Big Data, développement d’applications de DataVisualisation

Lien :
https://innov-systems.com/formation/big-data-developpement-dapplications-de-datavisualisation

DURÉE
5 jours (35h)

RÉFÉRENCE
BSI101

CATÉGORIE
Big Data - Analyse des Données et Datavisualisation

Innov Systems

OBJECTIFS DE LA FORMATION

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

- Comprendre les fondamentaux du Big Data
- Développer des applications avec Spark Streaming
- Utiliser la programmation parallèle sur un cluster
- Exploiter des données avec Spark SQL
- Comprendre et exploiter les algorithmes standard de Machine Learning

POUR QUI ?

- Data analysts
- Data scientists
- Experts en recherche opérationnelle
- Développeurs

Programme détaillé

1 / Comprendre le Big Data

- Les origines du Big Data
- Un déluge de données
- Données internes et externes
- Les dimensions en V du Big Data : les 3V (Vélocité, Variété et Volume)
- Les technologies essentielles
- Les grands acteurs (éditeurs, intégrateurs) et l'Opensource

2 / Aperçu des frameworks et des outils du Big Data

- Limites des outils et des paradigmes classiques
- Historique des technologies Big Data : Google's GFS & MapReduce
- Principaux Frameworks du Big Data : Hadoop, Spark, Storm, etc
- Classification des outils du Big Data selon leurs cas d'usage :
- Stockage et traitement de la donnée
- Les bases de données et la gestion des données
- La sérialisation
- Le management et le monitoring
- L'analyse et la visualisation des données
- Le transfert des données
- La sécurité et le contrôle d'accès
- Solutions Big Data sur le Cloud
- Les Frameworks et les outils Big Data utilisés : avantages, limites, évolutions

3 / Rappel sur les langages

- Python (application de la MapReduce)
- Les mathématiques de base (algèbre, statistiques, probabilités)

Innov Systems

- SQL

4 / Hadoop

- Comprendre Hadoop et ses composants
- HDFS : Hadoop distributed file system
- MapReduce : Traitement simplifié de données sur des grands clusters
- YARN : Gestionnaire de ressources du cluster Hadoop
- Différences entre les distributions (Hortonworks, Cloudera, MapR) pour le développeur
- Différentes catégories de bases NoSQL (clé/valeur, documents, colonnes, graphes)
- Elasticsearch : Indexer et rechercher des données
- Les visualiser à l'aide de KIBANA
- Quand utiliser le couplage Elasticsearch, Logstash, Kibana (ELK) ?
- Le moteur de recherche SolR
- SAS VA et autres solutions mixtes Cloud/On Premice pour explorer vos données
- IBM Watson (fédération des informations)
- Solution BI Classique

5 / Les algorithmes et analyse de données

- Les concepts du Machine Learning
- Les données supervisées ou non supervisées ? Quelles différences ?
- Régression
- Modèles linéaires
- Classification
- Naive Bayes
- K-NN
- K-Means clustering
- Réseaux de neurones
- Arbres de décisions et de régression
- Support vector machines
- Méthodes ensemblistes

Innov Systems

6 / Le développement (Mapreduce, Spark)

- MapReduce : Philosophie et contraintes
- Briques logicielles pour le développeur
- Exemple de pseudo-code pour les opérations map et reduce
- Limites de MapReduce et émergence de Spark
- Les différentes versions de Spark (Scala, Python et Java)
- Des traitements en mémoire et tolérants aux pannes RDD
- Les modes de travail en cluster de Spark

7 / Machine Learning avec : Spark et Spark Streaming

- Introduction à Apache Spark
- Faire du Machine Learning avec Apache Spark MLlib
- Traiter les données en temps réel
- Requêter en format SQL avec Apache Spark SQL
- Manipuler des graphes avec GraphX
- Stockage basé sur les graphes

8 / La visualisation des données

- La valeur de la visualisation
- Modèles de données et images
- Données multidimensionnelles
- Principes de design
- Perception graphique

9 / Restitution et pilotage

- Couleurs
- Itérations sur la perception
- Techniques d'interactions
- Visualisation Big Data

Innov Systems

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

18 au 22 Jan. 2027	Présentiel - Casablanca
15 au 19 Mars 2027	Présentiel - Casablanca
10 au 14 Mai 2027	Présentiel - Casablanca
05 au 09 Juil. 2027	Présentiel - Casablanca
30 Août au 03 Sep. 2027	Présentiel - Casablanca

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

Réservation & Renseignements

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

Innov Systems — Tous droits réservés