

# Formation les fondamentaux du big data et de la data science

## Description

Cette formation comprendre les fondamentaux du big data et de la data science est basée sur des cas pratiques afin de vous familiariser avec les concepts du big data et de la data science.

Vous apprendrez par des exemples à comprendre l'environnement du big data (Hadoop, Spark...) et ses applications (open data, internet des objets...).

Cette formation big data vise un public d'analystes, de chargés d'études voulant comprendre les enjeux liés au big data et ne demande aucun prérequis techniques.

## PLUS D'INFORMATIONS :

Formation disponible en intra ou en inter-entreprises

Inscription : <https://www.stat4decision.com/fr/formations/formation-introduction-au-big-data/>

## Durée

2.00 jours (14.00 heures)

## Profils des stagiaires et prérequis

### Profils :

- Analystes
- Chargés d'études
- Data scientist désirant avoir un état des lieux du domaine
- Tout public intéressé par la compréhension des fondamentaux du big data et de la data science

### Prérequis :

- Avoir quelques connaissances en traitement de données

## Objectifs pédagogiques

- Comprendre les principes du big data
- Maîtriser le vocabulaire lié au big data et à la data science
- Maîtriser les enjeux pour mettre en œuvre une stratégie big data dans votre entreprise
- Connaître les outils et leurs usages

## Programme détaillé

- Le big data :
  - Le big data – définitions et usages
  - Quelques cas d'usages de projets big data
  - Qui pour s'occuper des données dans votre structure : le data scientist, le data analyst, le data engineer, le big data architect...

Stat4decision

37-39 avenue Ledru Rollin Paris 75012

Tel. 01.72.25.40.82 | E-mail : [info@stat4decision.com](mailto:info@stat4decision.com) | [www.stat4decision.com](http://www.stat4decision.com)

Numéro SIRET: 81048985600015 | Numéro de déclaration d'activité: 11755352275 (auprès du préfet de région de: 75)

Organisme de formation certifié Qualiopi pour ses actions de formation

- Les données et leur gestion :
  - Les données non structurées
  - Le passage au big data : quels changements pour votre organisation ?
  - Le data lake et les formats de stockage (HDFS, in memory...), quelle solution choisir ?
  - Hadoop et son environnement (Hive, Pig, MapReduce, Kafka...)
- Les outils pour le stockage et la manipulation des données
  - Le cloud ou on premise ?
  - Les bases de données NoSQL (MongoDB, Cassandra, Redis)
  - Les bases de données basées sur des graphes : neo4j
  - Hadoop et son environnement (Hive, Pig, MapReduce, Apache Ranger pour la sécurité)
  - Kafka pour le traitement des flux de données
  - Spark pour le traitement de données et le data analytics
  - Les autres solutions pour les données sur le cloud (Snowflake, Redshift...)
- La data analytics (predictive analytics) et le machine learning pour l'intelligence artificielle :
  - Différence entre le traitement des données classiques et la data analytics pour le big data
  - Le machine learning : panorama des méthodes et des utilisations (supervisé, non supervisé, renforcement, deep learning)
- Présentation d'outils pour la data analytics de leurs forces et de leurs faiblesses
  - Python et R, les outils open source de la data science
  - Apache Spark, la boîte à outils pour le big data
  - Les plateformes (Dataiku DSS, Databricks, Alteryx, KNIME, Anaconda)
- Les étapes d'un projet big data :
  - La gouvernance des données
  - Le déroulé d'un projet suivant différents cas d'usage
  - L'évaluation de la réussite d'un projet data
- Les objectifs d'une stratégie big data :
  - Comment définir une stratégie big data dans votre entreprise.
  - Présentation de cas réels d'applications big data
  - Comment éviter les pièges liés à un projet big data
  - Exemples de déroulés de projets dans différents domaines (Gestion client, Détection de fraude, Manufacturing...)

## Organisation de la formation

### Moyens pédagogiques et techniques

- Accueil des stagiaires dans une salle dédiée à la formation.
- Documents supports de formation projetés.
- Exposés théoriques
- Etude de cas concrets
- Mise à disposition en ligne de documents supports à la suite de la formation.
- Outils utilisés : des outils du big data seront utilisés avec notamment l'utilisation d'outils du marché en cloud tels que Databricks

### Dispositif de suivi de l'exécution et d'évaluation des objectifs de la formation

- Mises en situation.
- Cas pratiques validés par le formateur.
- Évaluation des connaissances (quizz / tests).

Stat4decision

37-39 avenue Ledru Rollin Paris 75012

Tel. 01.72.25.40.82 | E-mail : [info@stat4decision.com](mailto:info@stat4decision.com) | [www.stat4decision.com](http://www.stat4decision.com)

Numéro SIRET: 81048985600015 | Numéro de déclaration d'activité: 11755352275 (auprès du préfet de région de: 75)

Organisme de formation certifié Qualiopi pour ses actions de formation

**Stat4decision**

37-39 avenue Ledru Rollin Paris 75012

Tel. 01.72.25.40.82 | E-mail : [info@stat4decision.com](mailto:info@stat4decision.com) | [www.stat4decision.com](http://www.stat4decision.com)

Numéro SIRET: 81048985600015 | Numéro de déclaration d'activité: 11755352275 (auprès du préfet de région de: 75)

*Organisme de formation certifié Qualiopi pour ses actions de formation*