



Big Data - Administration et infrastructure

Big Data - Architecture et infrastructure Hadoop On-Premise et sur le Cloud

5 jours (35h00) | ★★★★★ 4,6/5 | BIG-INFR | Évaluation qualitative de fin de stage |
Formation délivrée en présentiel ou distanciel

Formations Informatique > Big Data > Big Data - Administration et infrastructure

Contenu mis à jour le 13/10/2023. Document téléchargé le 23/05/2024.

Objectifs de formation

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

- Exploiter les architectures Big Data Hadoop On-Premise et Cloud
- Mettre en place des socles techniques complets pour des projets Big Data.

Modalités, méthodes et moyens pédagogiques

Formation délivrée en présentiel ou distanciel* (blended-learning, e-learning, classe virtuelle, présentiel à distance).

Le formateur alterne entre méthode** démonstrative, interrogative et active (via des travaux pratiques et/ou des mises en situation).

Variables suivant les formations, les moyens pédagogiques mis en oeuvre sont :

- Ordinateurs Mac ou PC (sauf pour certains cours de l'offre Management), connexion internet fibre, tableau blanc ou paperboard, vidéoprojecteur ou écran tactile interactif (pour le distanciel)
- Environnements de formation installés sur les postes de travail ou en ligne
- Supports de cours et exercices

En cas de formation intra sur site externe à M2i, le client s'assure et s'engage également à avoir toutes les ressources matérielles pédagogiques nécessaires (équipements informatiques...) au bon déroulement de l'action de formation visée conformément aux prérequis indiqués dans le programme de formation communiqué.

* nous consulter pour la faisabilité en distanciel

** ratio variable selon le cours suivi

Prérequis

Avoir une connaissance pratique du Shell Linux, des bases du réseau, des bases de données relationnelles.

Public concerné

Administrateurs systèmes, développeurs, architectes et chefs de projets techniques, consultants en Business Intelligence.

Cette formation :

- Est animée par un consultant-formateur dont les compétences techniques, professionnelles et pédagogiques ont été validées par des diplômes et/ou testées et approuvées par l'éditeur et/ou par M2i Formation
- Bénéficie d'un suivi de son exécution par une feuille de présence élargée par demi-journée par les stagiaires et le formateur.

Programme

Jour 1

Présentation du Big Data et d'Hadoop

- Introduction au Big Data : de quoi s'agit-il ?
- Perspectives offertes par le Big Data
- Les acteurs du Big Data
- Exemples pratiques
- Démystification du Big Data
- Les technologies concernées
- Les outils Hadoop
- Les différents métiers du Big Data
- Les distributions Big Data
- Démonstration de traitements avec Hadoop

Jour 2

Solutions Big Data avec Hadoop

- Projets, applications, plateformes
- Solutions de stockage des données
- Considérations Hardware
 - Stockage
 - CPU
 - Mémoire
 - Réseau
 - Systèmes distribués
- La science et l'art de l'analyse prédictive
- Aspects économiques (OPEX, CAPEX, TRI)

Jour 3

Architecture cluster applicative et technique avec Hadoop

- Les différentes étapes de gestion des données
- Les outils du marché
- L'ingestion
- Le stockage
- La transformation
- La restitution et la visualisation

Jour 4

Installation et configuration d'un cluster Hadoop On-Premise et sur le Cloud Hadoop

- Installation d'une plateforme Hadoop
 - Définir l'architecture
 - Installation d'Hadoop / HDFS et d'autres composants
 - Installation des principaux composants (Hive, Pig, Impala...)
- Configuration de l'environnement
 - Etude des fichiers de configuration
 - Modes de configuration (ligne de commandes, API Rest, serveur HTTP intégré, API natives)
 - Configuration des services (HDFS, YARN, logs...)
 - Configuration des droits d'accès

Configuration d'un cluster hybride On-Premise / Cloud

- Configuration d'Hadoop
- Configuration de l'infrastructure Cloud

Jour 5

Administration d'un cluster Hadoop

- Gestion des ressources (noeuds, disques, CPU, mémoire, réseau...)
- Administration de HDFS
- Administration de MapReduce
- Infrastructure physique des centres de données, serveurs, racks, réseaux
- Optimisation au niveau du cluster (HDFS, MapReduce), noeuds / OS et réseau (binding, jumbo frames)
- Planification des stratégies de sécurité et de récupération des données et de reprise sur panne

Exemples de travaux pratiques (à titre indicatif)

- *Installation, configuration et administration pas à pas d'un cluster Hadoop On-Premise et de l'infrastructure Cloud associée*
- *Data Warehousing avec Sqoop, Hue, Hive, Impala*
- *Conception et optimisation d'architectures cluster matérielles (racks / noeuds / CPU / RAM / disques...)*
- *Comparatif technico-économique entre une solution On-Premise et une solution Cloud*

Examen M2i (en option)

- Prévoir l'achat de l'examen en supplément
- L'examen (en français) sera passé le dernier jour, à l'issue de la formation et s'effectuera en ligne
- Il s'agit d'un QCM dont la durée moyenne est d'1h30 et dont le score obtenu attestera d'un niveau de compétence
- L'examen n'est pas éligible au CPF, mais permettra néanmoins de valider vos acquis

Le contenu de ce programme peut faire l'objet d'adaptation selon les niveaux, prérequis et besoins des apprenants.

Modalités d'évaluation des acquis

- En cours de formation, par des études de cas ou des travaux pratiques
- Et, en fin de formation, par un questionnaire d'auto-évaluation et/ou un examen M2i

Les + de la formation

Un examen M2i permettant de valider vos acquis à l'issue de la formation est disponible sur demande (coût : 120€).

Accessibilité de la formation

Le groupe M2i s'engage pour faciliter l'accessibilité de ses formations. Les détails de l'accueil des personnes en situation de handicap sont consultables sur la page Accueil et Handicap.

Modalités et délais d'accès à la formation

Les formations M2i sont disponibles selon les modalités proposées sur la page programme. Les inscriptions sont possibles jusqu'à 48 heures ouvrées avant le début de la formation. Dans le cas d'une formation financée par le CPF, ce délai est porté à 11 jours ouvrés.