

Durée : 2 jours soit 14 heures

Référence : IF-BGD

**Public visé :**

- Chefs de projets, architectes, développeurs, data scientists et toute personne souhaitant connaître les outils et solutions pour concevoir et mettre en oeuvre une architecture Big Data

**Pré-requis :**

- Connaissances de base du système GNU/Linux

**Objectifs pédagogiques :**

- Analyser le principe du Big Data avec les technologies phares
- Définir les technologies : le cluster Hadoop, MongoDB, Neo4J, InfluxDB, ElasticSearch
- Analyser les difficultés propres à un projet Big Data

**Modalités pédagogiques :**

Session dispensée en présentiel ou téléprésentiel, selon la modalité inter-entreprises ou intra-entreprises sur mesure.

La formation est animée par un(e) formateur(trice) durant toute la durée de la session et présentant une suite de modules théoriques clôturés par des ateliers pratiques validant l'acquisition des connaissances. Les ateliers peuvent être accompagnés de Quizz.

L'animateur(trice) présente la partie théorique à l'aide de support de présentation, d'animation réalisée sur un environnement de démonstration.

En présentiel comme en téléprésentiel, l'animateur(trice) accompagne les participants durant la réalisation des ateliers.

**Moyens et supports pédagogiques :**

**Cadre présentiel**

Salles de formation équipées et accessibles aux personnes à mobilité réduite.

- Un poste de travail par participant
- Un support de cours numérique ou papier (au choix)
- Un bloc-notes + stylo
- Vidéoprojection sur tableau blanc
- Connexion Internet
- Accès extranet pour partage de documents et émargement électronique

**Cadre téléprésentiel**

Session dispensée via notre solution iClassroom s'appuyant sur Microsoft Teams.

- Un compte Office 365 par participant
- Un poste virtuel par participant
- Un support numérique (PDF ou Web)
- Accès extranet pour partage de documents et émargement électronique

**Modalités d'évaluation et suivi :**

**Avant**

Afin de valider le choix d'un programme de formation, une évaluation des prérequis est réalisée à l'aide d'un questionnaire en ligne ou lors d'un échange avec le formateur(trice) qui validera la base de connaissances nécessaires.

**Pendant**

Après chaque module théorique, un ou des ateliers pratiques permettent la validation de l'acquisition des connaissances. Un Quizz peut accompagner l'atelier pratique.

**Après**

Un examen de certification si le programme de formation le prévoit dans les conditions de l'éditeur ou du centre de test (TOSA, Pearson Vue, ENI, PeopleCert)

**Enfin**

Un questionnaire de satisfaction permet au participant d'évaluer la qualité de la prestation.

**Description / Contenu**

**Module 1 : INTRODUCTION**

- La philosophie NoSQL
- Différences entre bases relationnelles et NoSQL
- Architecture distribuée
- Le format JSON comme format d'échange
- Exemples de bases NoSQL :
  - Base orientée clés/valeurs : Cassandra
  - Base orientée « colonnes » : Hadoop
  - Base volatile : memcache
  - Base orientée « documents » : MongoDB
  - Base de données orientées graphes
  - Base de données temps-réel
  - Base de données avec recherche "full-text"

**Module 2 : PRESENTATION DE NEO4J**

- (Base de données orientée graphes)
- Installation rapide
- Langage d'interrogation
- Prise en main avec TP

**Module 3 : PRESENTATION DE INFLUXDB**

- (Base de données temps-réel)
- Installation rapide
- Langage d'interrogation
- Prise en main avec TP

**Module 4 : PRESENTATION DE ELASTICSEARCH**

- (Base de données de recherche full-text)
- Installation rapide
- Prise en main avec TP



- Présentation de Kibana
- Langage de requêtage

**Module 5 : PRISE EN MAIN DE HADOOP-v2**

- Mise en œuvre d'un Cluster
- Le gestionnaire YARN
- Utilisation de HDFS
- 1er exemple de MapReduce

**Module 6 : ECRITURE D'UN ALGO.MAP/REDUCE**

- Application Java
- Les différentes étapes
- Utilisation d'autres langages

**Module 7 : APPLICATIONS CONNEXES**

- Présentation de HBase
- Présentation de Hive
- Présentation de Pig
- Collecte de données avec FLUME
- Présentation et mise en œuvre de Spark

**Module 8 : PRISE EN MAIN DE MONGODB**

- Concepts de base
- Documents
- Types intégrés
- Collections
- Bases de données
- Manipulation de documents
- Indexation

**Module 9 : REQUETES D'AGREGATION**

- Agrégation simple
- Algorithme MapReduce
- Framework d'agrégation
- Pipelines

**Module 10 : REPLICATION ET CLUSTER**

- Principe
- Mise en œuvre d'un Replica Set
- Définition du sharding et des chunks
- Mise en œuvre d'un cluster

**Module 11 : PRISE EN MAIN D'ELASTICSEARCH**

- Principe de fonctionnement
- Indexation
- Requête
- L'interface Kibana