Big Data - Enjeux et opportunité

Mise à jour janv. 2023

Durée : 2 jours - 14 heures

OBJECTIFS PÉDAGOGIQUES

- 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

PRÉREQUIS

- Connaissances de base des architectures techniques

PARTICIPANTS

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

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.
- Alternance entre apports théoriques et exercices pratiques (en moyenne 30 à 50%)
- Remise d'un support de cours.

MODALITÉS D'ÉVALUATION

- Feuille de présence signée en demi-journée
- Évaluation des acquis tout au long de la formation
- Questionnaire de satisfaction
- Attestation de stage à chaque apprenant
- Positionnement préalable oral ou écrit
- Évaluation formative tout au long de la formation
- Évaluation sommative faite par le formateur ou à l'aide des certifications disponibles

MOYENS TECHNIQUES EN PRÉSENTIEL

- Accueil des stagiaires dans une salle dédiée à la formation, équipée d'ordinateurs, d'un vidéo projecteur d'un tableau blanc et de paperboard

MOYENS TECHNIQUES DES CLASSES À DISTANCE

- A l'aide d'un logiciel comme Teams, Zoom... un micro et éventuellement une caméra pour l'apprenant,
- suivez une formation en temps réel et entièrement à distance. Lors de la classe en ligne, les apprenants interagissent et communiquent entre eux et avec le formateur
- Les formations en distanciel sont organisées en Inter-Entreprises comme en Intra-Entreprise
- L'accès à l'environnement d'apprentissage (support de cours, labs) ainsi qu'aux preuves de suivi et d'assiduité (émargement, évaluation) est assuré
- Les participants recevront une convocation avec lien de connexion
- Pour toute question avant et pendant le parcours, une assistance technique et pédagogique est à disposition auprès de notre équipe par téléphone au 03 25 80 08 64 ou par mail à secretariat@feep-entreprises.fr

ORGANISATION

- Les cours ont lieu de 9h à 12h30 et de 13h30 à 17h

PROFIL FORMATEUR

- Nos formateurs sont des experts dans leurs domaines d'intervention
- Leur expérience de terrain et leurs qualités pédagogiques constituent un gage de qualité.

ACCESSIBILITÉ

- Notre organisme peut vous offrir des possibilités d'adaptation et/ou de compensations spécifiques si elles sont nécessaires à l'amélioration de vos apprentissages sur l'ensemble de nos formations. Aussi, si vous rencontrez une quelconque difficulté, nous vous invitons à nous contacter directement afin d'étudier ensemble les possibilités de suivre la formation

Programme de formation

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.
- 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.