Management:** Miria is designed to grow with facility demands, optimizing tape-based workflows and improving data storage and retrieval processes.
- **Efficient On-demand Data Transfer:** Miria provides a user-friendly web interface to transfer data securely between systems, with pre-defined policies available to scientists for efficient and secure data management.
- **Customer-Centric Deployment:** Atempo's collaborative approach to implementation ensured seamless integration into the existing infrastructure and processes.

This partnership resulted in an effective data management system and a solid foundation for future collaboration. Atempo is undertaking the development of quota management in Miria to address the DiRAC's specific requirements.

In the future, DiRAC aims to enhance its data management processes and expand the use of Atempo's solution. Plans include broader implementation of the REST-API for advanced process automation and exploring third-party solutions like Storj Decentralised Cloud Storage. DiRAC will also work on developing a standardized API to streamline system communication and increase efficiency.

> "The tape solution offers an easy, effective and much needed sustainable way to offload data, with the archival/retrieval part of the REST API being the most important in our context."

> "The standardized REST API and interface offered by Miria makes it a good solution for IT departments looking to federate solutions and manage data across the facility."

KEY BENEFITS

- Improved archival space management and streamlined data offloading
- Intuitive interface and REST API for seamless user data management
- Exceptional performance and successful integration with DiRAC's systems
- Close relationships, flexibility, and customized solutions through collaborations with companies like ATOS/EVIDEN
- Large storage compatibility guide
- Ready to accommodate additional workflows

BUSINESS BENEFITS

Efficient on-demand data transfer for empowered users	**Scalable and flexible solution for HPC centers with changing needs**	**Easy deployment with a large open ecosystem**

WEBSITE
www.atempo.com