

BIG DATA – L'ESSENTIEL

CODE STAGE : BD003

OBJECTIFS

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

DURÉE

2 jours

PUBLIC

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

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 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 : gestion des données en cycles, de l'acquisition à la gouvernance

TECHNOLOGIES DE RÉFÉRENCE DU BIG DATA À CONNAÎTRE

Stockage des données à traiter : fichiers, blocs et objets

Différents types de base des données NoSQL (Not Only SQL)

Architecture de cluster et composants économiques

Traitement parallèle des données (Grid)

Hadoop : un modèle d'open source du Big Data adopté par les grands acteurs de l'informatique (IBM, Oracle, Microsoft, Amazone, EMC, Google...)

Principaux composants d'Hadoop : HDFS (Hadoop Distributed File System), MapReduce ...

Ecosystème et technologies associées à Hadoop : Pig, Flume, Zookeeper, H-BASE, Lucine, Hive, Oozie, Cassandra, Machine Learning...

Exemple de traitement en temps réel : traitement des données à la volée (Data Streaming)

Analyse de données (Data Analytics et Business Intelligent)

INTRODUCTION AUX ARCHITECTURES DES SOLUTIONS DE CALCUL DISTRIBUÉ

Stockage objets (pas de verrouillage de fichier dans la cadre des multiutilisateurs)

Serveurs NoSQL et HDFS (Fichiers distribués)

Scalabilité horizontale

Enjeux des architectures distribuées selon l'organisme CSA (Cloud Security Alliance) dédié au Big Data : Sécurité, gestion des données en grandes quantités

Limitations en termes d'usages (Analytiques)

Impacts des choix de technologies et d'architectures sur les usages (traitement des données en batch, temps réel, streaming)

PLATES-FORMES CLOUD PUBLIC BIG DATA APAAS (DATA AS A SERVICE) À EXPLOITER

IBM Analytics de la plate-forme Bluemix

Amazone 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

TROIS APPROCHES DE DÉPLOIEMENT DU BIG DATA : SUR SITE ET DANS LE CLOUD DAAS

Causes des nombreux échecs de projets Big Data selon des cabinets d'étude du secteur

Trois approches de déploiement « sur site » : Hadoop et son écosystème à télécharger, Big Data en versions

distribuées et Data as a Service

Déploiement sur site : définition des objectifs, choix des solutions d'analyse et d'intégration, présentation des informations (Data Visualization) / revue des fournisseurs de composants Big Data

Déploiement sur site en version distribuée : Hortonworks, MapR, Cloudera, IBM

Déploiement dans les plates-formes Cloud Big Data et les précautions à prendre (métriques de qualité)

QUALITÉ DES DONNÉES

Les 11 principales étapes de traitement des données selon les organismes internationaux

Processus de qualification des données (temporel, contextuel, liens aux autres données...) / cadres juridiques (CNIL, usages libres, payants...), formats ouverts et propriétaires

Approche d'enrichissement avec l'Open Data / WikiData.org

SÉCURITÉ DES DONNÉES ET CONFIDENTIALITÉ DU BIG DATA

Loi européenne et CNIL (protection de la vie privée)

Recommandation des bonnes pratiques de l'organisme international CSA (Cloud Security Alliance) pour le Big Data

Panorama des moyens conventionnels de sécurité des données et d'accès au Datacenter (cryptage et DLP : Data Lost Prevention....)

IMPACTS DU BIG DATA À ANTICIPER

Évolution des données (Internet des objets, mobilité...)

Impacts sur les compétences des équipes informatiques, de DRH, du Management...

Rôle de la DSI face à la montée du Big Data et des solutions numériques

