Les fondamentaux d'Hadoop

DESCRIPTION
Cette formation est une initiation aux fondamentaux d'Hadoop. Elle donne aux participants une connaissance théorique et pratique de la plateforme, au travers de plusieurs exercices pratiques appliqués à des cas réels. A l'issue de la session, les participants seront en capacité d'utiliser les outils de l'écosystème Hadoop pour explorer des données stockées sur un entrepôt Big Data.

OBJECTIFS PEDAGOGIQUES

- Appréhender le fonctionnement d'Hadoop
- Identifier l'écosystème : quels outils pour quels usages ?
- Manipuler les principales commandes shell d'interaction avec Hadoop
- Emettre des requêtes SQL avec Hive et HCatalog
- Créer des traitements de données avec Pig

PUBLIC CIBLE

- Analyste
- Data scientist
- Architecte
- Développeur

PRE-REQUIS
Connaissances de base en programmation ou en scripting.

METHODE PEDAGOGIQUE
Apports théoriques préalables présentant la plateforme Hadoop, son fonctionnement et son écosystème. Travaux pratiques pour une mise en application immédiate. Les travaux pratiques s'effectueront sur un cluster Apache Hadoop 2.x.

PROFIL DES INTERVENANTS
Toutes nos formations sont animées par des consultants-formateurs expérimentés et reconnus par leurs pairs.

Stage pratique
Data Engineering

Code :
HDP01

Durée :
2 jour(s) (14,00 heures)

Exposés : **60.00 %**
Cas pratiques : **30.00 %**
Echanges d'expérience : **10.00 %**

Inter-entreprises :
Prochaines sessions disponibles sur notre site web.
Tarif : 1 680,00 € HT / participant

Intra-entreprise :
Tarifs et dates sur demande.

OCTO Academy est une marque d'OCTO Technology, référencé au DATADOCK – 34, avenue de l'Opéra, 75002 Paris France - SAS au Capital de 509 525.30 €
Tél : +33 1 70 36 18 67 – Fax :+33 1 58 56 10 01 –academy@octo.com –https://www.octo.academy –academy@octo.com
SIRET : 418 166 096 00069 – RCS Paris B 418 166 096 - NAF 6202A – Organisme de formation N° 11 75 48 936 75

MODALITÉS D'ÉVALUATION ET FORMALISATION À L'ISSUE DE LA FORMATION

L'évaluation des acquis se fait tout au long de la session au travers des ateliers et des mises en pratique. Une évaluation à chaud sur la satisfaction des stagiaires est réalisée systématiquement en fin de session et une attestation de formation est délivrée aux participants mentionnant les objectifs de la formation, la nature, le programme et la durée de l'action de formation ainsi que la formalisation des acquis.

PROGRAMME PEDAGOGIQUE DETAILLE

Jour 1

INTRODUCTION AU BIG DATA

Qu'est-ce que Big Data ?

Les grands enjeux métier

Paysage technologique : les architectures Big Data

INTRODUCTION À HADOOP

Historique succinct

Le cœur de la plateforme : HDFS et YARN

L'écosystème Hadoop

- Frameworks et algorithmes
- Bases de données
- Traitements des données
- Intégration

Hadoop et la sécurité des données

MANIPULER LA LIGNE DE COMMANDE HADOOP

Présentation des principales commandes

Mise en pratique "Manipulation et transfert de fichiers en ligne de

OCTO Academy est une marque d'OCTO Technology, référencé au DATADOCK – 34, avenue de l'Opéra, 75002 Paris France - SAS au Capital de 509 525.30 €
Tél : +33 1 70 36 18 67 – Fax :+33 1 58 56 10 01 –academy@octo.com –https://www.octo.academy –academy@octo.com
SIRET : 418 166 096 00069 – RCS Paris B 418 166 096 - NAF 6202A – Organisme de formation N° 11 75 48 936 75

commande

UNE INTERFACE UTILISATEUR POUR HADOOP : HUE

Présentation de Hue et de ses modules

Mise en pratique "Manipulation interactive de données"

INTERROGER HADOOP AVEC DU SQL : HIVE

Présentation de Hive

Mise en pratique "Manipulation de données avec SQL"

- Créer un modèle de données
- Importer des fichiers sources
- Requêter les données

Jour 2

TRANSFORMER DES DONNÉES : LE LANGAGE PIG

Présentation de Pig

Mise en pratique "Transformation de données avec Pig"

- Charger des données semi-structurées
- Croiser avec des données Hive
- Sauvegarder le résultat dans HDFS

ECRITURE DE TRAITEMENTS AVANCÉS

Présentation du framework Hadoop Streaming

Présentation express du langage Python et du squelette de programme pour l'exercice

Mise en pratique "Ecriture d'un programme de manipulation complexe"

COMPOSITION ET ORDONNANCEMENT DE TRAITEMENTS

OCTO Academy est une marque d'OCTO Technology, référencé au DATADOCK – 34, avenue de l'Opéra, 75002 Paris France - SAS au Capital de 509 525.30 €
Tél : +33 1 70 36 18 67 – Fax :+33 1 58 56 10 01 –academy@octo.com –https://www.octo.academy –academy@octo.com
SIRET : 418 166 096 00069 – RCS Paris B 418 166 096 - NAF 6202A – Organisme de formation N° 11 75 48 936 75

Présentation d'Oozie

Mise en pratique "Création d'un pipeline de traitement de données"

OCTO Academy est une marque d'OCTO Technology, référencé au DATADOCK – 34, avenue de l'Opéra, 75002 Paris France - SAS au Capital de 509 525.30 €
Tél : +33 1 70 36 18 67 – Fax :+33 1 58 56 10 01 –academy@octo.com –https://www.octo.academy –academy@octo.com
SIRET : 418 166 096 00069 – RCS Paris B 418 166 096 - NAF 6202A – Organisme de formation N° 11 75 48 936 75