

Mis à jour le 27/07/2023

S'inscrire

Formation Dagster : Orchestration des pipelines de données dans le cloud

3 jours (21 heures)

Présentation

La formation [Dagster](#), vous apprendra à utiliser cet orchestrateur de données. Vous saurez manier avec perfection ce cadre pour la création de pipelines de données et les expédier à une vitesse extraordinaire.

Durant cet apprentissage, votre équipe saura ingérer des données provenant de sources externes ou d'autres ressources de données. Vous pourrez également combiner et transformer les données de manière significative.

Dagster vous sera utile, car il dispose de trois points clés qui le rendent donc unique. Votre entreprise sera apte à développer localement et effectuer des tests automatisés. Vous pourrez notamment calculer la valeur d'un actif grâce à une fonction Python pure déclarative. Grâce à son découplage des canalisations de l'environnement, vous accèderez à un grand nombre de fonctionnalités.

Comme toutes nos formations, celle-ci vous présentera la [dernière version](#) en date et ses nouveautés : Dagster 1.1.

Objectifs

- Savoir intégrer divers outils dans Dagster
- Savoir où utiliser un actif stocké
- Maîtriser l'ensemble des concepts
- Comprendre l'architecture de déploiement et la mettre en place

Public visé

- Data Scientists
- Data Analysts
- Ingénieurs d'analyse
- Ingénieurs de données
- Ingénieurs DevOps

Pré-Réquis

- Connaissances générales sur les données
- Connaissance du langage Python

Programme de notre formation Dagster

Introduction

- Présentation de Dagster
- Installation
- Premier pas avec Dagster
- Télémétrie

Création d'un nouveau projet

- Utilisation du squelette de projet
- Démarrage du serveur Web Today
- Ajout de nouvelles dépendances Python
- Test unitaire
- Variables d'environnement et secrets

Actifs

- Définir un actif
- Construction de graphes de ressources
- Définir des ressources sans valeurs de retour ou d'arguments
- Test des actifs
- Automatisation de la matérialisation des actifs
- Partitionnement des actifs
- Personnalisation du stockage des actifs
- Travaux hors actifs

Concepts

- Opérations
 - Crochets
 - Évènements
 - Tentatives
- Graphiques
 - Dynamique
 - Emboîtement
- Travaux
 - Métadonnées et balises
 - Exécution des travaux
- Programmes
 - Capteurs
 - Cloisons
 - Lancement des remplissages
- Lancement de l'interface utilisateur Dagit

Déploiement

- Architecture de déploiement
- Services de longue
- Configuration
- Flux d'exécution des tâches
- Dagster Cloud

Intégrations

- Utiliser Dagster avec Airflow
- Spark
 - Exécution du code PySpark dans les opérations
 - Soumission d'opérations PySark sur EMR
- Validation des DataFrames Pandas avec les types Pandas
- Snowflake et Dagster
 - Utilisez Python pour analyser vos données stockées dans Snowflake sans requête SQL
 - Intégrez vos tables Snowflake à d'autres outils

Sociétés concernées

Cette formation s'adresse à la fois aux particuliers ainsi qu'aux entreprises, petites ou grandes, souhaitant former ses équipes à une nouvelle technologie informatique avancée ou bien à acquérir des connaissances métiers spécifiques ou des méthodes modernes.

Méthodes pédagogiques

Stage Pratique : 60% Pratique, 40% Théorie. Support de la formation distribué au format numérique à tous les participants.

Organisation

Le cours alterne les apports théoriques du formateur soutenus par des exemples et des

séances de réflexions, et de travail en groupe.

Validation

À la fin de la session, un questionnaire à choix multiples permet de vérifier l'acquisition correcte des compétences.

Sanction

Une attestation sera remise à chaque stagiaire qui aura suivi la totalité de la formation.