



FORMATION EN PRESENTIEL ET/OU EN DISTANCIEL (SELON LE THEME)

DÉVELOPPER AVEC SPARK

Dernière mise à jour le 14/10/2024.

OBJECTIF

Cette formation vous permettra de maîtriser les concepts fondamentaux de Spark.

COMPÉTENCES DÉVELOPPÉES

- Découvrir les différents modules de Spark
- Découvrir le Machine Learning avec MLLib
- Manipuler des graphes avec GraphX
- Analyser des données en temps réel avec Spark Streaming
- Comprendre les fondamentaux de Spark
- Manipuler des données avec Spark SQL

PUBLIC CONCERNÉ

Développeurs - Architectes

MODALITÉS ET DÉLAIS D'ACCÈS

Vous pouvez vous inscrire pour suivre une de nos formations jusqu'à la veille de la date de démarrage si la formation est financée directement par votre entreprise ET si le nombre maximum de participants n'est pas atteint. Si la formation est financée via un OPCO, vous devez au préalable avoir obtenu un accord de ce dernier.

MODALITÉS DE DÉROULEMENT DE L'ACTION DE FORMATION

Formation présentielle ou distancielle dispensée par un formateur expérimenté. La formation alterne des exposés théoriques, des démonstrations et la mise en pratique au travers d'exercices et de cas concrets.

Tarif

2190 € HT
par participant

Réf.

DBGSK

Contact

0465260114
contact@univlearn.fr
<https://univlearn.fr/>

PRÉ-REQUIS

- Bonnes connaissances du langage Java ou du langage Python

MÉTHODES ET MOYENS MOBILISÉS

Evaluation des acquis tout au long de la formation à travers des Tps, des Quizz ; Evaluation de satisfaction de fin de formation ; Attestation de fin de formation précisant les modules acquis et en cours d'acquisition ; Support de cours remis en fin de session.



Niveau

**Autres formations
professionnelles continues**



Modalité

Présentiel



Effectif par session

2 mini > 8 maxi



Durée

21 heures / 3 jour(s)



PROGRAMME

Module 1 : PRÉSENTATION D'APACHE SPARK

Introduction et historique

Les langages de programmation supportés par Spark (Scala, Python et Java)

Comparaison avec l'environnement Apache Hadoop

Les modules de Spark (Spark SQL, Spark Streaming, MLlib, GraphX)

Module 2 : RESILIENT DISTRIBUTED DATASET (RDD)

C'est quoi un RDD ?

Créer, manipuler et réutiliser des RDD

Accumulateurs et variables broadcastées

Utiliser des partitions

Module 3 : 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

Module 4 : SPARK SUR UN CLUSTER

Les différents types d'architectures : 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

Module 5 : SPARK STREAMING ET L'ANALYSE EN TEMPS RÉEL DES DONNÉES

C'est quoi Spark streaming ?

C'est quoi Discretized Streams (DStreams)?

Les différents types de sources

Manipulation de l'API

Comparaison avec Apache Storm

Module 6 : MANIPULER DES GRAPHES AVEC GRAPHX

C'est quoi GraphX ?

Création de graphes

Opérations sur les graphes

Vertex and Edge RDD

Différents algorithmes

Module 7 : MACHINE LEARNING AVEC SPARK

Introduction au Machine Learning

Présentation de SparkML et MLlib

Les algorithmes d'apprentissages disponibles

Implémentations de ces différents algorithmes



INTERVENANT(S)

Consultant formateur ayant la double compétence pédagogique et technique

ÉVALUATION

Chaque module de cours est concrétisé par un TP afin de permettre l'acquisition d'un vrai savoir-faire sur tous les points abordés, Questions posées par le formateur tout au long de la formation à l'oral ou à travers un QCM

ATTRIBUTION FINALE

Attestation individuelle de formation