Formation Big Data - L'essentiel De la technologie aux usages du Big Data

14 heures
SEM34

Objectifs pédagogiques
Comprendre le concept du Big Data
Être capable d'identifier l'écosystème et comprendre les technologies associées
Savoir anticiper son intégration dans les activités informatiques de l'entreprise
Être en mesure de l'exploiter dans le respect des règles de sécurité et de confidentialité

Public(s)
Responsables de la DSI s'interrogeant sur les apports et le déploiement du Big Data
Chefs de projets, Responsables de métiers et consultants souhaitant aborder les projets
Toute personne impliquée dans la réflexion et l'étude du Big Data

Pré-requis
Connaissances sommaires en informatique

Modalités pédagogiques
Alternance théorie et pratique

Moyens et supports pédagogiques
Support(s) de formation par apprenant

Modalités d'évaluation et de suivi

Cette formation ne fait pas l'objet d'un contrôle des acquis via une certification

Formateur

Programme

EXEMPLES D'USAGE PERTINENT DU BIG DATA
Réseaux sociaux : Google, Twitter, Youtube
Gestion des clients (CRM) : Vue 360° des clients / Multicanal
Sécurité informatiques (étude de logs) : identification des tentatives d'attaques
Analyse des logs d'Internet (Web)
Profiling d'individus : ADN numérique
Synthèse des critères de succès d'un projet Big Data et de nombreuses causes d'échec

DÉFINITION COMMUNE DU BIG DATA SELON LES GRANDS ACTEURS DU MARCHÉ
Caractéristiques techniques des 3V de Gartner (Vélocité, Variété et Volume) et les variantes (Véracité, Valeur, Validité...)
Collecte et traitement des données structurées, semi-structurées et non-déstructurées
Transformation des données en informations
Création de la valeur à partir des données / Exemple de monétisation
Exemple de processus : gestion des données en cycles, de l'acquisition à la gouvernance

INTRODUCTION AUX ARCHITECTURES DES SOLUTIONS DE CALCUL DISTRIBUÉ DU BIG DATA
Principe
Scalabilité horizontale et verticale
Architecture de cluster et composants économiques
Traitement parallèle des données
Enjeux des architectures distribuées selon l'organisme CSA (Cloud Security Alliance) dédié au Big Data : sécurité, gestion des données en grandes quantités

TECHNOLOGIES DE RÉFÉRENCE DU BIG DATA À CONNAÎTRE
Stockage des données à traiter : Data Lake
Différents types de base des données NoSQL
Hadoop : un modèle de traitement distribué du Big Data adopté par les grands acteurs de l'informatique : HDFS, YARN, MapReduce...
Ecosystème de Hadoop : Pig, Flume, Zookeeper, HBase, Oozie ...
Analyse de données : 4 types d'analytique
Machine Learning / Intelligence artificielle pour l'analytique

APPROCHES DE DÉPLOIEMENT DU BIG DATA
Déploiement sur site : définition des objectifs, choix des solutions d'analyse et d'intégration, présentation des informations / revue des fournisseurs de composants Big Data
Déploiement sur site en version distribuée : Hortonworks, MapR, Cloudera (choix des intégrateurs)

Déploiement dans les plates-formes Cloud Big Data
Précautions à prendre (métriques de qualité, système fermé ...)
Difficultés techniques à anticiper
Validation de la pertinence d'une plate-forme

PLATES-FORMES CLOUD PUBLIC DU BIG DATA À EXPLOITER
IBM Analytics de la plate-forme Bluemix
Amazon Web Services (stockage des données et plates-formes d'analytiques)
Google Platform Big Data
Microsoft Azure Big Data
Points communs et différents entre les plates-formes Big Data

QUALITÉ DES DONNÉES
Bonne pratique internationale de gouvernance des données
Qualification des données (temporel, contextuel, liens aux autres données...) / cadres juridiques, formats ouverts et propriétaires
Enrichissement avec l'Open Data / WiKiData

SÉCURITÉ DES DONNÉES ET CONFIDENTIALITÉ DU BIG DATA
Règlementation RGPD à respecter
Recommandation des bonnes pratiques de l'organisme international CSA (Cloud Security Alliance) pour le Big Data
Panorama des moyens techniques de sécurité des données et d'accès au Datacenter (cryptage et DLP : Data Lost Prevention....)
Recommandations de sécurité dans la manipulation des données

IMPACTS DU BIG DATA À ANTICIPER
Évolution des données (Internet des objets, mobilité...)
Remise en question des bonnes pratiques actuelles
Impacts sur les compétences des équipes informatiques en place
Nouveaux métiers (Data Scientist, Data Steward...) et compétences nouvelles à acquérir
Rôle de la DSI face à la montée du Big Data et des activités numériques de l'entreprise