+33 (0)1 48 10 09 46
formation@ambient-it.net
140 Avenue Jean Lolive 93500 Pantin - France

Mis à jour le 27/07/2023

S'inscrire

Formation Hadoop : Développement

3 jours (21 heures)

Présentation

Hadoop est un framework open source développé par Google et destiné au stockage de données et à l'exécution des applications Hadoop applique des algorithmes "MapReduce (MR)" dans lesquels les données sont traitées en parallèle avec d'autres ensembles de données. Il offre un stockage massif pour tout type de données, une énorme puissance de traitement et la possibilité de traiter un nombre illimité de tâches ou de travaux simultanés. Notre formation Hadoop : Développement vous enseignera les techniques pour traiter de grands volumes de données. Pendant cette formation, vous apprendrez l'écosystème Hadoop, les principes du framework et le développement des algorithmes parallèles avec MapReduce. Vous verrez également comment grâce aux tâches Hadoop, extraire des éléments pertinents de l'ensemble de données et charger des données non structurées des systèmes HBase et HDFS. À l'issue de cette formation, vous saurez développer des applications compatibles avec la plateforme Hadoop d'Apache pour traiter des données Big Data. Comme pour toutes nos formations, celle-ci vous présentera la toute dernière version d'Apache Hadoop 3.3.

Objectifs

- Maîtriser l'écosystème Hadoop Cloudera/Hortonwork
- Mettre en œuvre les fonctionnalités du framework Hadoop
- Extraire des éléments pertinents d'ensembles de données volumineux et variés grâce aux tâches d'Hadoop
- Développer des algorithmes parallèles efficaces avec MapReduce
- Charger des données non structurées des systèmes HDFS et HBase

Public visé

- Développeurs
- Chefs de projets
- Data-scientists
- Architectes

Pré-requis

Avoir la connaissance d'un langage de programmation objet comme Java et du scripting.

Programme de notre formation Hadoop : Développement

Présentation du framework Hadoop

- Installation d'Hadoop
- Objectif du projet Hadoop
- Principes de base du framework
- Fonctionnalités essentielles
- Cas d'utilisation dans les domaines différents
- Plateforme Coudera et Hortonworks

MapReduce

- Implémentation MapReduce par le framework Hadoop
- Principe de programmation MapReduce
- Fonction Map() et Reduce()
- Utiliser des technologies MapReduce
- Développer des algorithmes parallèles efficaces
- Créer, personnaliser et déployer des tâches
- Synthétiser les données avec MapReduce
- Meilleures pratiques de développement des applications MapReduce

L'écosystème Hadoop

- Vue d'ensemble d'écosystème
- Fonctionnalités Hadoop vue d'ensemble
- Architecture d'Hadoop
  - HDFS MapReduce FIL
- Nœud de nom
- Nœud de données
- Nœud de nom secondaire
- Blocs
- Différence entre SGBDR et Hadoop

Hadoop YARN

- Utilisation MapReduce à travers Yarn
- Utilisation d'un cluster
- Gestion de cluster du cloud
- Différentes applications sur le même cluster

- Composants d'YARN

Base de données relationnelle avec Hadoop

- Qu'est-ce qu'Hive
- Syntaxe de base
- Intégration de MySQL à Hadoop
- Simplifier les requêtes
- Extension du HiveQL
- User-Defined-Functions (UDF)
- Utilisation de Sqoop pour importer des données de MySQL vers HFDS/Hive
- Utilisation de Sqoop pour exporter des données de Hadoop vers MySQL

Programmer Hadoop avec Pig

- Définition et utilisation
- Meilleures pratiques map/reduce
- Développement et intégration en Java
- Extension avec UDF

Hadoop avec Spark

- Pourquoi choisir Spark ?
- Architecture de Spark
- Composants essentiels
- Ensembles de données distribuées résilients (RDD)
  - Opérations
  - Persistence
  - Shared Variables
- Fonctions intégrées

Stockage de données sur HDFS

- Système de fichier Hadoop Distributed File System
- Charger des données non structurées de HDFS
- Différents types de donnés XML
- Paralléliser des calculs sur de larges volumes de données
- Fonctionnement en mode distribué

Stockage de données avec HBase

- Charger des données non structurées d'HBase
- Fonctionnement de cluster HBase
- Fonctionnement indépendant
  - HRegionServer
  - HMaster
  - ZooKeeper
- Mécanismes de sécurité en Hadoop
- Gestion de l'authentification

Hadoop Streaming

- Configuration d'Hadoop
- Définition de MapReduce à Streaming
- Langage Python avec Hadoop Streaming
- Créer un job MapReduce en Python
- Suivie d'un job MapReduce en streaming

Pour aller plus loin

Sociétés concernées

Cette formation s’adresse à la fois aux particuliers ainsi qu’aux entreprises, petites ou grandes, souhaitant former ses équipes à une nouvelle technologie informatique avancée ou bien à acquérir des connaissances métiers spécifiques ou des méthodes modernes.

Méthodes pédagogiques

Stage Pratique : 60% Pratique, 40% Théorie. Support de la formation distribué au format numérique à tous les participants.

Organisation

Le cours alterne les apports théoriques du formateur soutenus par des exemples et des séances de réflexions, et de travail en groupe.

Validation

À la fin de la session, un questionnaire à choix multiples permet de vérifier l’acquisition correcte des compétences.

Sanction

Une attestation sera remise à chaque stagiaire qui aura suivi la totalité de la formation.