



Big Data : concepts et enjeux

VOUS ÊTES

Chefs de projets, architectes, data-scientists, et toute personne souhaitant comprendre les impacts du BigData sur l'entreprise au niveau du traitement des données, des architectures, de l'organisation

VOS OBJECTIFS

Comprendre les concepts et les apports du BigData, les impacts sur l'organisation de l'entreprise

VOTRE FORMATION



DURÉE : 1 jour
7 heures



PROCHAINE SESSION :
27 mai 2020



LIEU : En distanciel



PRIX : 690 €
net de taxes

PRÉ-REQUIS

Il est demandé aux participants d'avoir une bonne culture générale sur les systèmes d'information.

MODALITÉS

La formation est accessible à distance en classe virtuelle : accès à l'infrastructure de travaux pratiques, machines physiques, outils pédagogiques, échanges avec le formateur

De 4 à 12 participants

Financement éligible au FNE Formation pour tout salarié d'entreprise en activité partielle

VOTRE CONTACT :



Andrea FALLOURD
Conseillère en formation
06 74 51 44 97
afalourd@itescia.fr

ITESCIA - Campus de Pontoise
8 rue Pierre de Coubertin
95300 PONTOISE

www.itescia.fr



VOTRE PROGRAMME

Introduction

A l'origine du BigData : traitement de volumes importants de données non structurées, traitements optimisés de flux de données au fil de l'eau, liés aux nouvelles technologies et aux nouveaux usages.

Domaines concernés : recherche scientifique, médical, e-commerce, sécurité, prédictif, ...

Exemples : lutte contre la criminalité, fraude, santé, ressources énergétiques

Apports des évolutions techniques sur différents aspects : stockage, indexation/recherche, calcul.

Concepts clés : ETL, Extract Transform Load, CAP, 3V, 4V, données non structurées, prédictif, Machine Learning.

Quelques applications : Amazon Rekognition, Polly

Le positionnement des technologies de cloud, BigData et noSQL, de data-mining.

Éléments d'architecture.

Gouvernance des données : importance de la qualité des données, fiabilité, durée de validité, sécurité des données

Aspects législatifs : sur le stockage, la conservation de données, etc... sur les traitements, la commercialisation des données, des résultats

Stockage distribué

Caractéristiques NoSQL

Les différents modes et formats de stockage. Besoin de distribution.

Définition de la notion d'élasticité.

Principe du stockage réparti, définitions : réplication, sharding, gossip protocol, hachage,

Systèmes de fichiers distribués : GFS, HDFS, Ceph

Les bases de données : Cassandra, HBase, MongoDB, Accumulo, CouchBase, Riak, BigTable, ..

Calcul et restitution

Apport des outils de calculs statistiques

Langages adaptés aux statistiques, liens avec les outils BigData.

Outils de calcul et visualisation : R, SAS, Spark, Tableau, QlikView, ...

Caractéristiques et points forts des différentes solutions.

Evolutions

Liens vers les nouveaux métiers : Hadoop scientists, Data scientists, CDO.

Analyse des données au service de l'entreprise

Rôle de la DSI dans la démarche BigData.

Ouverture sur l'OpenData : principe, la démarche publique, les licences.

Exemple : portail data.gouv.fr

Les offres Saas BigData comme Google BigQuery.

Les limites. Les nouveautés annoncées.

Calculs distribués, machine learning, I.A.

Rejoignez nos réseaux sociaux

