Durée de formation
4 jours (28 heures)

Participants
Développeurs, Consultant business intelligence, Chef de projet, Statisticiens.

Pré-requis
Connaissances de base des modèles relationnels, des statistiques, des langages de programmation et des concepts de la Business Intelligence.

Coût: 2000 € HT

Dates des sessions

A définir

PRESENTATION DE LA FORMATION

L'accroissement continuel des données numériques dans les organisations a conduit à l'émergence du Big Data. Ce concept recouvre les questions de stockage mais aussi celles liées au gisement potentiel de valeur que représentent ces masses de données. Cette formation vous permettra de comprendre les enjeux et les apports du Big Data ainsi que les technologies pour le mettre en œuvre. Vous apprendrez à intégrer des volumétries massives de données structurées et non structurées, puis à les analyser grâce à des modèles statistiques et des dashboards dynamiques.

OBJECTIFS PÉDAGOGIQUES

- Comprendre les concepts du Big Data par rapport aux enjeux métiers.
- Mettre en place l'écosystème nécessaire à l'implémentation d'un projet Big Data
- Acquérir les compétences techniques pour gérer des flux de données complexes, non structurés et massifs.
- Appréhender un outil de data visualisation.

1) Définition du big data

- Les principes d'un outil d'infrastructure as code
- Les différents providers

2) L'écosystème Hadoop

- Description de l'architecture et des composants Hadoop.
- Les modes de stockage (NoSQL, HDFS)
- Principes de fonctionnement de MapReduce et Spark.
- Principales distributions de Hadoop.
- Installer une plateforme Hadoop.

3) Les modes de stockage du big data

- Les bases de données NoSQL
- Les bases de données Clé-Valeur : Redis
- Les Bases de données Document : MongoDB
- Les bases de données colonnes : Cassandra et HBase
- Les bases de données Graphes

4) L'écosystème SPARK

- Les différents modes de travail avec Spark
- Les trois systèmes de gestion de cluster
- Modes d'écriture des commandes Spark
- Les quatre API Langage de Spark
- Le machine learning avec Spark
- Spark SQL
- Le moteur d'exécution SQL
- La création d'une session Spark
- Spark Dataframes
- Spark ML
- L'API pipeline
- Travail sur les variables prédictives
- Classification, régression, clustering et filtrage coopératif

5) Traitement en flux : KAFKA et NIFI

- Architectures types de traitement de Streams Big Data
- Apache NIFI
- Apache KAFKA
- Articulation NIFI et KAFKA (NIFI ON KAFKA)
- Apache STORM
- Articulation KAFKA et STORM (KAFKA ON STORM)
- Apache SPARK Streaming et Structured Streaming
- Articulation KAFKA et SPARK

6) Data visualisation

- Définition de besoin de la data visualisation.
- Analyse et visualisation des données.
- Les outils DataViz du marché.

7) Gouvernance des données

- Challenges Big Data pour la gouvernance des données
- L'écosystème des outils de gouvernance Big Data
- Les piliers de la gouvernance Big Data
- Mise en perspective dans une architecture Big Data
- Management de la qualité des données Big Data
- Tests ²de validation de données dans Hadoop

METHODES PÉDAGOGIQUES

Stage Pratique : 60% Pratique, 40% Théorie
Support de la formation distribué au format numérique à tous les participants

ORGANISATION

Le cours alterne les apports théoriques du formateur soutenus par des exemples et des séances de réflexions, et de travail en groupe.

VALIDATION

À la fin de la session, un questionnaire à choix multiple permet de vérifier l'acquisition correcte des compétences.

SANCTION

Une attestation sera remise à chaque stagiaire qui aura suivi la totalité de la formation.
Nous offrons également la possibilité de préparer la certification