

BIG DATA - ENJEUX ET OPPORTUNITÉ

| | | | |
|--------------|----------------|----------------------------|------------------|
| Durée | 2 jours | Référence Formation | 4-PP-BDBA |
|--------------|----------------|----------------------------|------------------|

Objectifs

- Découvrir les principaux concepts du Big Data
- Identifier les enjeux économiques
- Evaluer les avantages et les inconvénients du Big Data
- Comprendre les principaux problèmes et les solutions potentielles
- Identifier les principales méthodes et champs d'application du Big Data

Participants

DSI, directeurs techniques, chefs de projet, architectes, responsables SI

Pré-requis

Connaissances de base des architectures techniques

Moyens pédagogiques

- Réflexion de groupe et apports théoriques du formateur
- Travail d'échange avec les participants sous forme de réunion-discussion
- Utilisation de cas concrets issus de l'expérience professionnelle
- Validation des acquis par des questionnaires, des tests d'évaluation, des mises en situation et des jeux pédagogiques.
- Remise d'un support de cours.

PROGRAMME

Introduction

- Les origines du Big Data : un monde de données numériques, l'e-santé, chronologie
- Une définition par les quatre V : la provenance des données
- Une rupture : changements de quantité, de qualité, d'habitudes
- La valeur de la donnée : un changement d'importance
- La donnée en tant que matière première
- Le quatrième paradigme de la découverte scientifique

Big Data : traitements depuis l'acquisition jusqu'au résultat

- L'enchaînement des opérations
- L'acquisition
- Le recueil des données : crawling, scraping
- La gestion de flux événementiels (Complex Event Processing, CEP)
- L'indexation du flux entrant
- L'intégration avec les anciennes données
- La qualité des données : un cinquième V ?
- Les différents types de traitement : recherche, apprentissage (machine learning, transactionnel, data mining)
- D'autres modèles d'enchaînement : Amazon, e-Santé
- Un ou plusieurs gisements de données ? De Hadoop à l'in-memory
- De l'analyse de tonalité à la découverte de connaissances

Relations entre Cloud et Big Data

- Le modèle d'architecture des clouds publics et privés

CAP ÉLAN FORMATION

www.capelanformation.fr - Tél : 04.86.01.20.50

Mail : contact@capelanformation.fr

Organisme enregistré sous le N° 76 34 0908834

version 2024

- Les services XaaS
- Les objectifs et avantages des architectures Cloud
- Les infrastructures
- Les égalités et les différences entre cloud et Big Data
- Les clouds de stockage
- Classification, sécurité et confidentialité des données
- La structure comme critère de classification : non structurée, structurée, semi-structurée
- Classification selon le cycle de vie : données temporaires ou permanentes, archives actives
- Difficultés en matière de sécurité : augmentation des volumétries, la distribution
- Les solutions potentielles

Introduction à l'Open Data

- La philosophie des données ouvertes et les objectifs
- La libération des données publiques
- Les difficultés de la mise en œuvre
- Les caractéristiques essentielles des données ouvertes
- Les domaines d'application
- Les bénéfices escomptés

Matériel pour les architectures de stockage

- Les serveurs, disques, réseau et l'usage des disques SSD, l'importance de l'infrastructure réseau
- Les architectures cloud et les architectures plus traditionnelles
- Les avantages et les difficultés
- Le TCO
- La consommation électrique : serveurs (IPNM), disques (MAID)
- Le stockage objet : principe et avantages
- Le stockage objet par rapport aux stockages traditionnels NAS et SAN
- L'architecture logicielle
- Niveaux d'implantation de la gestion du stockage
- Le "Software Defined Storage"
- Architecture centralisée (Hadoop File System)
- L'architecture Peer-to-Peer et l'architecture mixte
- Les interfaces et connecteurs : S3, CDMI, FUSE, etc.
- Avenir des autres stockages (NAS, SAN) par rapport au stockage objet

Protection des données

- La conservation dans le temps face aux accroissements de volumétrie
- La sauvegarde, en ligne ou locale ?
- L'archive traditionnelle et l'archive active
- Les liens avec la gestion de hiérarchie de stockage : avenir des bandes magnétiques
- La réplication multisites
- La dégradation des supports de stockage

Méthodes de traitement et champs d'application

- Classification des méthodes d'analyse selon le volume des données et la puissance des traitements
- Hadoop : le modèle de traitement Map Reduce
- L'écosystème Hadoop : Hive, Pig. Les difficultés d'Hadoop
- OpenStack et le gestionnaire de données Ceph
- Le Complex Event Processing : un exemple ? Storm
- Du BI au Big Data
- Le décisionnel et le transactionnel renouvelés : les bases de données NoSQL
- Typologie et exemples
- L'ingestion de données et l'indexation
- Deux exemples : splunk et Logstash
- Les crawlers Open Source

- Recherche et analyse : Elasticsearch
- L'apprentissage : Mahout. In-memory
- Visualisation : temps réel ou non, sur le Cloud (Bime), comparaison QlikView, Tibco Spotfire, Tableau
- Une architecture générale du data mining via le Big Data

Cas d'usage à travers des exemples et conclusion

- L'anticipation : besoins des utilisateurs dans les entreprises, maintenance des équipements
- La sécurité : des personnes, détection de fraude (postale, taxes), le réseau
- La recommandation. Analyses marketing et analyses d'impact
- Analyses de parcours. Distribution de contenu vidéo
- Big Data pour l'industrie automobile ? Pour l'industrie pétrolière ?
- Faut-il se lancer dans un projet Big Data ?
- Quel avenir pour les données ?
- Gouvernance du stockage des données : rôle et recommandations, le Data Scientist, les compétences d'un projet Big Data