

Internship Position

Création d'une nouvelle source de données ultra-rapide pour Atoti

internships@activeviam.com

“Concevoir et implémenter une nouvelle source de données optimisée pour un chargement en mémoire ultra-rapide dans Atoti, en explorant des formats innovants et des techniques avancées de traitement des données”

● Contexte

ActiveViam est à la pointe de l'innovation dans l'analyse de données avec Atoti, notre plateforme qui permet à nos clients de traiter des volumes massifs de données à des vitesses inégalées. Les sources de données actuelles, telles que CSV, Parquet ou SQL, sont robustes, mais la demande croissante en rapidité de chargement et de traitement des données nous pousse à explorer de nouvelles méthodes. Le but est de repousser les limites actuelles du chargement in-memory, en cherchant à maximiser les performances en s'affranchissant de certaines contraintes des formats traditionnels.

● Descriptif du stage

Lors de ce stage, vous serez chargé(e) de créer une nouvelle source de données optimisée pour Atoti, en visant un temps de chargement in-memory minimal, essentiellement limité par la vitesse de lecture sur disque. Contrairement aux formats traditionnels, cette nouvelle source pourra contourner plusieurs limitations comme :

- La vérification de la duplication des clés
- Le partitionnement dynamique des données
- Les contraintes liées à des formats linéaires

Vous explorerez différentes approches, notamment en testant des formats de données existants, tels que Arrow, ou en concevant un format sur-mesure plus adapté aux besoins d'Atoti. Vous réaliserez également des benchmarks sur ces différentes solutions pour comparer les performances en termes de vitesse de chargement et de gestion mémoire

● Objectifs

- Proposer et concevoir un nouveau format ou une nouvelle technique de chargement des données optimisée pour des performances maximales dans Atoti.
- Tester des solutions de formats préexistants et les benchmarker par rapport aux solutions actuelles.
- Mettre en place des benchmarks pour mesurer l'impact de cette nouvelle source sur les temps de chargement et la gestion des données.
- Proposer des améliorations techniques à partir des résultats obtenus

● Technologies

Java, API I/O, systèmes de fichiers, formats de données (CSV, Parquet, Avro, etc.), JVM performance tuning, cloud benchmarking.

Ce stage, d'une durée de 5 à 6 mois, se déroulera à Paris, et pourrait aboutir à une offre de CDI dans notre équipe R&D, vous permettant d'intégrer directement vos travaux à la plateforme Atoti.