

retengr

Architecture Big Data FOAD

Durée : 6.5 jours



Méthode pédagogique

Composée à 70% de pratique, cette formation utilise des exercices illustrés, présentés et accomplis avec le formateur, et des exercices à réaliser en autonomie.

Une journée se décompose de la façon suivante :

Matin : 2h de théorie en visioconférence, 10 à 15 minutes de présentation d'exercices en visio, 1h de TP en autonomie avec possibilité de solliciter le formateur (partage d'écran à distance pour une assistance efficace).

Après-midi : 2h de théorie en visioconférence, 10 à 15 minutes de présentation d'exercices en visio, 1h de TP en autonomie avec possibilité de solliciter le formateur (partage d'écran à distance pour une assistance efficace).

Une évaluation quotidienne de l'acquisition des connaissances de la veille est effectuée.

Une synthèse est proposée en fin de formation.

Une évaluation à chaud sera proposée au stagiaire à la fin du cours.

Un support de cours (version électronique) sera remis à chaque participant comprenant les slides sur la théorie, les exercices. Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de présence de connexion pour chaque participant.

Enfin, une attestation de formation sera envoyée si le stagiaire a bien assisté à la totalité de la session.

En ce qui concerne le matériel informatique du stagiaire, il est seulement préconisé un ordinateur et une connexion internet. Nous nous chargeons du reste.

Chaque participant se verra attribuer une Machine Virtuelle qui sera exécutée dans le Cloud d'Amazon. Il disposera alors de la puissance et des outils nécessaires pour le bon déroulement de la formation. Aucune installation de la part du participant n'est requise avant la formation.



Présentation

Si le Big Data est LE sujet du moment, il est souvent considéré comme une boîte noire dans laquelle il est très difficile de se retrouver. Large regroupement de pratiques, d'objectifs et de technologies différentes, il demeure l'objet de nombreux questionnements :

A partir de quand parlons-nous de Big Data ?

Quels outils pour gérer de gros volume de données ?

Traitement batch ou traitement en continue ?

La data science implique t'elle nécessairement une approche Big Data ?

Quelle compétence pour un data scientist ?

Objectifs

- Définir les concepts et identifier l'apport du Big Data
- Déterminer l'écosystème technologique
- Organiser la collecte des données
- Choisir une technologie de stockage de données
- Connaître les technologies pour traiter les gros volumes de données
- Définir et comprendre le rôle du datascientist

Audience

Architectes, Chefs de projet, DSI, Data Scientists, Développeurs



Le formateur

Le formateur est un expert du domaine qui intervient sur le sujet depuis plusieurs années en formation mais aussi en conseil.

Doté d'une grande qualité d'écoute, sa pédagogie et sa compétence technique vous permettront d'acquérir les compétences sur les architectures Big Data.

Pré-requis

Les compétences professionnelles suivantes sont souhaitables : la connaissance d'un langage de programmation structuré et les bases du monde relationnel.

Programme

JOUR 1 [3.5h]

Comprendre les concepts et les enjeux du

BigData [3.5h]

- Origines et définition du BigData.
- Les 3 V : Volume, Vitesse et Variété
- Diversité dans les cas d'usage : données chaudes, données froides
- Big Data : Une approche réservée aux GAFAs ?
- Un exemple d'architecture Big Data.
- Exercice / Démo : Parcourir différentes sources de données accessibles via le WEB (API)



Expliquer les technologies du Big Data [3.5h]

- Définir les outils de collecte de données
- Anticiper les moyens de stockage en fonction des usages
- Le datalake : votre référentiel de données
- Paralléliser ou traiter vos données en continue ?
- S'approprier les données avec des analyses visuelles : la dataviz

Stocker des données [3.5h]

- État de l'art : Le BigData, sonne t'il le glas des bases de données relationnelles ?
- Le triangle de CAP
- Pourquoi le NoSql ?
- Les différentes approches : document / wide column / key-value
- Tour d'horizon des solutions à disposition : MongoDB, Cassandra, HBase...
- Exercice / Démo : définir et mettre en place un modèle de stockage de type document avec MongoDB

Collecter les données [3.5h]

- Comprendre les différentes sources de données : IoT / SI / Réseau sociaux / API : D'où viennent les données ?
- Gérer des formats de données différents : JSON, XML, CSV, binaires, ...
- De l'importance des connecteurs...
- Tour d'horizon des outils du marché : NIFI / Node Red / Flume / Sqoop
- Exercice / Démo : Utiliser NIFI pour collecter les données d'une API publique



JOUR 2 [3.5h]

Hadoop [3.5h]

