

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, exécuter
 - 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

- Alternance d'apports théoriques et de mises en pratique contextualisées.

- Progression pédagogique structurée, de la compréhension à la mise en œuvre opérationnelle.
- Pédagogie immersive favorisant une mise en application immédiate des compétences.
- Formation interactive et participative s'appuyant sur des ateliers pratiques, études de cas et échanges collectifs (+ 60 % du temps de formation.)
- Séquences pédagogiques rythmées et évaluations formatives régulières pour renforcer l'ancrage des apprentissages.
- Supports de cours, ressources documentaires et références mis à disposition par le formateur

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

En amont de la formation *(évaluation en ligne dans l'extranet apprenant)*

- Recueil des besoins et des attentes des participants.
- Auto-positionnement permettant d'évaluer le niveau initial des apprenants.

Tout au long de la formation

- Évaluations continues réalisées par le formateur à travers des questions orales, exercices, QCM, études de cas et mises en situation pratiques.

À l'issue de la formation *(évaluations en ligne dans l'extranet apprenant)*

- Auto-positionnement afin de mesurer l'acquisition des compétences visées.
- Validation de l'acquisition des compétences par le formateur
- Questionnaire de satisfaction « à chaud » à l'issue de la formation
- Questionnaire de satisfaction "à froid", 45 jours après la formation afin d'évaluer l'ancrage des acquis et leur mise en application en situation professionnelle.

Accessibilité

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