Big Data : synthèse technique

Lien : https://innov-systems.com/formation/big-data-synthese-technique

DURÉE
3 jours (21h)

RÉFÉRENCE
BSI04

CATÉGORIE
Big Data : fondamentaux et architecture

Innov Systems

OBJECTIFS DE LA FORMATION

À l'issue de cette formation, vous serez capable de :

- Découvrir les principaux concepts du Big Data
- Comprendre les spécificités propres aux projets Big Data
- Apprendre à mettre en œuvre une plateforme complète pour gérer vos données
- Travailler avec Hadoop, Spark et Talend pour la datavisualisation

POUR QUI ?

- Dataminers
- Chargés d'études statistiques
- Développeurs
- Chefs de projet

Programme détaillé

1 / Comprendre le Big Data

- Définitions, historique
- Concepts fondamentaux et caractéristiques clés du Big Data
- Les origines du Big Data
- La provenance des données
- Changements de quantité, de qualité, d'habitudes
- La valeur de la donnée : un changement d'importance
- La donnée en tant que matière première
- Le quatrième paradigme de la découverte scientifique
- 3V de Gartner (Vélocité, Variété et Volume) et les variantes (Véracité, Valeur, Validité…)
- Collecte et traitement des données structurées, semi-structurées et non-déstructurées
- Transformation des données en informations
- Création de la valeur à partir des données
- Différences et apports par rapport aux technologies traditionnelles
- Exemples d'application et usages des technologies Big Data

2 / L'implémentation d'un projet Big Data : méthodologie

- Les besoins métiers et objectifs d'un projet
- Spécificités d'un projet Big Data
- Sélectionner les bons outils
- Poser les bonnes questions
- Obtenir les données
- Explorer et prétraiter les données
- Analyser les données : planifier et construire un modèle viable
- Communiquer les résultats
- Transformer les résultats en décisions et en actions

Inov Systems

3 / Technologies et solutions associées au Big Data

- Architecture et composants de la plateforme Hadoop 2
- Les modes de stockage (NoSQL, HDFS)
- Fonctionnement de MapReduce et Yarn…
- Principales distributions Hadoop : Hortonworks, Cloudera, MapR…
- Les technologies émergentes : Spark, Storm, Machine Learning Azure…
- Installation d'une plateforme Hadoop
- Panorama des solutions de datavisualisation pour le Big Data (Talend, Tableau, Qlikview…)

4 / Collecte, stockage et traitement des données

- Connaître les différentes sources de données (internes/externes, publiques/privées)
- Analyser les caractéristiques d’un jeu de données
- Créer des flux de données massives avec un ETL (Extract Transform Load) : présentation et prise en main de Talend
- Gérer les spécificités des données semi et non-structurées
- Rappels des principes du stockage distribué
- Bases de données NoSQL (Cassandra, Neo4j, MongoDB…)
- Prise en main d’HDFS et du modèle MapReduce
- Requêter et traiter les données avec Pig et Hive
- ElasticSearch pour l’indexation et la recherche de données

5 / Analyse et traitements des données issues du Big Data

- Différentes méthodes et techniques d’analyse
- Notions d’analyse statistique et prédictive
- L’impact de l’augmentation des volumes de données
- Analyser un ensemble de données avec l’environnement Apache Spark
- Comprendre les apports de la convergence entre Big Data et intelligence artificielle
- Introduction aux principes de Machine Learning
- Les familles de modèles : régression, classification, clustering, reinforcement learning, NLP
- Technique de feature engineering pour préparer les données

Innov Systems

- Sélectionner, entraîner et tester un algorithme de Machine Learning
- Principes du Deep Learning : apprentissage profond et réseaux neuronaux

6 / Data visualisation : Apprendre à communiquer sur les données

- Les enjeux et objectifs d'une bonne communication
- Principes fondamentaux de la représentation visuelle
- La Datavisualisation tout au long du processus Big Data
- Techniques et outils principaux pour la visualisation de données (Tableau, Qlik, etc)
- Sélectionner une solution de datavisualisation selon l'objectif recherché
- Bonnes pratiques de Dataviz
- Visualisation interactive

Approche pédagogique

- Support Ecrit et Projection
- Exposés Interactifs, Podcasts et Vidéos
- Brainstorming et Jeux de Rôle
- Cas Pratiques et Labs inclus pour leur impact opérationnel
- Test de Validation des Acquis des Connaissances

Prochaines dates programmées

24 au 26 Fév. 2027	Présentiel - Casablanca
21 au 23 Avr. 2027	Présentiel - Casablanca
16 au 18 Juin 2027	Présentiel - Casablanca
18 au 20 Août 2027	Présentiel - Casablanca
20 au 22 Oct. 2027	Présentiel - Casablanca

Autres dates possibles sur demande. Contactez-nous pour organiser une session intra-entreprise.

Inov Systems

Réservation & Renseignements

Téléphone : +212 522 247 210
Email : contact@innov-systems.com
Web : https//www.innov-systems.com

Innov Systems — Tous droits réservés

Innov Systems