Formation L'apprentissage Machine avec Spark (Spark ML)

Utiliser le Framework de la fondation Apache

N° ACTIVITÉ : 11 92 18558 92 | TÉLÉPHONE : 01 85 77 07 07 | E-MAIL : inscription@hubformation.com

Apparu dans la version 1.2 de Spark, la bibliothèque d'apprentissage automatique (Machine Learning) contient tous les algorithmes nécessaires à l'exploration de données (Data Mining) par apprentissage statistique. L'ensemble des API proposés par Spark ML permet de créer et régler des flux de travaux complets (Pipelines) combinant plusieurs algorithmes d'apprentissage automatique. Durant cette formation de 3 jours, les participants découvriront l'étendue des possibilités offertes par Spark ML et s'approprieront les différentes méthodes de mises en oeuvre pour un déploiement adapté aux besoins de leur entreprise.

Référence	DEC052
Durée	3 jours (21h)
Tarif	2 150 €HT

SESSIONS PROGRAMMÉES

A DISTANCE (FRA)

du 6 au 8 octobre 2025

du 24 au 26 novembre 2025

VOIR TOUTES LES DATES

Objectifs

| Être capable de mettre en oeuvre Spark ML pour de l'apprentissage machine et l'analyse de données
| Identifier comment récupérer les données et les préparer pour un traitement efficient
| Savoir modéliser les données et optimiser les modèles
| Être en mesure d'industrialiser et de déployer une application de prédiction

Public

| Chefs de projet et spécialistes big data souhaitant mettre en oeuvre Spark ML

Prérequis

| Avoir suivi la formation Les bases de l'apprentissage Machine (Machine Learning) (BI105)

Programme de la formation

L'écosystème SPARK et l'apprentissage Big Data
| Enjeux machine learning et Big Data
| L'écosystème Apache Spark
| Les différentes briques de base
| Focus SPARK SQL
| Dataframes et Datasets
| Lab : Mise en oeuvre de l'écosystème SPARK pour l'apprentissage machine Big Data

Le chargement de données d'entraînement massives
| Chargement générique de données
| Chargement de fichiers de formats spécifiques
| Interrogation de bases HIVE
| Interrogation de bases externes
| Lab : Chargement de données de sources diverses sur un cluster SPARK

L'exploration de données d'entraînement massives
| Réalisation de statistiques de base avec SPARK
| Exploitation des librairies graphiques statistiques dans un cadre Big Data
| Lab : Exploration de données d'entraînement sur un cas concret

HUB Formation - 14 avenue de l'Opéra - 75001 Paris
01 85 77 07 07 - www.hubformation.com - SIREN 525 287 785

Le Pipelining
| Le concept de Pipeline Spark
| Les composants d'un Pipeline
| Le fonctionnement d'un Pipeline
| La gestion des paramètres
| Persistance et chargement de Pipelines
| Lab : Création d'un premier pipeline d'apprentissage machine avec SPARK

Le prétraitement et l'ingénierie des variables prédictives
| Extraction de variables prédictives
| Transformation de variables
| Sélection de variables prédictives
| Hachage de variables
| Lab : Prétraitement et ingénierie des variables prédictives sur un cas concret

La création de modèles d'apprentissage Big Data
| Classification de données massives
| Régression de données massives
| Clustering de données massives
| Systèmes de recommandation Big Data
| Règles d'association de données massives
| Lab : Réalisation de modèles d'apprentissage sur des cas concrets Big Data

L'optimisation du réglage des modèles d'apprentissage
| Réglage des hyper-paramètres des modèles
| Validation croisée
| Séparation des données (entraînement, validation)
| Lab : Optimisation du réglage de modèles d'apprentissage sur cas concrets

Déploiement de modèles d'apprentissage Big Data
| Création d'application prédictive en batch
| Création d'application prédictive en streaming
| Mise en oeuvre concrète sur un cluster Big Data
| Bonnes pratiques de déploiement
| Lab : Création d'applications en batch et en streaming sur cas concrets

Méthode pédagogique

Un tour d'horizon de l'écosystème Spark ML. Un focus sur la mise en oeuvre d'algorithmes à l'aide de la brique Machine Learning de Spark qui permet de réaliser de l'apprentissage sur un socle Big Data. Des ateliers pour expérimenter le framework sur des cas concrets.

Méthode d'évaluation

Tout au long de la formation, les exercices et mises en situation permettent de valider et contrôler les acquis du stagiaire. En fin de formation, le stagiaire complète un QCM d'auto-évaluation.

Suivre cette formation à distance

Voici les prérequis techniques pour pouvoir suivre le cours à distance :

| Un ordinateur avec webcam, micro, haut-parleur et un navigateur (de préférence Chrome ou Firefox). Un casque n'est pas nécessaire suivant l'environnement.
| Une connexion Internet de type ADSL ou supérieure. Attention, une connexion Internet ne permettant pas, par exemple, de recevoir la télévision par Internet, ne sera pas suffisante, cela engendrera des déconnexions intempestives du stagiaire et dérangera toute la classe.
| Privilégier une connexion filaire plutôt que le Wifi.
| Avoir accès au poste depuis lequel vous suivrez le cours à distance au moins 2 jours avant la formation pour effectuer les tests de connexion préalables.
| Votre numéro de téléphone portable (pour l'envoi du mot de passe d'accès aux supports de cours et pour une messagerie instantanée autre que celle intégrée à la classe virtuelle).

HUB Formation - 14 avenue de l'Opéra - 75001 Paris
01 85 77 07 07 - www.hubformation.com - SIREN 525 287 785

| Selon la formation, une configuration spécifique de votre machine peut être attendue, merci de nous contacter.
| Pour les formations incluant le passage d'une certification la dernière journée, un voucher vous est fourni pour passer l'examen en ligne.
| Pour les formations logiciel (Adobe, Microsoft Office...), il est nécessaire d'avoir le logiciel installé sur votre machine, nous ne fournissons pas de licence ou de version test.
| Horaires identiques au présentiel.

Accessibilité

Les sessions de formation se déroulent sur des sites différents selon les villes ou les dates, merci de nous contacter pour vérifier l'accessibilité aux personnes à mobilité réduite.
Pour tout besoin spécifique (vue, audition...), veuillez nous contacter au 01 85 77 07 07.