

Data Science

Data Science avec Spark ML

3 jours (21h00) | ★★★★★ 4,6/5 | BIG-SML | Évaluation qualitative de fin de stage |
Formation délivrée en présentiel ou distanciel ⁽¹⁾

Formations Informatique > Big Data > Data Science



À l'issue de ce stage vous serez capable de :

- Réaliser des applications incluant des implémentations d'algorithmes d'intelligence artificielle en mode Standalone ou clusterisés sur Spark avec Scala.

Niveau requis

Avoir de fortes connaissances en algorithmes d'intelligence artificielle, une maîtrise de Scala et de l'écosystème Hadoop.

Public concerné

Statisticiens, consultants Big Data, data analysts, data scientists.

Cette formation :

- Est animée par un consultant-formateur dont les compétences techniques, professionnelles et pédagogiques ont été validées par des diplômes et/ou testées et approuvées par l'éditeur et/ou par M2i Formation
- Bénéficie d'un suivi de son exécution par une feuille de présence émargée par demi-journée par les stagiaires et le formateur.

(1) Modalité et moyens pédagogique :

Formation délivrée en présentiel ou distanciel * (e-learning, classe virtuelle, présentiel à distance). Le formateur alterne entre méthodes ** démonstrative, interrogative et active (via des travaux pratiques et/ou des mises en situation). La validation des acquis peut se faire via des études de cas, des quiz et/ou une certification.

Les moyens pédagogiques mis en oeuvre (variables suivant les formations) sont : ordinateurs Mac ou PC (sauf pour les cours de l'offre Management), connexion internet fibre, tableau blanc ou paperboard, vidéoprojecteur ou écran tactile interactif (pour le distanciel). Environnements de formation installés sur les postes de travail ou en ligne. Supports de cours et exercices.

* Nous consulter pour la faisabilité en distanciel. ** Ratio variable selon le cours suivi.

Programme

Jour 1

Introduction à Big Data et Apache Spark

- Introduction au Big Data
- Les challenges du Big Data
- Batch vs le temps réel dans le Big Data Analytics
- Analyse en Batch Hadoop
- Vue d'ensemble de l'écosystème
- Les options de l'analyse en temps réel
- Streaming Data - Spark
- In-memory Data - Spark
- Présentation de Spark
- Ecosystème Spark
- Les modes de Spark
- Installation de Spark
- Vue d'ensemble de Spark en cluster
- Spark Standalone cluster
- Spark Web UI

Exemples de travaux pratiques (à titre indicatif)

- *Installation*
- *Configuration*

Jour 2

Les opérations communes sur Spark

- Utilisation de Spark Shell
- Création d'un contexte Spark
- Chargement d'un fichier en Shell
- Réalisation d'opérations basiques sur un fichier avec Spark Shell
- Présentation de l'environnement de développement SBT
- Créer un projet Spark avec SBT
- Exécuter un projet Spark avec SBT
- Le mode local
- Le mode Spark
- Le caching sur Spark
- Persistance distribuée

Exemple de travaux pratiques (à titre indicatif)

- *Data Preprocessing*

Jour 3

Spark Machine Learning

- Introduction au Machine Learning
- Les terminologies communes au Machine Learning
- Applications du Machine Learning
- Machine Learning dans Spark
- Spark ML API
- DataFrames
- Transformateurs et estimateurs
- Les pipelines
- Travailler avec un pipeline
- DAG Pipelines
- La vérification pendant l'exécution
- Passage de paramètres
- General Machine Learning Pipeline
- Sélection de modèles via une validation croisée
- Les types supportés, les algorithmes et les utilitaires
- Les types de données
- Les fonctionnalités d'extraction et les statistiques basiques
- Clustering
- K-Means
- Mettre en place le Clustering en utilisant K-Means
- Gaussian Mixture
- Power Iteration Clustering (PIC)
- Latent Dirichlet Allocation (LDA)
- Le filtrage collaboratif
- Classification
- Régression
- Exemple de régression
- Mettre en place une classification en utilisation la régression
- Linéaire
- Mettre en place un système de recommandations utilisant le filtrage collaboratif

Exemple de travaux pratiques (à titre indicatif)

- *Cas d'usage de Machine Learning sous Spark*

Modalités d'évaluation des acquis

L'évaluation des acquis se fait :

- En cours de formation, par des études de cas ou des travaux pratiques
- Et, en fin de formation, par un questionnaire d'auto-évaluation ou une certification (M2i ou éditeur)