



Optimisation des données de calcul haute performance pour la recherche scientifique

Miria, la solution de gestion des données flexible déployée au centre HPC de l'université d'Édimbourg

LA RENCONTRE

Nous avons eu l'occasion d'échanger avec le Professeur Antonin Portelli, spécialiste de physique des hautes énergies à l'**Université d'Édimbourg**, sur l'utilisation de la solution de Data Management Miria au sein de leur centre de calcul intensif. Le Professeur Portelli est Directeur Technique du DiRAC, l'un des principaux supercalculateurs nationaux du Royaume-Uni dédiés à la science fondamentale.

L'ORGANISATION

Le DiRAC est le **principal fournisseur de ressources et services destinés à la communauté scientifique** de la STFC (Science and Technology Facilities Council) dans le domaine de la physique des particules, la physique des astroparticules, l'astrophysique, la cosmologie, la science du système solaire et des planètes, ainsi que la physique nucléaire.

Les données générées par le DiRAC sont réparties sur quatre sites de calcul : Université de Cambridge, Université de Leicester, Université de Durham et Université d'Édimbourg. Principalement dédié à la recherche en physique des particules, le site d'Édimbourg est le plus important des quatre.

Les services proposés sont notamment la modélisation, la simulation, l'analyse et le stockage des données. Le site d'Édimbourg héberge le système Tursa du DiRAC, un système dit « Extreme Scaling » doté de **724 GPU NVIDIA A100** connectés à un stockage **DDN Lustre de 5 PO**, et à une librairie de bandes **Spectra de 16 PO**.

LE PROJET

Le DiRAC, dont les supercalculateurs de pointe sont répartis dans tout le pays, a été confronté à un défi de taille : gérer les grands volumes de données nécessaires à la recherche en physique des particules et en cosmologie.

Consciente des challenges liés au stockage et à la gestion des données auxquels sont confrontés les utilisateurs du DiRAC dans le cadre de leurs recherches, l'équipe HPC d'Édimbourg a cherché des solutions innovantes pour faire face à l'augmentation constante des besoins de calcul.

Leurs objectifs sont les suivants :

- Mettre en place une **solution de protection des données évolutive** pour sauvegarder ces données scientifiques coûteuses à produire.
- **Placer l'utilisateur au centre des services de données offerts**, rendant les scientifiques autonomes pour déplacer et archiver leurs données, et ce, de manière fiable tout en économisant du temps et de l'énergie, et en contribuant à de nouvelles avancées scientifiques.

66 «Les alliances stratégiques, la compréhension des besoins du HPC et la souplesse de Miria, ont permis à Atempo d'offrir une solution de gestion de données flexible et 100% adaptée à nos besoins.»

99

Des utilisateurs autonomes pour transférer efficacement leurs données à la demande

66

« Miria, c'est à la fois une gestion des données accessible aux utilisateurs grâce à son interface web, et une capacité d'automatisation grâce à son API REST. C'est la solution idéale pour les centres de HPC confrontés à l'évolution rapide des demandes et aux besoins extrêmes. »

99

Antonin Portelli, Professeur de physique des hautes énergies, Université d'Édimbourg, Directeur Technique de l'installation HPC DiRAC

MIRIA : UN DÉPLOIEMENT RÉUSSI

L'équipe d'Édimbourg a choisi la solution de gestion de données Miria d'Atempo, proposée par le GSI ATOS/EVIDEN, pour sa capacité à répondre à toutes les exigences en matière d'évolutivité et de flexibilité.

Le déploiement a commencé par **la sécurisation du stockage Lustre** (4 PO mis sur bande, par **Miria Sauvegarde** en mode haute disponibilité et FastScan), pour les données existantes. L'ajout de **Miria Archivage** a permis la conservation longue durée efficace de 8 PO de données ainsi que leur restitution.

Atempo et ATOS/EVIDEN se sont associés pour proposer une solution adaptée aux besoins de DiRAC, offrant les avantages suivants :

- **Amélioration des capacités d'automatisation** : L'API REST permet une intégration simple de Miria dans des commandes personnalisées, utilisées par les scientifiques en ligne de commande dans leurs flux de travail existants. C'est un gain de temps, pour les scientifiques et pour les équipes IT.
- **Évolutivité et gestion haute performance des bandes** : Miria est conçue pour évoluer en fonction des besoins du centre, optimisant les flux de mise sur bande, les processus de stockage et de restitution des données.
- **Transfert efficace de données à la demande** : l'interface web conviviale de Miria transfère les données en toute sécurité entre systèmes, via des procédures mises à disposition des scientifiques pour la sécurisation de leurs données.
- **Déploiement centré sur le client** : Atempo a déployé Miria avec une approche collaborative permettant une intégration transparente de l'infrastructure et des processus existants.

66 « Grâce à son interface et à son API REST standardisées, Miria est une solution idéale pour les services informatiques qui cherchent à fédérer les services de gestion de données offerts à l'ensemble de leurs établissements. » **99**

Ce partenariat met en place une gestion des données efficace et constitue une base solide pour les collaborations futures. Atempo travaille sur le développement de la gestion des quotas dans Miria pour répondre aux attentes spécifiques du DiRAC.

66 « La mise sur bande est un moyen simple, efficace et durable de conserver nos données. De plus, la capacité à archiver/restituer ces données depuis l'API REST est décisive dans notre contexte. » **99**

À l'avenir, le DiRAC souhaite améliorer ses processus de gestion des données et étendre les capacités de la solution d'Atempo : élargir l'utilisation et l'automatisation de l'API REST et explorer la solutions de stockage Cloud décentralisé (Storj Decentralised Cloud Storage). Le DiRAC travaille aussi au développement d'une API standard pour simplifier la communication entre systèmes et accroître les performances.

LES BÉNÉFICES MÉTIER



Transfert à la demande, autonomie des utilisateurs



Solution flexible qui répond aux besoins évolutifs du HPC



Large éventail de technologies supportées



Consulter la fiche produit
<https://links.atempo.com/DM-presentation>



Contacter un expert Atempo
<https://links.atempo.com/DM-contact>

LES BÉNÉFICES

- **Amélioration de la gestion de l'espace d'archivage et simplification du transfert de données**
- **Interface Web intuitive et API REST pour une gestion transparente des données des utilisateurs**
- **Performance exceptionnelle et intégration réussie** avec les systèmes utilisés par le DiRAC
- **Relations de proximité, flexibilité et solutions personnalisées** à travers des collaborations avec des entreprises telles qu'ATOS/EVIDEN
- **Capacité à absorber des flux de travail supplémentaires**

