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 - Architecture et infrastructure

REF : MOD_2022363

DUREE : 21h

Mixte Classe virtuelle

PUBLIC

Cette formation Architecture et infrastructure est destinée aux Administrateurs système, exploitants,Architectes d'infrastructures Big Data.

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 Architecture et infrastructure nécessite des connaissances dans l'architecture Linux (navigation et structure) et les fondamentaux de la gouvernance des données.

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 Architecture et infrastructure vous permettra de :

- Définir et identifier le contexte spécifique des projets Big Data
- Connaitre le panorama technologique et enjeux socio-économiques
- Mesurer l'impact des choix technologiques en matière d'analyse et de visualisation
- Connaitre la problématique et les enjeux de l'architecture distribuée
- Intervenir sur la qualité de la donnée en respectant les bonnes pratiques
- Respecter la propriété de la donnée, connaitre l'environnement juridique du traitement et mettre en œuvre la sécurité des données
- Mettre en œuvre une architecture et du calcul distribué
- 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

Certification DIGITT en option, Code CPF 235908

(Financement possible Action Collective ATLAS, ex-fafiec)

- Sécurité

Impact des choix technologiques en matière d'infrastructure et d'architecture Big Data

- Les impacts organisationnels
- Comment positionner le Big Data face à l'existant ?
- Quelles sont les possibilités offertes par le Big Data
- Quels sont les contraintes techniques du Big Data ?
- Quelles stratégies de conservation des données (chaudes, froides, "gelées") dans le temps ?
- Exemples de mise en œuvre d'architectures Big Data
- Faut-il partir sur du commodity Hardware ou sur des appliances

Qualité des données dans les projets Big Data

- Données, information, connaissance
- Le cycle de vie de l'information
- Les données.
- Qu'est-ce que la qualité des données ?
- **La fraîcheur**
- La disponibilité
- La cohérence
- La traçabilité
- La sécurisation
- L'exhaustivité.
- La démarche qualité dans le Big Data
- **Motivation : Les besoins de qualité engendrés par le DataLake**
- La recherche de plus d'efficacité métier
- Facteurs clés de succès et bonnes pratiques
- **Les 7 piliers de la qualité des données**
- Les bonnes questions à se poser
- Le pilotage du projet
- La gestion des grands volumes
- Les outils en charge de la gouvernance et du cycle de vie des données des Big Data

Vers des architectures distribuées

- Rappels des principes de base des architectures distribuées
- Le stockage distribué dans HDFS
- La data localité
- YARN, le super-chef d'orchestre des applications distribuées
- Les différents frameworks distribués
- Découvrir Hive, Pig, Spark, R et Python
- Calculs et traitements distribués de la donnée
- Avantages et inconvénients des architectures distribuées
- Les performances liées aux architectures distribuées

Etude de cas

Mise en œuvre d'une architecture Big Data, conception d'une application de reconnaissance d'images utilisant des Frameworks distribués (Python, Spark)

Version du : 31/05/2022