

ANALYSE DE DONNÉES ET MACHINE LEARNING AVEC SPARK EN PYTHON

IIA21

3 jours (21h) 2 280,00 €^{HT}

Objectifs

Comprendre le principe de fonctionnement d'Apache Spark • Utiliser PySpark pour manipuler des données avec Python • Traiter les flux de données en temps réel avec Spark Streaming

Profil Stagiaire(s)

Cette formation s'adresse à toute personne intéressée par la technologie Spark et sa mise en oeuvre avec le langage Python.

Pré-requis

Cette formation ne nécessite aucun pré-requis même si une connaissance de base du langage Python est recommandée.

Profil Animateur(s)

Cette formation est animée par un consultant-formateur en Intelligence Artificielle, enseignant-chercheur en IA.

Contactez-nous

Conseiller formation
09 88 66 10 00
inscriptions@demoss.fr

new

PROGRAMME

Introduction au big data

Environnements big data : Hadoop, Mapreduce, Spark, etc.

Découvrir l'environnement Spark

Utilisation Spark : Quels avantages ?

Comment installer Apache Spark

Quelle infrastructure pour utiliser Spark en entreprise ?

Les principes et fonctionnement de l'environnement Spark : RDD, DataFrame, DataSet...

Manipuler des données avec Python, Pyspark

Rappels sur Python et la manipulation des données

Utilisation de SparkSQL et des DataFrames pour manipuler des données

Bonnes pratiques de Spark pour la manipulation de volume important de données

Architecture et composants de la plateforme Hadoop

Chargement des données depuis Hadoop, depuis des fichiers csv...

Transformer des données (création de DataFrames, ajout de colonnes, filtres...)

Cas pratiques de chargement et de modifications de données avec PySpark

Comprendre MLlib avec les principaux algorithmes d'apprentissage et méthodes statistiques (régression, discrimination, forêt aléatoires, etc.)

Utiliser spark.ml pour la machine learning

Traitement de données textuelles

Automatiser vos analyses avec des pipelines

Analyser en temps réel avec Spark streaming

Principe de fonctionnement

Présentation des Discretized Streams

Les différents types de sources

Manipulation de l'API

Comparaison avec Apache Storm

ANALYSE DE DONNÉES ET MACHINE LEARNING AVEC SPARK EN PYTHON

IIA21

3 jours (21h) 2 280,00 €^{HT}

Méthode pédagogique

Présentation de la technologie Big Data Apache Spark et de sa mise en oeuvre avec Python, le langage de programmation le plus populaire actuellement.

Les entreprises Google, Facebook, Netflix, Airbnb, Amazon, la NASA et bien d'autres encore utilisent la technologie Spark pour le traitement de très gros volumes de données.

La capacité d'analyser d'énormes quantités de données est une compétence incontournable à l'heure de l'Intelligence Artificielle et de la Data Science.

Le cours présente Spark DataFrames en s'appuyant sur le langage Python avec la bibliothèque PySpark, la bibliothèque Machine Learning MLib ainsi que les technologies Spark SQL, Spark Streaming.

Moyens pédagogiques et techniques de mise en œuvre

Nos formateurs DEMOS sont recrutés conformément aux spécifications mentionnées pour chaque formation. Ce sont des professionnels en activité et/ou des experts dans leur domaine.

Ils utilisent des méthodes et outils appropriés aux formations qu'ils dispensent et adaptent leur pédagogie au public accueilli.

Par ailleurs, nos centres de formation DEMOS sont tous équipés :

- Salles de formation lumineuses, spacieuses, design, ergonomiques, mobiles et équipées d'écrans plasma, de PC portables/fixes si nécessaire à la formation suivie.
- Accès au wifi haut débit sur l'ensemble des lieux
- Espace de co-learning et webcorners
- Cafés, collations et rafraîchissement sont offerts

Et tous sont accessibles aux personnes à mobilité réduite

Le dispositif de suivi et d'évaluation

Pour assurer un suivi individuel, Demos a mis en place 2 types d'évaluation :

- Des évaluations des acquis en cours et en fin de formation

Elles peuvent être faites de différente manière selon le contenu de la formation suivie :

Quiz, exercice pratique, étude de cas, jeu de rôles, mise en situation, soutenance devant un jury pour les formations à finalité certifiante.

- Une évaluation de la satisfaction de chaque stagiaire est réalisée en ligne. Cette évaluation est complétée par l'appréciation du formateur à l'issue de chaque session.

En complément

Avant cette formation vous pouvez suivre :

- Programmation en Python