Architecture des données : stockage et accès

DESCRIPTION
Si les algorithmes de Machine Learning ont connu des avancées majeures ces dernières années, c'est avant tout grâce à la quantité d'information disponible pour les entraîner. Accumuler toute cette donnée, la traiter, et la rendre disponible sont les enjeux principaux du mouvement Big Data.

Au cours de cette formation, nos consultants mettent à disposition les connaissances issues de leurs retours d'expériences auprès de nos clients, et vous font découvrir les bases des architectures permettant de répondre à ces enjeux de stockage et d'accès.

OBJECTIFS PEDAGOGIQUES

- Découvrir les notions centrales de stockage de données
- Appréhender les enjeux des nouvelles architectures de données (Hadoop, NoSQL, Spark), et positionner leurs usages au sein de l'univers Big Data
- Savoir manipuler ces technologies et les bases de données de façon conjointe, pour mener à bien des analyses efficaces

PUBLIC CIBLE
Analyste

Statisticien

Développeur

PRE-REQUIS
Notions de programmation sur la base d'un langage quelconque

Manipulation basique de la ligne de commande Linux

METHODE PEDAGOGIQUE
Formation avec apports théoriques, échanges sur les contextes des participants et retours d'expérience pratique du formateur, complétés de travaux pratiques et de mises en situation.

Stage pratique
Data Science

Code :
DSARC

Durée :
3 jour(s) (21,00 heures)

Exposés : **50 %**
Cas pratiques : **40 %**
Echanges d'expérience : **10 %**

Inter-entreprises :
Prochaines sessions disponibles sur notre site web.
Tarif : 2 390,00 € 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

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

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

1.Introduction

2.Accès aux données

2.1.Les fichiers

2.1.1.Arborescence

2.1.2.Formats

2.2.Les bases de données

2.2.1.Notion de Catalogue

2.2.2.Le langage SQL

2.2.3.Data Warehouses et Data Lake

2.2.4.Bases d'analyse

2.3.API

2.3.1.Définition

2.3.2.Web Scraping

2.4.Traitements en mémoire

3.Architecte de données

3.1. Limites des systèmes traditionnels

3.1.1.Limites des fichiers

3.1.2.Limites des SGBD

3.2.Les architectures distribuées

3.2.1.Patterns d'accès

3.2.1.1.OLTP

3.2.1.2.OLAP

3.2.2.Distribution vs Réplication

3.3.Concepts essentiels

3.3.1.Disponibilités

3.3.2.Cohérence

3.3.3.Tolérance à la partition

3.4.Le théorème CAP

3.5.Quorums

4.Bases NoSQL

4.1.Avantages et inconvénients

4.2.Modèles de données

4.2.1.Key-Value

4.2.2.Documents

4.2.3.Column-Family

4.2.4.Graph

4.3.Exemple : MongoDB

4.4.Les moteurs de recherche

JOUR 2

5.Hadoop

5.1.Introduction à Hadoop

5.1.1.Histoire

5.1.2.Ecosystème

5.2.HDFS

5.3.Map-Reduce

5.3.1.Les phases de Map-Reduce

5.3.2.Notion de job

5.3.3.Exemple

5.4.YARN

5.5.Les distributions

5.6.La ligne de commande

5.7.Administration d'un cluster

5.7.1.Resource Manager

5.7.2.History Server

5.7.3.Hue

6.Études de cas

6.1.Traitements de courbes de charge

6.1.1.Contexte et hypothèses

6.1.2.Raisonnements

6.2.Analyse de logs

6.2.1.Contexte et Hypothèses

6.2.2.Raisonnements

7.Conclusion

7.1.Rappels des points abordés

7.2.Questions et réponses

7.3.Synthèse

Jour 3

8.Découverte de Spark

8.1. Spark Core

8.1.1. RDD

8.1.2.Transformations

8.1.3. Pair RDD

8.2.Spark SQL

8.3.Spark Streaming

8.4.Structured Streaming

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