

SPARK

Spark - Traitement de données

REF : SIHA010

DUREE : 21h

Présentiel Classe virtuelle

PUBLIC

Cette formation s'adresse aux :

Développeurs informatiques,

Chefs de projet,

Data Scientists,

Consultants en business intelligence,

Responsables système d'informations.

Modalités et délais d'accès : les inscriptions sont fermées 24h avant la 1ère journée de formation.

Accessibilité : Si vous avez des contraintes particulières liées à une situation de handicap, veuillez nous contacter au préalable afin que nous puissions, dans la mesure du possible, adapter l'action de formation.

PREREQUIS

Connaissance de langage orienté objet (Python, Java, C++).

MODALITES PEDAGOGIQUES

1 poste et 1 support par stagiaire

8 à 10 stagiaires par salle

Remise d'une documentation pédagogique papier ou numérique pendant le stage

La formation est constituée d'apports théoriques, d'exercices pratiques, de réflexions et de retours d'expérience

MODALITES D'EVALUATION

Auto-évaluation des acquis par le stagiaire via un questionnaire en ligne

Attestation de fin de stage remise au stagiaire

OBJECTIFS PEDAGOGIQUES

Cette formation permet de :

- Se former à Spark, l'utiliser en ligne de commande et en développement
- L'interfacer avec d'autres composants Big Data comme Kafka et Cassandra
- Comprendre les notions de base des RDD (Resilient Distributed Datasets) et le partitionnement, le pipelining et les calculs de données
- Comprendre les implications et les optimisations de performance lors de l'utilisation de Spark.

PROGRAMME

Introduction au Big Data

- Apache Spark
- Spark version MapReduce
- Architecture de Spark
- Interagir avec Spark
- Resilient Distributed Datasets (RDD)

Poste de travail

- Différents composants (Kafka, Spark, Cassandra)
- Outils de développement
- Travaux Pratiques

Introduction à Scala

- Les types de données
- La manipulation des données (Map, FlatMap, For)
- La programmation fonctionnelle
- L'approche immutable
- Exercices

Kafka, fonctionnement

- TP Kafka
- Spark Streaming
- Développement Spark Streaming
- Générer une source de données
- Faire un traitement

Présentation SQL

- Concevoir un modèle de données
- Ecrire des requêtes
- Spark SQL
- TP simples

Support Cassandra

- Description rapide de l'architecture Cassandra
- Mise en œuvre depuis Spark.
- Exécution de travaux Spark s'appuyant sur une grappe Cassandra.

Version du : 25/01/2024



