# Test de data lakes et architectures big data (Spark, Hadoop)

*Lien :*
*https://innov-systems.com/formation/test-de-data-lakes-et-architectures-big-data-spark-hadoop*

DURÉE
**5 jours (35h)**

RÉFÉRENCE
**TLA41**

CATÉGORIE
**Tests de Données, BI et Big Data**

Innov Systems

## OBJECTIFS DE LA FORMATION

À l'issue de cette formation, vous serez capable de :

- Comprendre les principes des Data Lakes et des architectures Big Data
- Concevoir une stratégie de test adaptée aux plateformes distribuées
- Vérifier la qualité, l'intégrité et la cohérence des données
- Tester les pipelines ETL/ELT et les traitements Spark et Hadoop
- Évaluer les performances, la scalabilité et la résilience des plateformes Big Data
- Automatiser les tests des données et des traitements distribués
- Intégrer les tests dans les pipelines CI/CD et DataOps
- Mettre en œuvre des mécanismes de monitoring et d'observabilité
- Garantir la sécurité, la gouvernance et la conformité des données
- Réduire les risques liés aux traitements de données massives
- Industrialiser les processus de validation des architectures Big Data
- Réaliser un projet complet de test d'un Data Lake avec Spark et Hadoop

## POUR QUI ?

- Testeurs logiciels (QA Engineers)
- Test Automation Engineers
- Data Engineers
- Big Data Engineers
- Data Analysts
- Data Architects
- Développeurs Spark
- Développeurs Hadoop
- Ingénieurs DataOps
- Responsables Qualité des Données
- Chefs de projet Data
- Toute personne souhaitant tester, valider et industrialiser des plateformes Big Data et des Data Lakes

Innov Systems

## Programme détaillé

### 1 / Introduction aux architectures Big Data

- Concepts fondamentaux des Data Lakes et des plateformes Big Data
- Architecture Hadoop, Spark et écosystèmes distribués
- Enjeux de qualité des données et de validation des traitements

### 2 / Architecture d'un Data Lake

- Zones Bronze, Silver et Gold
- Formats de données (Parquet, ORC, Avro, Delta Lake)
- Gouvernance, catalogage et métadonnées

### 3 / Élaboration d'une stratégie de test Big Data

- Planification des campagnes de tests
- Identification des risques et des critères d'acceptation
- Gestion des jeux de données volumineux

### 4 / Tests de qualité et d'intégrité des données

- Validation des données en entrée et en sortie
- Contrôle de la complétude, de l'unicité et de la cohérence
- Vérification des transformations ETL/ELT

### 5 / Tests des traitements distribués avec Spark et Hadoop

- Validation des jobs Spark et MapReduce
- Tests des pipelines de traitement batch et streaming
- Vérification des performances des traitements distribués

Innov Systems

### 6 / Tests de performance et de montée en charge

- Tests de volumétrie et de scalabilité
- Analyse des temps de traitement et de la consommation des ressources
- Optimisation des performances des clusters

### 7 / Automatisation des tests Data

- Frameworks d'automatisation des tests de données
- Validation automatique des pipelines de données
- Intégration des tests dans les pipelines CI/CD et DataOps

### 8 / Sécurité, gouvernance et conformité

- Contrôle des accès aux données
- Validation des mécanismes de sécurité et de chiffrement
- Gouvernance des données et conformité réglementaire

### 9 / Monitoring, observabilité et gestion des incidents

- Surveillance des pipelines Big Data
- Détection des anomalies et suivi des KPI
- Analyse des erreurs et amélioration continue

### 10 / Cas pratiques et projet de bout en bout

- Validation complète d'un Data Lake basé sur Spark et Hadoop
- Automatisation des tests de qualité des données
- Bonnes pratiques d'industrialisation et de gouvernance des tests Big Data

Inov Systems

## Approche pédagogique

- Support Ecrit et Projection
- Exposés Interactifs, Podcasts et Vidéos
- Brainstorming et Jeux de Rôle
- Cas Pratiques et Labs inclus pour leur impact opérationnel
- Test de Validation des Acquis des Connaissances

## Prochaines dates programmées

| | |
|---|---|
| 12 au 18 Jan. 2027 | Présentiel - Casablanca |
| 08 au 12 Mars 2027 | Présentiel - Casablanca |
| 03 au 07 Mai 2027 | Présentiel - Casablanca |
| 28 Juin au 02 Juil. 2027 | Présentiel - Casablanca |
| 23 au 27 Août 2027 | Présentiel - Casablanca |

Autres dates possibles sur demande. Contactez-nous pour organiser une session intra-entreprise.

## Réservation & Renseignements

**Téléphone :** +212 522 247 210
**Email :** contact@innov-systems.com
**Web :** https//www.innov-systems.com


Innov Systems — Tous droits réservés