Hadoop, développer des applications pour le Big Data

Programme *(Mis à jour le 12/02/2021)*

Le Big Data

- Définition du périmètre du Big Data
- Le rôle du projet Hadoop
- Les concepts de base des projets Big Data
- Présentation du Cloud Computing
- Différence entre Cloud Computing privé et public
- Les architectures Big Data à base du projet Hadoop

Collecte de données et application de Map Reduce

- Analyse des flux de données dans l'entreprise
- Données structurées et non-structurées
- Les principes de l'analyse sémantique des données d'entreprise
- Graphe des tâches à base de MapReduce
- La granularité de cohérence des données
- Transfert de données depuis un système de persistance dans Hadoop
- Transfert de données d'un Cloud dans Hadoop

Le stockage des données avec HBase

- Plusieurs types de base de données XML
- Patterns d'usages et application au Cloud
- Application de Hadoop database au sein d'un workflow
- Utilisation des projets Hive/Pig
- Utilisation du projet HCatalog
- L'API Java HBase

Le stockage des données sur HDFS

- Patterns d'usages et application au Cloud
- Architecture et installation d'un système HDFS, journal, NameNode, DataNode
- Opérations, commandes et gestion des commandes
- L'API HDFS Java
- Analyse de données avec Apache Pig
- Le langage Pig Latin. Utiliser Apache Pig avec Java
- Requêtage avec Apache Hive
- Réplication de données. Partage de données sur une architecture HDFS

Spring Data Hadoop

- Introduction à Spring et Spring Data
- Le namespace Hadoop pour Spring
- Utiliser Spring pour simplifier la configuration Hadoop
- Configuration du cache distribué
- Définition des Jobs et dépendance entre Jobs
- Intégration des outils (Pig, Hive…)

Référence

THBI1994

Durée

3 jours / 21 heures

Prix HT / stagiaire

2175€

Objectifs pédagogiques

- Développer un programme à base de Map Reduce
- Intégrer Hadoop HBase dans un workflow d'entreprise
- Manipuler Apache Hive et Pig depuis HDFS
- Utiliser un graphe de tâches avec Hadoop

Niveau requis

- Développer avec un langage de programmation objet comme Java

Public concerné

- Chefs de projets, développeurs, data scientists et toute personne souhaitant comprendre les techniques de développement avec MapReduce dans l'environnement Hadoop

Formateur

Les formateurs intervenants pour Themanis sont qualifiés par notre Responsable Technique Olivier Astre pour les formations informatiques et bureautiques et par Didier Payen pour les formations management.

Conditions d'accès à la formation

Délai : 3 mois à 1 semaine avant le démarrage de la formation dans la limite des effectifs indiqués

Moyens pédagogiques et techniques

Salles de formation (les personnes en situation de handicap peuvent avoir des besoins spécifiques pour suivre la formation. N'hésitez pas à nous contacter pour en discuter) équipée d'un ordinateur de dernière génération par stagiaire, réseau haut débit et vidéo-projection UHD
Documents supports de formation projetés
Apports théoriques, étude de cas concrets et exercices
Mise à disposition en ligne de documents supports à la suite de la formation

Dispositif de suivi de l'éxécution de l'évaluation des résultats de la formation

Feuilles d'émargement (signature électronique privilégiée)
Evaluations formatives et des acquis sous forme de questions orales et/ou écrites (QCM) et/ou mises en situation
Questionnaires de satisfaction (enquête électronique privilégiée)