- Comprendre le périmètre de Hadoop : Stockage et traitement
- Une plateforme de traitement batch et de stockage de données froides
- Architecture et composants de la plateforme Hadoop.
- HDFS, YARN et Mapreduce : les 3 piliers
- Un écosystème complexe et complet : Hive, HBase
- Exercice / Démo : Manipuler des fichiers via Hue, mise en place de tables et requêtes Hive sur une plateforme Hadoop

Spark [3.5h]

- Un framework pour paralléliser des traitements
- Positionnement Spark / Hadoop
- Quelle infrastructure de déploiement
- Comprendre la complexité de la parallélisation des traitements
- SparkML : une librairie pour la datascience
- Exercice / Démo : Mise en place et analyse d'un traitement simple

Stream processing [3.5h]

- Le besoin de traitement au fil de l'eau
- Streaming ETL
- Streaming analytics
- Prise de décision en temps réel
- Les approches et outils de streaming : Spark Streaming / Kafka Streaming / Flink...
- Exercice / Démo : analyse en continue d'un flux de données simple



JOUR 3 [3.5h]

Transporter vos données : Kafka [3.5h]

- Définir le besoin d'un bus de données
- Les middleware Orienté Messages dans un contexte BigData
- Définir les acteurs : Producers & Consumers
- Comprendre les composants : Messages, brokers, topics, ...
- Un outil taillé pour les performances
- Kafka Connect : Connectez vos outils à Kafka
- Exercice / Démo : Mise en place d'un bus Kafka pour permettre à Elasticsearch de manipuler des données extraites via NIFI

Big Data et Machine Learning [3.5h]

- Présentation du Machine Learning
- Positionnement de la datascience dans un contexte Big Data
- Les différentes approches : Clusterisation, classification, régression
- Les implémentations : Scikit Learning / SparkML
- Spark et DASK : des frameworks de distribution des traitements
- Le « Deep learning »
- Le « Online learning » ou machine learning en streaming
- Démo : processus complet d'un projet de datascience (analyse des données, sélection de données, apprentissage, scoring)

Datavisualisation [3.5h]

- Pourquoi faire ?
- Dataviz pour comprendre les données
- L'écosystème de la Dataviz : outils et API
- Exercice / Démo : Analyse visuelle d'un jeu de données



Modalités et délais d'accès à la formation

Les inscriptions sont possibles jusqu'à 48 heures ouvrées avant le début de la formation, en interentreprises, dans la limite des places disponibles. Pour les formations organisées en intra entreprise, la liste des participants peut être modifiée jusqu'à 24h ouvrées avant le début de la formation.

Accessibilité

RETENGR facilite l'accessibilité de ses formations.

Cette formation est accessible aux personnes en situation de handicap.

Si vous avez un besoin d'accès spécifique, contactez Céline BOURREIL (celine.bourreil@retengr.com) qui étudiera avec Handifiel's (notre référent handicap) votre demande et vous proposera les meilleures solutions

The background is a vibrant, abstract composition. It features large, overlapping organic shapes in shades of purple, yellow, and red. On the right side, there is a pattern of small yellow dots arranged in a grid that tapers off towards the top. In the lower half, there are more shapes, including a light pink circle and a purple shape with a teal section. A dashed blue line with a circular loop is also visible in the lower left.

**Vous allez nous adorer si
comme nous vous pensez que...**

Une formation doit être au service de la performance du collaborateur et de l'entreprise

Ceci nécessite une quête constante d'excellence de la part de l'organisme formateur avec une adaptation systématique aux enjeux de l'entreprise, la mise à jour régulière des supports de cours et une veille technologique indispensables pour toujours être à la pointe du domaine.



L'expertise technique est aussi importante que les qualités pédagogiques



Nos formateurs sont tous des experts de leur domaine. Mais qu'ont-ils de plus que les autres ? Nous les sélectionnons en plus pour leurs qualités de pédagogue et leurs méthodes d'enseignements. Nous plaçons les qualités pédagogiques au même niveau que l'expertise afin que nos stagiaires tirent le meilleur de leurs formations.



re'engr

L'excellence naît de l'excellence

Beaucoup de nos clients se classent parmi les leaders de leurs industries respectives ou parmi les start-ups les plus prometteuses. Nous savons que former les collaborateurs de telles entreprises nécessite de prêter attention à chaque détail en prodiguant un accompagnement à la hauteur de l'ambition de nos stagiaires. C'est pourquoi nous savons faire des leaders d'aujourd'hui les champions de demain !





retengr

**Faire du leader
d'aujourd'hui, le champion
de demain**