EduGroupe
Accompagner pour réussir

Dataiku DSS

Référence : **PYDS034** Durée : **3 jours (21 heures)** Certification : **Aucune**

Connaissances préalables

- Connaissance des principes de la manipulation de données et du machine learning

Profil des stagiaires

- Chefs de projet, développeurs, data-scientists, utilisateurs de Dataiku DSS

Objectifs

- Savoir installer Dataiku DSS
- Être capable de configurer Dataiku DSS
- Apprendre à l'utiliser depuis l'interface web ou des API

Certification préparée

- Aucune

Méthodes pédagogiques

- 6 à 12 personnes maximum par cours, 1 poste de travail par stagiaire
- Remise d'une documentation pédagogique papier ou numérique pendant le stage
- La formation est constituée d'apports théoriques, d'exercices pratiques et de réflexions

Formateur

- Consultant-Formateur expert Bigdata

Méthodes d'évaluation des acquis

- Auto-évaluation des acquis par le stagiaire via un questionnaire
- Attestation des compétences acquises envoyée au stagiaire
- Attestation de fin de stage adressée avec la facture

Contenu du cours

1. Présentation, concepts DSS

- Fonctionnalités : Mise à disposition des méthodes et outils de data-sciences à partir d'une interface graphique ou de langages de requêtage ou de programmation (Python, SQL, R), principe de "Self Service Analytics"
- Cluster DSS : Design, Deployer, Api, Automation et Govern Nodes
- Notions de projet, data, dataset, recipes
- Outils collaboratif : Wiki, tags, discussions...
- Premiers pas avec Dataiku DSS : exemples

2. Connexion aux données

- Connecteurs disponibles, formats de fichiers, formats spécifiques Hadoop/Spark
- Chargement des fichiers, emplacement des fichiers
- Système de fichiers, HDFS, Amazon S3, stockage Google Cloud, etc.
- FTP, connexion ssh, connexion aux bases de données SQL
- Autres stockages : Cassandra, MongoDB, Elasticsearch, ...

3. Préparation des données

- Scripts de préparation, échantillonnage, exécution, les différents moteurs d'exécution: DSS, Spark, base de données SQL

4. Graphiques et statistiques

- Présentation des types de graphiques disponibles et configuration
- Echantillonnage et exécution, graphiques de base, tables, nuages de points, visualisation cartographique, etc.
- Réalisation d'exercices pratiques
- Fonctionnalités statistiques disponibles, démonstrations

5. Machine learning

- Le laboratoire
- Fonctionnalités disponibles : apprentissage supervisé et non supervisé, algorithmes utilisés, scoring, deep learning...
- Evaluation, optimisation, compréhension du modèle
- Déploiement et cycle de vie du modèle

6. Flow/Recipes

- Interface graphique de gestion des cycles de traitement, des datasets, des traitements, paramètrage de l'exécution en parallèle, des données géographiques, export en format pdf ou images, etc.
- Travaux pratiques

7. Interfaces de programmation

- Présentation des différentes interfaces : SQL, Python, R
- Exemples d'interrogations SQL et Python
- Interfaçage avec Spark

Notre référent handicap se tient à votre disposition au 01.71.19.70.30 ou par mail à referent.handicap@edugroupe.com pour recueillir vos éventuels besoins d'aménagements, afin de vous offrir la meilleure expérience possible.