

CONCEPTS BIG DATA

Big Data - Les techniques d'analyse et de visualisation

REF : SIBI008

DUREE : 21h

Mixte Classe virtuelle

PUBLIC

Cette formation Les techniques d'analyse et de visualisation est destinée aux Data Analysts, Business Analysts, Analystes Business Intelligence, Dataminers.

Modalités et délais d'accès : les inscriptions sont fermées 24h avant la 1ère journée de formation.

Accessibilité : Si vous avez des contraintes particulières liées à une situation de handicap, veuillez nous contacter au préalable afin que nous puissions, dans la mesure du possible, adapter l'action de formation.

PREREQUIS

Cette formation Les techniques d'analyse et de visualisation nécessite de connaître les principes de programmation et avoir une expérience dans le développement. Connaissance du langage SQL.

MODALITES PEDAGOGIQUES

1 poste et 1 support par stagiaire

8 à 10 stagiaires par salle ou en classe virtuelle

Remise d'une documentation pédagogique papier ou numérique pendant le stage

La formation est constituée d'apports théoriques, d'exercices pratiques, de réflexions et de retours d'expérience

MODALITES D'EVALUATION

Auto-évaluation des acquis par le stagiaire via un questionnaire en ligne

Attestation de fin de stage remise au stagiaire

OBJECTIFS PEDAGOGIQUES

Cette formation Les techniques d'analyse et de visualisation vous permettra de :

- Définir et identifier le contexte spécifique des projets Big Data
- Connaître le panorama technologique et enjeux socio-économiques
- Mesurer l'impact des choix technologiques en matière d'analyse et de visualisation
- Gérer la structure des données (structurées – non structurées ; internes – acquises)
- Intégrer des données par les techniques de collecte
- Mettre en œuvre un entrepôt et le stockage de données
- Maîtriser les méthodes d'analyse et de visualisation
- Consolider ses connaissances à travers un cas d'usage

PROGRAMME

Panorama technologique et enjeux socio-économiques

- Bâtir une vision Data Centric pour l'entreprise
- Etudier l'environnement concurrentiel de l'entreprise
- Comment créer de la valeur ou apporter de la valeur complémentaire aux données
- Comment utiliser les Big Data qui doivent être un levier technologique pour accompagner les enjeux métiers et non l'inverse
- Comprendre les acteurs du Big Data et leur positionnement
- Quelle vision à 3 ans
- Propriété de la donnée, environnement juridique du traitement, sécurité
- La nécessité de la gouvernance des données
- Qu'est-ce qu'un CDO

Aspects juridiques et éthiques : quelles données pour quels usages ?

- Données objectives
- Données à caractère personnel
- Quelle gestion des données personnelles ? (donnée se rapportant à une personne physique, qui peut être identifiée quel que soit le moyen utilisé)
- Quels Impact sur la vie privée
- Surveillance et sanction de la CNIL
- Déclaration préalable
- Exemples
- Présentation du socle (la finalité du traitement) et de 4 conditions
- Finalité explicite et légitime
- Loyauté dans la mise en œuvre du traitement
- Données pertinentes
- Durée de conservation non excessive
- Sécurité

Impact des choix technologiques en matière d'analyse et de visualisation



Certification DIGITT en option, Code
CPF 235908

(Financement possible Action
Collective ATLAS, ex-fafiec)

- Fonctionnement des solutions d'analyse et de visualisation aujourd'hui dans un environnement BI
- Qu'implique la mise en place d'une solution de type Big Data
- Quels outils utiliser pour l'analyse et la visualisation ?
- Les nouveaux outils natifs au Big Data
- Les outils classiques de Datavisualisation
- Quelles performances pour quels outils ?
- Les nouvelles solutions tout en un pour packager des applications Big Data (de l'intégration de la donnée à leur visualisation)
- Gérer la structure des données (structurées – non structurées ; internes – acquises)
- Le Data Lake
- Retour sur le La Datawarehouse historique : en tirer partie.
- Nouvelles approches Big Data : le Data Lake.
- Intégrer toutes les données avec le Data Lake
- Pourquoi faut-il vraiment tout conserver ?
- La nécessaire gouvernance de la donnée.
- Structuration des données (structurées – non structurées ; internes – acquises)
- Température des données (froides, tièdes, chaudes)
- Coloration des données (données blanches, données grises, données sombres)
- Quelle valeur par typologie de données

La collecte de données

- Intégration de données hétérogènes
- Intégration réelle (DataWarehouse) ou virtuelle (fédérateur de données).
- Médiateur, adaptateur et ETL.
- Intégration de schémas et d'ontologies : les conflits sémantiques, le problème de la résolution d'entité.
- Modèles de données pour l'intégration : relationnel, XML, JSON, clé-valeur.
- L'offre produits et services.
- Intégration des données du SI avec big data
- Besoins : acquisition, organisation, analyse.
- Place de Hadoop et Spark dans une architecture d'intégration
- Exemples de plateformes d'intégration : Microsoft HDInsight, IBM InfoSphere BigInsights, Oracle Bigdata Appliance.

L'entrepôt / le stockage de données

- Stockage d'objets
- Stockage en fichiers distribués
- Systèmes de fichiers distribués : Hadoop HDFS, Google File System, IBM GPFS, GlusterFS, Lustre.
- Stockage clé-valeur
- Systèmes clé-valeur : Amazon DynamoDB, Amazon SimpleDB, Apache Cassandra, LinkedIn Voldemort.

Méthodes d'analyse et de visualisation

- La Data Visualisation
- La Data Discovery
- Le self-service BI
- Les nouveaux enjeux
- Marchés mouvants et volatiles
- Accélération du cycle de décision
- Populations métiers plus matures
- Intégration avec le Big Data
- Que visualiser ?

- Exploration de données.
- Performance de modèles.
- Prédiction des modèles.
- Comment visualiser ?
- Types de représentations : courbes, histogrammes, diagrammes 3D
- Statique vs interactif.
- Outils et technologies
- La montée du JavaScript.
- Les solutions du marché
- Les solutions intégrées aux plateformes Big Data

Etudes de cas

- Mise en place d'une architecture Big Data orientée analyse des données et visualisation

Version du : 29/07/2021