



STERE

Programme de formation Dataiku



dataiku

Notre contact :

Jean-Noël SAUNOIS

 contact@stere-informatique.fr

 46, rue de Lagny
93100 Montreuil

PRESENTATION : Cette formation vous permet de découvrir Dataiku DSS, connecter et préparer vos données, créer des modèles ML, automatiser vos workflows et exploiter toute la puissance de Python, SQL, R et Spark grâce à une interface intuitive et collaborative.

OBJECTIFS

A l'issue de cette formation, les participants seront capables de :

- Maîtriser l'utilisation de Dataiku DSS pour gérer des projets de data science en mode collaboratif.
- Savoir connecter, importer et préparer des données à partir de multiples sources (bases SQL, fichiers, Big Data, Cloud).
- Concevoir, entraîner et déployer des modèles de Machine Learning supervisés et non supervisés.
- Automatiser les workflows, gérer les cycles de vie des projets et appliquer les bonnes pratiques de gouvernance.
- Exploiter efficacement les langages Python, SQL et R au sein de l'interface DSS et intégrer Spark pour le traitement distribué.

DURÉE DE LA FORMATION

- 3 jours, soit 21 heures de formation.

PRÉREQUIS À LA FORMATION

- Connaissance des principes de la manipulation de données et du machine Learning.

PUBLIC VISÉ

- Chefs de projet, Développeurs, Data-scientists, Utilisateurs de Dataiku DSS

CONTENU DE LA FORMATION

1. Présentation, concepts DSS

- 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, Déployer, Api, Automation et Govern Nodes
- Notions de projet, data, dataset, recipes
- Premiers pas avec Dataiku
- Outils collaboratifs : Wiki, tags, discussions...

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 etc.

3. Préparation des données

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

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

5. Machine Learning

- 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 dataset, des traitements.
- Paramétrage de l'exécution en parallèle, des données géographiques
- Export en format PDF ou images etc.

7. Interfaces de programmations

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

MOYENS PÉDAGOGIQUES ET TECHNIQUES

- Accueil des stagiaires dans une salle dédiée à la formation.
- Supports de formation projetés sur écran.
- Etude de cas concrets.
- Mise à disposition en ligne de documents supports à la suite de la formation.

Formation à distance : utilisation de la plateforme de visioconférence client et partage d'écran pour le support de formation.

Attention : Nous ne fournissons pas le matériel informatique pour la formation. Les stagiaires doivent être équipés d'ordinateur et d'une connexion internet.

MODALITÉS D'ÉVALUATION

- Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de QCM, mises en situation, travaux pratiques...
- Le participant complète également un test de positionnement en amont et en aval pour valider les compétences acquises.

MODALITÉS ET DÉLAIS D'ACCES

- L'inscription doit être finalisée 24 heures avant le début de la formation.

TARIFICATION

Inter entreprise :

- 1800€ HT par personne pour 3 jours de formation.

Intra entreprise :

- Contactez-nous pour un devis personnalisé.