

Fondamentaux de Hadoop et MapReduce

Découvrez les bases de Hadoop et de MapReduce dans cette formation complète.

Catégorie: Big Data · Durée: 2 jours

POUR QUI ?

- Analystes de données et data scientists
- Développeurs de tout horizon
- Accessible aux débutants : aucune expérience préalable requise

Cette formation apporte une compréhension approfondie de Hadoop et de MapReduce, en explorant les concepts et techniques clés d'un traitement de données efficace. Vous mettrez en pratique la mise en place de clusters Hadoop, l'écriture de programmes MapReduce et l'optimisation des flux de traitement de données. À l'issue de cette formation, vous disposerez des compétences nécessaires pour exploiter la puissance de Hadoop et de MapReduce dans l'analyse du big data.

Prérequis

Connaissances de base en programmation et familiarité avec la ligne de commande Linux.

Programme

- Introduction au Big Data et à Hadoop
- HDFS – Hadoop Distributed File System
- Concepts et principes de MapReduce
- Écriture de programmes MapReduce en Java
- Techniques MapReduce avancées
- Installation et configuration d'un cluster Hadoop
- Gestion de données dans Hadoop
- Traitement des données avec Hive
- Traitement des données avec Pig
- Travailler avec HBase
- Introduction à Apache Spark

- Optimisation des jobs MapReduce
- Supervision et débogage des applications Hadoop
- Sécurisation des clusters Hadoop
- Intégration de Hadoop avec les systèmes existants
- Bonnes pratiques de développement Hadoop
- Cas d'usage réels et études de cas
- Réglage et optimisation des performances
- Introduction à YARN – Yet Another Resource Negotiator
- Analyse big data et machine learning avec Hadoop