**Contactez-nous** Tél. : +33 (0)1 83 35 34 40
Email inscription@lepont-learning.com
Site web www.lepont-learning.com

## CONCEPTS BIG DATA

# Big Data - Enjeux et perspectives du Big Data

***REF : SIBI006***

***DUREE : 14h***

***Mixte Classe virtuelle***

### *PUBLIC*

Cette formation Enjeux et perspectives du Big Data est destinée aux Managers, Chefs et Directeurs de projet Directeurs d'études, Data Stewards.

Modalités et délais d'accés : les inscriptions sont fernées 24h avant la 1ére 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 Enjeux et perspectives du Big Data nécessite une bonne culture générale sur les systèmes d'information et/ou les organisations.

### *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

Certification DIGITT en option, Code CPF 235908

## *OBJECTIFS PEDAGOGIQUES*

Cette formation Enjeux et perspectives du Big Data vous permettra de :

- Identifier les enjeux, opportunités et évolutions du Big Data
- Définir les enjeux stratégiques et organisationnels des projets Big Data en entreprise
- Comprendre les technologies principalement utilisées selon les processus de mise en place des projets Big Data
- Déterminer les éléments de sécurité, d'éthique et les enjeux juridiques
- Consolider ses connaissances à travers un cas d'usage

## *PROGRAMME*

**Big Data : enjeux et opportunités / évolution des données**

- Le Big Data et ses 4 grands défis : volume, variété, vélocité, véracité.
- Données massives : Web, réseaux sociaux, Open Data, capteurs, données industrielles.
- Ouverture des données publiques : le mouvement Open Data.
- Variété, distribution, mobilité des données sur Internet.
- Vélocité et flux continus de données.
- Le Edge Computing, vers une informatique de flux

**Les enjeux pour les entreprises**

- Le processus de création de valeur grâce au Big Data
- Véracité des données provenant de sources variées pour la prise de décision.
- Analyses complexes sur Big Data, Big Analytics.
- Production d'informations en temps réel à partir de Big Data.
- Réactivité : traitement de flux de données en temps réel, Complex Event Processing (CEP).
- Définir sa stratégie de déploiement : Cloud, "on-premise" ou hybride

**Opportunités offertes par les progrès matériels**

- Le stockage : Disques durs HDD vs SSD "on premise" et stockage objet ou bloc dans le Cloud.
- L'impact de la baisse du coût de la mémoire : l'avènement du traitement de données in-memory
- Couplage CPU/GPU pour une meilleure optimisation du calcul parallèle.
- Les stratégies de stockage disque en réseau NAS/SAN/Isilon (stockage élastique) : Quels impacts sur les architectures de gestion de données ?
- Les architectures massivement parallèles (MPP) vs Hadoop
- Les appliances.

**Compréhension des technologies utilisées / motivations**

- Le théorème CAP (Consistency, Availability, Partition tolerance) : analyse et impact.
- La pile logicielle Big Data (la "stack Hadoop")

(Financement possible Action Collective ATLAS, ex-fafiec)

- Les fondamentaux des Big Data
- L'organisation fonctionnelle et les principaux composants : stockage, organisation, traitement, intégration, outils d'analyse.
- La gestion de clusters.
- Les "design pattern" d'Hadoop (architectures Lambda, Kappa, Omega, etc.), pour répondre à quels usages
- Comparaison avec les SGBD relationnels.

**Les différentes méthodologies adaptées au Big Data**

- La gouvernance des projets Big Data

**Qu'est-ce que la Gouvernance des données**

- Gouvernance des projets
- Organisation et rôles
- Les grands principes des méthodes agiles
- Le design thinking pourquoi est-il adapté aux enjeux du Big data ?

**Les principes de base du design thinking**

- Les méthodes de développement
- Les impacts sur l'organisation d'une DSI (privilégier la conception collaborative)
- Intégrer les métiers dès la phase de conception
- Définir les différents ateliers (modeling, intercation design, prototypage)

**La sécurité du Big Data**

- Les 5 cercles de défense du Big Data
- Protection de la donnée
- Sécurité périmétrique
- Gestion des authentifications
- Gestion des autorisations
- Sécurité du système d'exploitation

**Utilisation de la cryptographie**

- Comprendre les fondamentaux (cryptographie, cryptanalyse, kleptographie, etc.).
- Les trois approches de gestion des clés de chiffrement dans une architecture Big Data.
- Les enjeux du chiffrement homomorphe.
- Les outils de data management de la donnée
- Le chiffrement à la volée avec préservation de format.
- Les bonnes pratiques en matière de chiffrement et les recommandations de l'ANSSI et de l'ENISA.

**Autres techniques de protection des données**

- Comprendre le cycle de vie des données dans une architecture Big Data.
- Les outils en charge de la gouvernance et du cycle de vie des données
- La classification des données. L'importance de la classification dans les projets Big Data.
- L'anonymisation et la pseudonymisation des données.

**Éthique et enjeux juridiques / principes généraux de conformité et de responsabilités**

- Responsabilités sur le plan Informatique et Libertés
- Responsabilités sur le plan identités
- Responsabilités sur le plan régulation
- Responsabilités sur le plan économique
- Responsabilités sur le plan sécurité

- Responsabilités sur le plan patrimoine
- Responsabilités sur le plan de l'archivage
- Responsabilités sur le plan ressources humaines

**Cas d'usage, retours d'expériences**

Méthodologie pour aider l'entreprise à devenir Data Centric, c'est-à-dire en partant de zéro, comment se former, se transformer, choisir sa distribution Hadoop, l'installer, la paramétrer, préparer la gouvernance et préparer les premiers usages.

Version du : 25/01/2024