

# Formation Big Data, Méthodes et solutions pratiques pour l'analyse des données

## Présentation

Cette formation intermédiaire propose une approche pragmatique du Big Data appliqué à l'analyse des données. Sur trois jours, les participants apprennent à analyser des besoins, mettre en œuvre des pipelines d'ingestion et de préparation, configurer des traitements distribués (batch/stream) et produire des visualisations et indicateurs exploitables en contexte entreprise.

Durée : 21,00 heures (3 jours)

Tarif INTRA : Nous consulter

## Objectifs de la formation

- Analyser des besoins analytiques et définir un périmètre Big Data adapté
- Mettre en œuvre l'ingestion et la préparation à l'échelle (formats colonnes, partitionnement)
- Configurer des traitements distribués (batch/stream) et orchestrer des workflows
- Mettre en œuvre des visualisations et tableaux de bord orientés décision
- Analyser la performance et la qualité des données pour fiabiliser les usages

## Prérequis

- Connaissances de base en SQL
- Notions de manipulation de données (CSV, JSON)
- Familiarité avec un langage de script (Python de préférence)

## Public

- Data analysts, data engineers et développeurs évoluant vers la data à l'échelle
- Équipes BI/IT souhaitant industrialiser l'analyse de données volumineuses

## Programme de la formation



## Jour 1 – Ingestion, stockage et préparation à l'échelle

### Session du matin :

- Panorama Big Data pour l'analyse : cas d'usage, limites et bonnes pratiques
- Modélisation des zones de données (bronze/silver/gold), data lake et lakehouse
- Formats et optimisation : Parquet/ORC, partitionnement, gestion des schémas

### Session de l'après-midi :

- Ingestion multi-sources (fichiers, API) et validation de schémas
- Préparation des données : nettoyage, jointures, agrégations, enrichissement
- Contrôles de qualité : règles, métriques et traçabilité

### TP / Exercice :

- Construire un pipeline d'ingestion et de préparation (CSV/JSON → format colonne) avec contrôle qualité (livrables : scripts, diagramme de flux, rapport qualité)

### Points clés & takeaways :

- Choisir formats et stratégies d'ingestion adaptés au volume
- Mettre en place une chaîne de préparation reproductible et contrôlée

## Jour 2 – Traitements distribués et orchestration

### Session du matin :

- Traitements batch et streaming : principes, usages et arbitrages
- Performance : partitions, parallélisme, coût vs latence
- Tests et observabilité : logs, métriques, profiling de jobs

### Session de l'après-midi :

- Orchestration (Airflow, Prefect) : DAG, dépendances, retries, SLA
- Mise en œuvre d'un job distribué (Spark ou équivalent)
- Reprise sur incident, idempotence et fiabilité

**TP / Exercice :**

- Développer un workflow orchestré pour alimenter un datamart (batch + micro-batch) avec suivi de performance (livrables : DAG, code, métriques)

**Points clés & takeaways :**

- Exécuter et fiabiliser des traitements à l'échelle
- Mettre en place une observabilité utile à l'exploitation

**Jour 3 – Visualisation, KPIs et restitution décisionnelle**

**Session du matin :**

- Conception d'indicateurs : définitions, granularité, périodicité
- Bonnes pratiques de visualisation et storytelling
- Publication et gouvernance : data contracts, sécurité (aperçu)

**Session de l'après-midi :**

- Construction d'un dashboard de bout en bout (KPI, tendances, alertes)
- Documentation et checklist d'exploitation (qualité, coûts, rafraîchissement)
- Revue finale : limites, améliorations et feuille de route

**TP / Exercice :**

- Créer un dashboard alimenté par le pipeline (KPI, filtres, alertes) avec note d'exploitation

**Points clés & takeaways :**

- Produire des restitutions utiles et actionnables
- Relier pipeline, qualité des données et décision métier

## Organisation

### Formateur

Les formateurs de Docaposte Institute sont des experts de leur domaine, disposant d'une expérience terrain qu'ils enrichissent continuellement. Leurs connaissances techniques et pédagogiques sont rigoureusement validées en amont par nos référents internes. Riches de leur expérience sur le sujet, ils sauront accompagner vos collaborateurs dans leur montée en compétences.

### **Moyens pédagogiques et techniques**

- Apports des connaissances communes.
- Mises en situation sur le thème de la formation et des cas concrets.
- Méthodologie d'apprentissage attractive, interactive et participative.
- Equilibre théorie / pratique : 60 % / 40 %.
- Supports de cours fournis au format papier et/ou numérique.
- Ressources documentaires en ligne et références mises à disposition par le formateur.
- Pour les formations en présentiel dans les locaux mis à disposition, les apprenants sont accueillis dans une salle de cours équipée d'un réseau Wi-Fi, d'un tableau blanc ou paperboard. Un ordinateur avec les logiciels appropriés est mis à disposition (le cas échéant).

### **Dispositif de suivi de l'exécution et de l'évaluation des résultats de la formation**

En amont de la formation :

- Recueil des besoins des apprenants afin de disposer des informations essentielles au bon déroulé de la formation (profil, niveau, attentes particulières...).
- Auto-positionnement des apprenants afin de mesurer le niveau de départ.

Tout au long de la formation :

- Évaluation continue des acquis avec des questions orales, des exercices, des QCM, des cas pratiques ou mises en situation...

A la fin de la formation :

- Auto-positionnement des apprenants afin de mesurer l'acquisition des compétences.
- Evaluation par le formateur des compétences acquises par les apprenants.
- Questionnaire de satisfaction à chaud afin de recueillir la satisfaction des apprenants à l'issue de la formation.

- Questionnaire de satisfaction à froid afin d'évaluer les apports ancrés de la formation et leurs mises en application au quotidien.

### **Accessibilité**

Nos formations peuvent être adaptées à certaines conditions de handicap. Nous contacter pour toute information et demande spécifique.