

# Panorama et choix des briques d'exploitation d'un data lake

## Description

Une formation pour comprendre les concepts et les choix existants pour la construction d'un data lake.

## Durée

2.00 jours (14.00 heures)

## Profils des stagiaires et prérequis

### Profils :

- Chef de projet DSI désirant faciliter l'exploitation de données déjà présentes dans un data lake
- Manager souhaitant accompagner une maîtrise d'ouvrage technique

### Prérequis :

- Une première utilisation de données provenant d'un data lake
- Connaissances de base en traitement de données (statistique et tables de données)

## Objectifs pédagogiques

- Réaliser un panorama des différents outils et thématiques de data engineering (ETL, Datalake, SQL noSQL)
- Faire une revue des différentes méthodologie de choix d'une technologie
- Proposer et décrire une architecture répondant à un cas d'usage

## Programme détaillé

- Qu'est-ce que le data engineering?
  - Quelle est sa fonction à l'intérieure de l'entreprise
  - Qu'est-ce qu'un data lake
  - Retour sur le rôle des différentes briques qui composent un data lake
- Comprendre les différentes manières de stocker sa donnée, les bases des technologies et les différences d'approches
  - SQL et noSQL
  - HDFS
  - GCS , S3, Azure Data Lake Storage
- Orchestration des transformations et ingestion de la donnée avec un ETL
- Data Processing avec Spark, Flink et Hadoop
- Data Lineage et Data management
- Gestion des accès et de la sécurité
- Analyser sa donnée avec des outils de Data Analytics
- Proposer et documenter une architecture
  - Description visuelle d'une architecture

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

- Utilisation d'une RFC
- Monitorer et manager les performances du data lake
- Mise en places de principes de data gouvernance

## Organisation de la formation

### Moyens pédagogiques et techniques

- Support PDF avec de nombreuses références et cas pratiques
- Utilisation d'une plateforme de test des outils du data lake

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

- Cas pratiques validés par le formateur
- Quiz en classe

**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