

## Formation Programmer Hadoop en Java : Initiation

<b>Durée :</b>	3 jours
<b>Public :</b>	Développeurs Java, Administrateurs (DBA ou systèmes)
<b>Pré-requis :</b>	Maîtrise de la programmation orientée objets en Java
<b>Objectifs :</b>	- Développer des algorithmes parallèles efficaces - Analyser des fichiers non structurés et développer des tâches Java MapReduce -
<b>Sanction :</b>	Attestation de fin de stage mentionnant le résultat des acquis
<b>Taux de retour à l'emploi:</b>	Aucune donnée disponible
<b>Référence:</b>	BUS101999-F
<b>Note de satisfaction des participants:</b>	Pas de données disponibles

### Comprendre le contexte d'utilisation d'Hadoop

- Évaluer la valeur que peut apporter Hadoop à l'entreprise
- Examiner l'écosystème d'Hadoop
- Choisir un modèle de distribution adapté

### Défier la complexité de la programmation parallèle

- Examiner les difficultés liées à l'exécution de programmes parallèles : algorithmes, échange des données
- Évaluer le mode de stockage et la complexité du Big Data

### Programmation parallèle avec MapReduce

- Fragmenter et résoudre les problèmes à grande échelle
- Découvrir les tâches compatibles avec MapReduce
- Résoudre des problèmes métier courants

### Appliquer le paradigme Hadoop MapReduce

- Configurer l'environnement de développement
- Examiner la distribution Hadoop
- Étudier les démons Hadoop
- Créer les différents composants des tâches MapReduce
- Analyser les différentes étapes de traitement MapReduce : fractionnement, mappage, lecture aléatoire et réduction

### Créer des tâches MapReduce complexes

Choisir et utiliser plusieurs outils de mappage et de réduction, exploiter les partitionneurs et les fonctions map et reduce intégrées, analyser les données en séries temporelles avec un second tri, rationaliser les tâches dans différents langages de programmation

### **Résoudre les problèmes de manipulation des données**

Exécuter les algorithmes : tris, jointures et recherches parallèles, analyser les fichiers journaux, les données des média sociaux et les courriels

### **Mise en œuvre des partitionneurs et des comparateurs**

Identifier les algorithmes parallèles liés au réseau, au processeur et aux E/S de disque  
Répartir la charge de travail avec les partitionneurs  
Contrôler l'ordre de groupement et de tri avec les comparateurs  
Mesurer les performances avec les compteurs