

BIG DATA – SYNTHÈSE TECHNIQUE

Durée : 2 jours (14 heures)

CONNAISSANCES PREALABLES

- Connaissances générales des bases de données relationnelles
- Compréhension des architectures informatiques distribuées
- Connaissances de base en SQL
- Une première expérience dans le domaine Data ou décisionnel est recommandée.

PROFIL DES STAGIAIRES

- Architectes Data
- Data Engineers
- Administrateurs de bases de données
- Administrateurs systèmes et infrastructures
- Consultants Data et BI
- Chefs de projets Data
- Responsables techniques souhaitant comprendre les architectures Big Data
- Développeurs impliqués dans des projets de traitement de données massives.

OBJECTIFS

À l'issue de cette formation, les participants seront capables de :

- Comprendre les concepts fondamentaux du Big Data
- Identifier les architectures et technologies majeures du marché
- Différencier les traitements Batch et Streaming
- Comprendre les principes des plateformes Hadoop et Spark
- Identifier les différents modèles de stockage Big Data
- Concevoir une architecture Data adaptée aux besoins métiers
- Comprendre les enjeux de gouvernance, de sécurité et de performance
- Évaluer les technologies adaptées à un projet Big Data.

CERTIFICATION PREPAREE

Aucune

METHODES PEDAGOGIQUES

- Présentations illustrées par des retours d'expérience
- Démonstrations d'architectures Big Data
- Études de cas réels
- Ateliers de conception d'architectures
- Travaux pratiques de découverte des composants Big Data

- Échanges autour des retours terrain et bonnes pratiques

FORMATEUR

- Consultant expert Big Data et architectures Data modernes, spécialisé dans les plateformes distribuées, les traitements massifs de données et les écosystèmes Cloud Data.

METHODE D'EVALUATION DES ACQUIS

- Quiz de validation des connaissances
- Exercices pratiques
- Études de cas
- Ateliers d'architecture
- Évaluation continue par le formateur
- Validation des acquis en fin de formation.

CONTENU DU COURS

Module 1 – Comprendre les fondamentaux du Big Data (2h)

Objectifs

- Comprendre les enjeux du Big Data
- Identifier les évolutions des architectures de données

Contenu

- Définition du Big Data
- Les 5V du Big Data :
 - Volume
 - Vélocité
 - Variété
 - Véracité
 - Valeur
- Limites des architectures décisionnelles traditionnelles
- Évolution des usages Data
- Cas d'usage métiers :
 - Marketing
 - Industrie
 - IoT
 - Finance
 - Cybersécurité

Mise en pratique

- Brainstorming sur les usages Big Data dans l'entreprise
- Analyse de cas réels de transformation Data
- Identification des problématiques de volumétrie

Module 2 – Architectures Big Data et écosystèmes technologiques (2h)

Objectifs

- Comprendre les architectures distribuées
- Identifier les principaux composants d'une plateforme Big Data

Contenu

- Principes des architectures distribuées
-

- Scalabilité horizontale et verticale
- Stockage distribué
- Calcul distribué
- Présentation des architectures Data modernes
- Data Lake
- Data Warehouse moderne
- Data Lakehouse
- Data Mesh
- Data Fabric

Mise en pratique

- Analyse comparative de plusieurs architectures
- Atelier de choix d'architecture selon un besoin métier

Module 3 – Hadoop et les fondements du traitement distribué (2h)**Objectifs**

- Comprendre le fonctionnement de l'écosystème Hadoop
- Identifier les principaux composants

Contenu

- Historique et principes Hadoop
- HDFS
- YARN
- MapReduce
- Hive
- HBase
- Sqoop
- Oozie
- Limites de l'écosystème Hadoop classique

Mise en pratique

- Analyse d'une architecture Hadoop
- Étude d'un flux de traitement distribué
- Cartographie des composants

Module 4 – Apache Spark et les traitements Big Data modernes (2h)**Objectifs**

- Comprendre l'évolution vers Spark
- Identifier les avantages des traitements en mémoire

Contenu

- Présentation d'Apache Spark
- Architecture Spark
- Spark SQL
- DataFrames
- Spark Streaming
- MLlib
- GraphX
- Comparaison Hadoop / Spark

Mise en pratique

- Démonstration d'un traitement Spark
- Analyse d'un pipeline de données
- Étude comparative de performances

Module 5 – Stockage et bases de données Big Data (2h)

Objectifs

- Comprendre les différents modèles de stockage
- Choisir une solution adaptée aux besoins

Contenu

- Limites du modèle relationnel
- Bases NoSQL :
 - Clé/Valeur
 - Colonnes
 - Documents
 - Graphes
- MongoDB
- Cassandra
- Elasticsearch
- Redis
- Object Storage
- Formats de données :
 - Parquet
 - ORC
 - Avro

Mise en pratique

- Analyse de scénarios de stockage
- Choix des technologies adaptées selon les cas d'usage

Module 6 – Traitements temps réel et streaming (1h30)

Objectifs

- Comprendre les architectures temps réel
- Identifier les usages du Streaming Data

Contenu

- Batch versus Streaming
- Architecture événementielle
- Apache Kafka
- Kafka Connect
- Apache Flink
- Spark Streaming
- Cas d'usage IoT et temps réel

Mise en pratique

- Analyse d'un pipeline temps réel
- Étude de cas sur des flux événementiels

Module 7 – Gouvernance, sécurité et industrialisation des données (1h30)

Objectifs

- Comprendre les enjeux de gouvernance Big Data
- Sécuriser les plateformes de données

Contenu

- Gouvernance des données
- Qualité des données
- Catalogage et métadonnées
- Data Lineage
- Sécurité des données

- Gestion des accès
- Conformité RGPD
- FinOps et maîtrise des coûts

Mise en pratique

- Atelier de définition d'une gouvernance Data
- Analyse des risques de sécurité

Module 8 – Big Data, Cloud et architectures de nouvelle génération (0h30)**Objectifs**

- Comprendre l'évolution des plateformes Data modernes
- Identifier les services Cloud adaptés

Contenu

- Big Data dans le Cloud
- Services managés :
 - AWS
 - Azure
 - GCP
- Modern Data Platform
- DataOps
- MLOps
- Tendances du marché

Mise en pratique

- Panorama comparatif des plateformes Cloud Data

Module 9 – Atelier de synthèse : concevoir une architecture Big Data (0h30)**Objectifs**

- Mettre en œuvre l'ensemble des connaissances acquises
- Concevoir une architecture répondant à un besoin métier

Contenu

- Analyse d'un besoin métier
- Choix des composants techniques
- Définition de l'architecture cible
- Identification des contraintes

Mise en pratique

- Cas fil rouge complet
- Présentation des architectures proposées
- Débriefing collectif et recommandations

Notre référent handicap se tient à votre disposition au [01.71.19.70.30](tel:01.71.19.70.30) ou par mail à <mailto:referent.handicap@edugroupe.com> pour recueillir vos éventuels besoins d'aménagements, afin de vous offrir la meilleure expérience possible.