Formation Spark, développer des applications pour le Big Data

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

Vous développerez des applications en Java en vue de traiter en temps réel des données issues du Big Data. Vous collecterez, stockerez et traiterez avec Spark des données de formats hétérogènes afin de mettre en place des chaînes de traitement intégrées à votre Système d'Information.

Référence	SPK
Durée	3 jours (21h)
Tarif	2 330 €HT
Repas	repas inclus

Objectifs

| Maîtriser les concepts fondamentaux de Spark
| Développer des applications avec Spark Streaming
| Faire de la programmation parallèle avec Spark sur un cluster
| Exploiter des données avec Spark SQL
| Avoir une première approche du Machine Learning

Public

| Développeurs
| Architectes

Prérequis

| Bonnes connaissances du langage Java

Programme de la formation

Travaux pratiques
| Mise en pratique des notions vues en cours à l'aide du langage Java.

Présentation d'Apache Spark
| Historique du Framework.
| Les différentes versions de Spark (Scala, Python et Java).
| Comparaison avec l'environnement Apache Hadoop.
| Les différents modules de Spark.
| Travaux pratiques Installation et configuration de Spark. Exécution d'un premier exemple avec le comptage de mots.

Programmer avec les Resilient Distributed Dataset (RDD)
| Présentation des RDD.
| Créer, manipuler et réutiliser des RDD.
| Accumulateurs et variables broadcastées.
| Utiliser des partitions.
| Travaux pratiques Manipulation de différents Datasets à l'aide de RDD et utilisation de l'API fournie par Spark.

Manipuler des données structurées avec Spark SQL
| SQL, DataFrames et Datasets.
| Les différents types de sources de données.
| Interopérabilité avec les RDD.
| Performance de Spark SQL.
| JDBC/ODBC server et Spark SQL CLI.
| Travaux pratiques Manipulation de Datasets via des requêtes SQL. Connexion avec une base externe via JDBC.

SESSIONS PROGRAMMÉES

A DISTANCE (FRA)
du 21 au 23 juillet 2025

PARIS
du 1er au 3 octobre 2025

AIX-EN-PROVENCE
du 21 au 23 juillet 2025

BORDEAUX
du 21 au 23 juillet 2025

LILLE
du 21 au 23 juillet 2025

LYON
du 21 au 23 juillet 2025

NANTES
du 21 au 23 juillet 2025

SOPHIA-ANTIPOLIS
du 21 au 23 juillet 2025

STRASBOURG
du 21 au 23 juillet 2025

TOULOUSE
du 21 au 23 juillet 2025

VOIR TOUTES LES DATES

Spark sur un cluster
| Les différents types d'architecture : Standalone, Apache Mesos ou Hadoop YARN.
| Configurer un cluster en mode Standalone.
| Packager une application avec ses dépendances.
| Déployer des applications avec Spark-submit.
| Dimensionner un cluster .
| Travaux pratiques Mise en place d'un cluster Spark.

Analyser en temps réel avec Spark Streaming
| Principe de fonctionnement.
| Présentation des Discretized Streams (DStreams).
| Les différents types de sources.
| Manipulation de l'API.
| Comparaison avec Apache Storm.
| Travaux pratiques Consommation de logs avec Spark Streaming.

Manipuler des graphes avec GraphX
| Présentation de GraphX.
| Les différentes opérations.
| Créer des graphes.
| Vertex and Edge RDD.
| Présentation de différents algorithmes.
| Travaux pratiques Manipulation de l'API GraphX à travers différents exemples.

Machine Learning avec Spark
| Introduction au Machine Learning.
| Les différentes classes d'algorithmes.
| Présentation de SparkML et MLlib.
| Implémentations des différents algorithmes dans MLlib.
| Travaux pratiques Utilisation de SparkML et MLlib.

Méthode pédagogique

Chaque participant travaille sur un poste informatique qui lui est dédié. Un support de cours lui est remis soit en début soit en fin de cours. La théorie est complétée par des cas pratiques ou exercices corrigés et discutés avec le formateur. Le formateur projette une présentation pour animer la formation et reste disponible pour répondre à toutes les questions.

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).
| 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.

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

| 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.