

Formation Python pour la Data Science, Le Langage Python - PY3

Présentation

Spark est un des principaux frameworks d'analyse de données Big Data. Il est massivement déployé sur les infrastructures de type Data Lake, soit directement, soit au travers d'autres outils tels que Dataiku DSS. PySpark est une bibliothèque permettant d'utiliser ce framework en Python.

Cette formation est un prolongement de la formation « PY2 – Exploration de données en Python ». Elle aborde principalement la manipulation de données dans un contexte Big Data, et plus spécifiquement le traitement de données distribuées sur un cluster de calcul avec la bibliothèque PySpark. Elle s'adresse donc à des personnes ayant suivi la formation PY2 ou ayant déjà une bonne compréhension des bases de la manipulation de données en Python avec Pandas.

Les supports de formation sont rédigés en anglais, toutefois les sessions sont données en français.

Durée : 7,00 heures (1 jours)

Tarif INTRA : [Nous consulter](#)

Objectifs de la formation

- Reconnaître les spécificités du traitement de données en environnement Big Data.
- Savoir expliquer les notions principales liées au traitement de données distribué sur un cluster de calcul
Spark :
 - Driver, master, exécuteur
 - Partitionnement, shuffle
- Appliquer les principaux types de manipulation de données avec la bibliothèque PySpark

Prérequis



- Toute personne ayant déjà manipulé des données en Python notamment avec la bibliothèque Pandas et désirant transposer ces traitements avec PySpark.
- Lecture de l'anglais technique
- Utilisation de Python, des Notebooks Jupyter et de la bibliothèque Pandas ou avoir suivi la formation « Python pour la Data Science : L'exploration de données en Python - PY2 »

Public

Toute personne ayant un profil technique ou ayant validée la formation « Le langage Python – PY1 » et

souhaitant s'initier à la manipulation et à l'exploration de données en Python.

Programme de la formation

- Big Data & Spark clusters
- Lecture et écriture de données (partitionnement)
- Exécution paresseuse et récupération de données
- Spark Dataframes et manipulation de données de base
- Structure d'une application Spark (Jobs, Stages, Tasks) +notion de shuffle
- Spark UI
- Group by et Jointure
- Spark et SQL

Organisation

Formateur

Les formateurs de Docaposte Institute sont des experts de leur domaine, disposant d'une expérience terrain qu'ils enrichissent continuellement. Leurs connaissances techniques et pédagogiques sont rigoureusement validées en amont par nos référents internes.

Riches de leur expérience sur le sujet, ils sauront accompagner vos collaborateurs dans leur montée en compétence

Moyens pédagogiques et techniques

- Exposé théorique sur le langage et exercices pratiques

- Fil rouge autour d'un projet concret
- Utilisation des cahiers (notebooks) Jupyter pour rassembler supports et exercices pratiques dans un même document

- Apports didactiques pour apporter des connaissances communes.
- Mises en situation de réflexion sur le thème du stage et des cas concrets.
- Méthodologie d'apprentissage attractive, interactive et participative.
- Equilibre théorie / pratique : 70 % / 30 %.

- Supports de cours fournis au format papier et/ou numérique.
- Ressources documentaires en ligne et références mises à disposition par le formateur.

- Pour les formations en présentiel dans les locaux de Docaposte Institute, les stagiaires sont accueillis dans une salle de cours équipée d'un réseau Wi-Fi, d'un tableau blanc ou paperboard. Un carnet de notes est offert. Un ordinateur avec les logiciels appropriés est mis à disposition (le cas échéant).

Dispositif de suivi de l'exécution et de l'évaluation des résultats de la formation

En amont de la formation

- Recueil des besoins des apprenants afin de disposer des informations essentielles au bon déroulé de la formation (profil, niveau, attentes particulières...).
- Auto-positionnement des apprenants afin de mesurer le niveau de départ.

Tout au long de la formation

- Évaluation continue des acquis avec des questions orales, des exercices, des QCM, des cas pratiques ou mises en situation...

A la fin de la formation

- Auto-positionnement des apprenants afin de mesurer l'acquisition des compétences.
- Evaluation par le formateur des compétences acquises par les apprenants.
- Questionnaire de satisfaction à chaud afin de recueillir la satisfaction des apprenants à l'issue de la formation.

- Questionnaire de satisfaction à froid afin d'évaluer les apports ancrés de la formation et leurs mises en application au quotidien.

Accessibilité

Nos formations peuvent être adaptées à certaines conditions de handicap. Nous contacter pour toute information et demande spécifique.