

# Formation Big Data : Les techniques d'Analyse et de Visualisation

## Description de la formation Big Data Analyse

### Visualisation:

L'objectif de cette formation est de vous rendre autonome dans l'analyse et la visualisation des données dans un contexte Big Data.

Un premier rappel permet de replacer les technologies dans le contexte du Big Data afin de clairement comprendre pourquoi de nouveaux outils apparaissent par rapport aux standards SQL ou de visualisation.

Ensuite, la progression logique de la formation vous permettra de savoir structurer vos données, les alimenter, les analyser et enfin de les visualiser avec le bon outil. On se situe ainsi dans un scénario cohérent et réaliste autour du cycle de vie des données.

Les nombreuses manipulations réalisées durant la formation vous permettront non seulement de disposer d'une vue générale très précise des différents concepts et outils, y compris sur le plan méthodologique, mais aussi d'être véritablement opérationnel sur les standards du marché tant en terme de langage que de framework de traitement de la donnée. Vous saurez également choisir le bon outil de visualisation pour restituer des présentations dynamiques sur des analyses plus ou moins complexes en communiquant de manière efficace et accessible (Dataviz, Data Storytelling).

# Objectifs de la formation Big Data Analyse

## Visualisation:

### Objectifs opérationnels :

Être autonome dans l'analyse et la visualisation des données dans un contexte Big Data.

### Objectifs pédagogiques :

À l'issue de cette **formation Big Data Analyse Visualisation**, vous aurez acquis les connaissances et compétences nécessaires pour :

- Comprendre les spécificités du Big Data
- Connaître les concepts fondamentaux et technologies associées au Big Data
- Gérer, collecter et explorer des données
- Analyser et visualiser ces données

## À qui s'adresse cette formation ?

### Public :

Cette formation est destinée aux statisticiens, analystes de données (Data Analyst), consultants en informatique décisionnelle (BI), dataminers, développeurs, chefs de projets.

### Pré requis :

Des connaissances de base sur le langage SQL, le développement et les statistiques (de niveau scolaire) sont nécessaires pour suivre ce cours.

## Contenu du cours Big Data Analyse Visualisation

### Comprendre les spécificités du Big Data

Les origines du Big Data

Les données au cœur des enjeux (volume, diversité, IoT etc.)

Les limites des architectures actuelles et de la BI  
Sécurité, éthique, environnement juridique (données personnelles, CNIL, accords internationaux, etc.)

[Les concepts fondamentaux et technologies associées du Big Data \(stockage, recherche, visualisation\)](#)

Le concept de Datalake

L'écosystème Hadoop, quel outil pour quel usage ?

Comprendre Hadoop et ses composants

Le système de fichiers répartis HDFS

Philosophie MapReduce

Quelles différences entre les distributions pour l'analyste ?

Différentes catégories de bases NoSQL (clé/valeur, documents, colonnes, graphes)

La philosophie open source et les alternatives propriétaires

Python, langage phare du Big Data

Solution Cloud ou on Premise, quels impacts ?

Moteur de recherche et Big Data, quand utiliser la stack Elasticsearch, Logstash, Kibana (ELK) ?

Impacts techniques et organisationnel

[Gestion des données structurées ou non](#)

Exemples de données non structurées

Manipuler les données avec différents formats de fichiers :

HDFS

Fichiers plats : CSV, JSON

Structures optimisées : Parquet, Avro

Organisation relationnelle : Hive metastore

Le Master Data Management (MDM) pour réconcilier les référentiels

Structurer vos données avec Hive (bases, tables, etc.)

## La collecte et exploration des données

Intégrer les données avec un ETL

La phase de collecte des données

Les outils dédiés à la collecte:

Utiliser Scoop pour intégrer les données depuis une base de données relationnelle

Travailler en streaming avec Kafka et Spark Streaming

Utiliser le langage Pig latin pour intégrer les données

Utiliser Spark pour la collecte et l'exploration

## L'analyse des données

Analyse en contexte Big Data

Vue générale des différentes méthodes d'analyse (exploration, segmentation, classification, estimation, prédiction)

Data Science et Machine Learning:

quels rôles pour quels usages ?

phase d'un projet de Data Science

domaines d'application

Interagir avec Hadoop en temps réel (traitement parallèle avec Impala, interroger les données Hive avec Spark)

Les frameworks orientés « analyse de données »

## La visualisation des données (Dataviz)

Ce que les statistiques ne disent pas

Les objectifs de la visualisation

Quels graphes pour quels usages ?

Représentation de données complexes (encodage visuel, visualisation interactive)

Exemple de visualisations avec les standards du marché (Tableau, Power BI)

Savoir communiquer sur les analyses de données (Data Storytelling)