FORMATION · DATA MANAGEMENT

Talend Big Data

Durée de la formation : 3 jours
Tarif : 1850 €
Réf :BI100
Délai d'accès : 2 à 4 semaines

Prérequis

Avoir suivi la formation Talend DI et avoir des notions en BIG DATA.

Public concerné

Gestionnaires de données, architectes, consultants en informatique décisionnelle.

Objectifs pédagogiques

Vous connecter à un duster Hadoop depuis un Job Talend. Utiliser des variables de contextes et des métadonnées. Lire et écrire des fichiers dans HDFS ou dans HBase, dans des lobs Big Data Batch ou Big Data Streaming. Utiliser l'API Twitter avec les composants Talend Lire et écrire des messages dans un topic Kafka en temps réel.Lire et écrire des messages dans un topic Kafka en temps réel.Configurer un Job Big Data Batch.

Objectifs opérationnels

Savoir utiliser VARN. Connaître le fonctionnement de Spark. Être opérationnel sur le traitement des logs. Travail en autonomie. Maîtriser le module.

Méthodes mobilisées et modalités d'évaluation

L'évaluation se fait en fin de parcours sous forme de quizz et d'étude de cas.

Talend Studio ESB Bases

Talend fournit un environnement de développement qui permet une interaction avec de nombreuses sources et destinations Big Data sans avoir à apprendre et écrire du code compliqué. Cette formation traite des Job Big Data Batch utilisant le framework MapReduce ou Spark. Elle couvre ensuite les Jobs Big Data Streaming utilisant le framework Spark Streaming.

Inscription

Contactez-nous au 01 86 95 71 60 ou à l'adresse formation@adbi.fr pour connaitre les dates de notre prochaine session

Modes de participation

Nos formations sont proposées sous 2 formats :

- Chez nous prêt de place d'Italie
- Chez vous dans vos locaux

Accessibilité

Les personnes en situation de handicap sont invitées à nous communiquer leurs besoins spécifiques. Nous ferons tout pour les mettre dans les meilleures conditions de suivi de la formation (compensation, accessibilité…)

Attestation et certification

Une attestation de fin de formation sera remise à tous les participants à l'issue de leur parcours

- **Big Data Advanced YARN**
    - Cas d'utilisation de flux de dics
    - Configuration de l'environnement de développement
    - Chargement des données dans HDFS
    - Enrichissement des logs
    - Calcul de statistiques
    - Conversion d'un Job standard en Job Big Data Batch
    - Compréhension d'un Job Map Reduce
    - Utilisation du Studio pour configurer les demandes de ressources à YARN
- **Cas d'utilisation d'analyse des sentiments**
    - Chargement des données de dictionnaire et de fuseaux horaires
    - Chargement des tweets dans HDFS
    - Traitement des tweets avec Map Reduce
    - Planification des exécutions de Jobs
- **Big Data Advanced Spark**
  Introduction à Kafka
    - Présentation des bases de Kafka
    - Publication de messages dans un topic Kafka
    - Consommation de messages
- **Introduction à Spark**
    - Présentation des bases de Spark
    - Analyse des données clients
    - Production et consommation des messages en temps réel
- **Scénario de traitements des logs – Génération de logs enrichis**
    - Introduction au scénario de traitements des logs
    - Génération de logs bruts
    - Génération de logs enrichis
    - Scénario de traitements des logs – Monitoring
        - Monitoring des logs enrichis
    - Scénario de traitement des logs – Reporting
        - Reporting d'informations utilisateurs
    - Scénario de traitements des logs – Analyse
        - Ingestions de flux de données
        - Analyse des logs avec un Job Batch
- **Scénario de classification (facultatif)**
    - Présentation de l'apprentissage automatique
    - Création d'un modèle de classification de SMS
    - Test du modèle de classification de SMS