

Mini Lab – Atelier Déploiement de Pipeline

Vous faites partie d'une équipe chargée de déployer et de gérer des pipelines de données sur une plateforme Big Data dans une entreprise de services financiers. Votre objectif est d'assurer la préparation et la mise à disposition efficace des données aux différentes équipes utilisatrices, telles que les data scientists et l'équipe B.I.

- Analyse des besoins des utilisateurs : En tant que membre de l'équipe, vous devez comprendre les besoins spécifiques des data scientists et de l'équipe B.I. Identifiez les types de données requis, les fréquences de mise à jour et les transformations nécessaires.
- Conception du pipeline de données : Concevez un pipeline de données qui répond aux besoins identifiés. Déterminez les étapes de prétraitement, de transformation et de chargement nécessaires.
- Sélection des outils et des technologies : Choisissez les outils et les technologies adaptés pour implémenter le pipeline de données. Cela peut
- Inclure des Framework tels que Hadoop, Apache Spark, ou des outils spécifiques de gestion de flux de données.
- Configuration du pipeline : Configurez et paramétrez les composants du pipeline de données, tels que les extracteurs, les transformateurs et les Chargeurs.
- Assurez-vous que les données sont collectées, nettoyées, transformées et chargées de manière cohérente et fiable.
- Tests du pipeline : Effectuez des tests approfondis pour valider le bon fonctionnement du pipeline de données. Vérifiez l'intégrité des données, la
- Conformité aux règles de transformation et les performances globales du pipeline.
- Surveillance et maintenance : Mettez en place des mécanismes de surveillance pour suivre le flux de données à travers le pipeline. Identifiez-les.
- Problèmes potentiels et effectuez des tâches de maintenance régulières pour assurer la disponibilité continue du pipeline.
- Documentation et formation : Documentez le pipeline de données, y compris les étapes, les composants utilisés et les règles de transformation.
- Fournissez une formation aux utilisateurs finaux sur l'utilisation du pipeline et sur la manière d'accéder aux données préparées.
- Améliorations et évolutivité : Identifiez les opportunités d'amélioration du pipeline de données et proposez des modifications ou des ajouts pour répondre aux futurs

besoins des utilisateurs. Assurez-vous que le pipeline est évolutif et peut gérer des volumes de données croissants