



# Big Data : architecture et technologies

## VOUS ÊTES

---

Chefs de projets, architectes, développeurs, data-scientists, et toute personne souhaitant connaître les outils et solutions pour concevoir et mettre en oeuvre une architecture BigData.

## VOS OBJECTIFS

---

Comprendre les concepts essentiels du BigData, et les technologies implémentées

Analyser les difficultés propres à un projet BigData, les freins, les apports, tant sur les aspects techniques que sur les points liés à la gestion du projet.

## VOTRE FORMATION

---



**DURÉE : 2 jours**  
14 heures



**PROCHAINE SESSION :**  
28 et 29 mai 2020



**LIEU :** En distanciel



**PRIX :** 1 476 €  
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

L'essentiel du BigData : calcul distribué, données non structurées. Besoins fonctionnels et caractéristiques techniques des projets. La valorisation des données. Le positionnement respectif des technologies de cloud, BigData et noSQL, et les liens, implications.  
Concepts clés : ETL, Extract Transform Load, CAP, 3V, 4V, données non structurées, prédictif, Machine Learning.

Exemple d'application : Amazon Rekognition, Polly, EMR.

L'écosystème du BigData : les acteurs, les produits, état de l'art. Cycle de vie des projets BigData.

Emergence de nouveaux métiers : Datascientists, Data labs, Hadoop scientists, CDO,...

Rôle de la DSI dans la démarche BigData. 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.

Les types de bases de données : clé/valeur, document, colonne, graphe. Besoin de distribution. Définition de la notion d'élasticité. Principe du stockage réparti.

Définitions : réplication, sharding, gossip, hachage,

Systèmes de fichiers distribués : GFS, HDFS, Ceph. Les bases de données : Redis, Cassandra, DynamoDB, Accumulo, HBase, MongoDB, BigTable, Neo4j, ..

Données structurées et non structurées, documents, images, fichiers XML, JSON, CSV...

Moteurs de recherche. Principe de fonctionnement. Méthodes d'indexation.

Recherche dans les bases de volumes importants.

Présentation de Elasticsearch et SolR

Principe du schemaless, schéma de stockage, clé de distribution, clé de hachage

### Calcul et restitution, intégration

Différentes solutions : calculs en mode batch, ou en temps réel, sur des flux de données ou des données statiques.

Les produits : langage de calculs statistiques, R Statistics Language, sas, RStudio; outils de visualisation : Tableau, QlikView

Ponts entre les outils statistiques et les bases BigData. Outils de calcul sur des volumes importants : Kafka/Spark Streaming/Storm en temps réel, Hadoop/Spark en mode batch.

Zoom sur Hadoop : complémentarité de HDFS et YARN. Restitution et analyse : logstash, kibana, elk, zeppelin

Principe de map/reduce et exemples d'implémentations, langages et sur-couches.

Présentation de pig pour la conception de tâches map/reduce sur une grappe Hadoop.

Rejoignez nos réseaux sociaux

