

The logo for 'retengr' is displayed in a white, lowercase, sans-serif font. The letters are clean and modern, with a slight shadow effect. The background is a vibrant red with abstract, overlapping organic shapes in various shades of red and pink. A decorative pattern of white dots is visible on the right side of the page.

retengr

Big Data Une journée pour tout comprendre

Durée : 5.5 jours



Méthode pédagogique

La formation se compose de travaux pratiques (mises en situation, débats, exercices).

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 fin de formation si le stagiaire a bien assisté à la totalité de la session.

Présentation

Limité au monde des GAFAs il y a quelques années encore, la gestion des données volumineuses est une préoccupation pour toutes les entreprises, quel que soit leur domaine d'activité. Le Big Data est un vaste sujet : stratégique, organisationnel, et technique, il requiert des adaptations importantes à la fois dans les technologies mais aussi dans les esprits. Ces arbitrages demeurent complexes, surtout face à un sujet dont l'étendue et la complexité en rendent une lecture difficile et souvent pleine de fausses idées.

Cette conférence se décompose en deux parties distinctes : la matinée est consacrée à une présentation générale du Big Data, et l'après-midi à un tour d'horizon des technologies du Big Data.



Objectifs

- Première partie : le matin
- Connaître les origines du BigData
- Découvrir les concepts qui se cachent derrière le terme BigData
- Déterminer les usages pertinents au travers d'exemples
- Définir les étapes qui jalonnent la mise en œuvre d'une approche BigData
- Mesurer l'impact du BigData sur notre quotidien
- Évaluer la relation entre BigData et Intelligence Artificielle
- Deuxième partie : l'après-midi
- Dégrossir le panorama technologique des outils autour du BigData
- Utiliser nifi et/ou node-red pour automatiser vos workflows de collecte
- Pourquoi utiliser Hadoop comme infrastructure de stockage et de traitement ?
- Comprendre l'usage des frameworks Spark et Dask
- Appréhender les différentes approches en Machine Learning et Deep Learning

Audience

DSI, Architectes, 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 une vue d'ensemble sur le Big Data.

Pré-requis

La connaissance d'un langage de programmation et du monde des bases de données est nécessaire.

Programme

Première partie : le matin [3.5h]

- Cette présentation utilise un fil rouge ludique et dont la compréhension est à la portée de l'ensemble de l'audience (aucun prérequis technique). Les différentes étapes d'un projet Big Data ainsi que les outils y afférent seront expliqués en s'appuyant sur un cas concret.

Origine des données [3.5h]

- Définir les différentes sources de données disponibles
- Données structurées / non structurées
- Comprendre les différents moyens d'y accéder
- Confidentialité des données, RGPD

Collecte [3.5h]

- Réflexion sur les besoins : origine, fiabilité, fréquence, coût
- Les différents formats de données
- Nettoyage et consolidation
- Problématiques techniques



Stockage [3.5h]

- Les limites des modèles actuels
- Le principe du datalake
- Les nouveaux modèles NoSql et NewSql
- Problématiques techniques du stockage

Traitement [3.5h]

- Comment traiter de gros volumes de données
- Paralléliser pour aller plus vite

Deuxième partie : l'après-midi [3.5h]

- Cette 2ème partie s'adresse à un public plus technique, qui nécessite un minimum de connaissances des technologies informatiques pour pouvoir suivre aisément. C'est surtout le vocabulaire qui va être plus poussé.

Collecte de données [3.5h]

- Nifi
- Node-red

Les nouveaux moyens de stockages [3.5h]

- Les nouveaux types de bases de données : key / value, document, wide-column
- MongoDB : une base de données de type document
- Scalability & Sharding
- Hadoop : Un framework de stockage et de traitement en parallèle

Hadoop [3.5h]

- Le framework BigData
- Les piliers : HDFS, Yarn et Mapreduce



- HDFS : architecture, découpage en block, réplication
- YARN : le chef d'orchestre des tâches Hadoop
- Mapreduce : un template de programmation pour les traitements en parallèles
- L'écosystème Hadoop : Hive, impala, ...

Spark [3.5h]

- Un autre framework pour paralléliser les traitements
- Plusieurs langages : Python, Scala, Java
- RDD et DAG : les principes fondamentaux de Spark
- Dataframe et Dataset : structuration et typage des données
- Infrastructure d'exécution : Spark sur Yarn, mesos

Intelligence Artificielle [3.5h]

- Machine learning et Deep Learning
- Les librairies : Scikit learn et sparkml
- Régression | Classification | Clustering
- Analyse, nettoyage et préparation des données
- Apprentissage et scoring

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. A prominent yellow shape on the left transitions into a red shape on the right. In the upper right, there is a pattern of small yellow dots arranged in a grid-like structure. At the bottom, there are more organic shapes in light pink, teal, and purple. A dashed blue line with a circular loop is visible in the lower-left quadrant.

**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**