

Big Data - Développement

Talend pour le Big Data

3 jours (21h00) | ★★★★★ 4,6/5 | BIG-TAL | Code Certif Info : 100085 | Certification M2i
Expertises Big Data (non incluse) | Évaluation qualitative de fin de stage | Formation délivrée
en présentiel ou distanciel ⁽¹⁾

Formations Informatique > Big Data > Big Data - Développement



À l'issue de ce stage vous serez capable de :

- Utiliser Talend dans un environnement Big Data.

Niveau requis

Avoir des connaissances en Hadoop, Spark et Kafka serait un plus pour la formation.

Public concerné

Consultants BI, architectes SI, chefs de projets.

Cette formation :

- Est animée par un consultant-formateur dont les compétences techniques, professionnelles et pédagogiques ont été validées par des diplômes et/ou testées et approuvées par l'éditeur et/ou par M2i Formation
- Bénéficie d'un suivi de son exécution par une feuille de présence émarginée par demi-journée par les stagiaires et le formateur.

(1) Modalité et moyens pédagogique :

Formation délivrée en présentiel ou distanciel * (e-learning, classe virtuelle, présentiel à distance). Le formateur alterne entre méthodes ** démonstrative, interrogative et active (via des travaux pratiques et/ou des mises en situation). La validation des acquis peut se faire via des études de cas, des quiz et/ou une certification.

Les moyens pédagogiques mis en oeuvre (variables suivant les formations) sont : ordinateurs Mac ou PC (sauf pour les cours de l'offre Management), connexion internet fibre, tableau blanc ou paperboard, vidéoprojecteur ou écran tactile interactif (pour le distanciel). Environnements de formation installés sur les postes de travail ou en ligne. Supports de cours et exercices.

* Nous consulter pour la faisabilité en distanciel. ** Ratio variable selon le cours suivi.

Programme

Concepts de bases

- Ouvrir un projet
- Monitorer un cluster Hadoop
- Créer un cluster de métadonnées

Lire et écrire des données en HDFS

- Stocker un fichier sur HDFS
- Stocker plusieurs fichiers de HDFS
- Lire les données de HDFS
- Utiliser HBase pour sauvegarder les données charger dans HDFS

Travailler avec les tables

- Importation de tables avec Sqoop
- Création de tables dans HDFS avec Hive

Traitement des données et des tables en HDFS

- Traitement des tables Hive avec des jobs
- Traitement des données avec Pig
- Traitement des données par lots

Guide de dépannage

- Dépannage de votre cluster

Cas d'utilisation de clickstream (flux de clics)

- Surveillance du cluster Hadoop
- Créer un environnement de développement
- Chargement des données dans HDFS
- Enrichissement des logs
- Calculer les statistiques
- Conversion d'un job standard en un lot Big Data
- Comprendre les jobs MapReduce
- Utilisation du studio pour configurer les resource requests vers YARN

Cas d'utilisation : l'analyse comportementale

- Chargement du dictionnaire et des données du fuseau horaire dans HDFS
- Chargement des tweets dans HDFS
- Traitement des tweets avec MapReduce
- Planification de l'exécution du job

Introduction à Kafka

- Surveillance du cluster Hadoop
- Comprendre les bases de Kafka
- Publication de messages sur un sujet Kafka
- Consommer des messages

Introduction à Spark

- Comprendre les bases de Spark
- Analyser les données des clients
- Produire et consommer des messages en temps réel

Cas d'utilisation de traitement des logs : génération des logs enrichis

- Introduction au cas d'utilisation du traitement des logs
- Génération de logs bruts
- Génération de logs enrichis

Traitement des logs cas d'utilisation : surveillance

- Surveillance des logs enrichis

Cas d'utilisation de traitement des logs : rapports

- Génération de rapports basés sur les fenêtres de données

Cas d'utilisation de traitement des logs : analyse des batchs

- Ingestion de flux de données
- Analyser les logs avec un batch job

Certification (en option)

- Prévoir l'achat de la certification en supplément
- L'examen (en français) sera passé le dernier jour, à l'issue de la formation et s'effectuera en ligne
- Il s'agit d'un QCM dont la durée moyenne est d'1h30 et dont le score obtenu attestera d'un niveau de compétence

Modalités d'évaluation des acquis

- En cours de formation, par des études de cas ou des travaux pratiques
- Et, en fin de formation, par un questionnaire d'auto-évaluation ou une certification (M2i ou éditeur)

Compétences visées

- Gérer des projets Big Data et communiquer auprès des directions et des utilisateurs
- Maîtriser les technologies, les langages et les outils du Big Data
- Développer de nouveaux processus d'importation pour améliorer la qualité des données
- Appliquer le Big Data dans différents domaines d'activités.