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

Compétences acquises à l'issue de la formation :

- Analyser le principe du Big-Data avec les technologies phares
- Définir les technologies : le cluster Hadoop, MongoDB, Neo4J (ou OrientDB), InfluxDB, ElasticSearch
- Analyser les difficultés propres à un projet Big Data, les freins et les apports, tant sur les aspects techniques que sur ceux liés à la gestion du projet

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.

Note relative au programme : Les ateliers, démonstrations, travaux pratiques et scénarios présentés dans ce programme sont fournis à titre d'exemples illustratifs et ne sont pas exhaustifs. Afin de garantir une adéquation permanente avec les évolutions technologiques, le contenu pédagogique ainsi que les environnements de travaux pratiques (TP) sont susceptibles d'être adaptés, ajustés ou mis à jour au cours du temps.

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
- 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 de 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 valide la base de connaissances nécessaires.

Les stagiaires disposent également, depuis leur extranet, d'un **outil d'auto-positionnement** leur permettant d'évaluer leur niveau de maîtrise des compétences visées par la formation. Cette auto-évaluation est réalisée avant le démarrage de la formation à l'aide d'un curseur de positionnement allant du niveau le plus faible au niveau le plus élevé.

Pendant

Après chaque module théorique, un ou plusieurs ateliers pratiques permettent de valider progressivement l'acquisition des connaissances. Un quiz peut accompagner les ateliers pratiques afin de mesurer la compréhension des notions abordées.

Après

À l'issue de la formation, les stagiaires réalisent une nouvelle **auto-évaluation** sur les mêmes compétences que celles évaluées avant la formation. Cette comparaison permet de mesurer leur perception de la progression réalisée et de visualiser les compétences développées au cours de la formation.

Lorsque le programme de formation le prévoit, un examen de certification est organisé conformément aux modalités définies par l'éditeur ou le centre de certification (TOSA, Pearson VUE, ENI, PeopleCert…).

Le formateur réalise également une **évaluation individuelle des compétences** de chaque stagiaire sur l'ensemble des compétences visées par la formation. Pour chacune d'elles, il indique si la compétence est :

- Acquise ;
- En cours d'acquisition ;
- Non acquise.

Lorsque la compétence est évaluée comme **en cours d'acquisition** ou **non acquise**, le formateur formule des observations précisant les difficultés rencontrées ainsi que les actions ou recommandations permettant au stagiaire de poursuivre sa montée en compétences.

Enfin

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

Informations sur l'accessibilité :

Nos formations sont, dans la mesure du possible, conçues pour être accessibles à toutes et à tous. Afin de garantir les meilleures conditions d'accueil et d'apprentissage pour les personnes en situation de handicap, nous vous invitons à contacter notre référente handicap certifiée :
Céline SOLATGES – 05 61 34 39 80 – csolatges@iform.fr
Nous vous remercions de bien vouloir nous communiquer toute information utile à ce sujet en amont de la formation, afin de mettre en place les adaptations nécessaires et d'assurer un accompagnement optimal.

Pour en savoir plus sur les dispositifs d'accompagnement existants, vous pouvez consulter les sites suivants :

- AGEFIPH
- FIPHFP
- MDPH de votre département

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 gaphes
- 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 oeuvre 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 oeuvre 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
• Requêtage
• L'interface Kibana