Databricks — Plateforme unifiée de data engineering et analytics

Lien : https://innov-systems.com/formation/databricks-plateforme-unifiee-de-data-engineering-et-analytics

DURÉE
5 jours (35h)

RÉFÉRENCE
BDI172

CATÉGORIE
Data Science et Data Engineering

Innov Systems

OBJECTIFS DE LA FORMATION

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

- Comprendre l’architecture de Databricks et du Lakehouse
- Maîtriser Apache Spark dans un environnement cloud
- Développer des pipelines data engineering robustes
- Utiliser Delta Lake pour la gestion fiable des données
- Construire des workflows analytiques et ML
- Optimiser les performances des traitements distribués
- Déployer des solutions data industrialisées
- Assurer la gouvernance et la sécurité des données
- Intégrer Databricks avec les clouds majeurs
- Réduire les coûts et améliorer la scalabilité
- Exploiter MLflow pour le machine learning
- Étudier des cas concrets d’architecture data moderne

POUR QUI ?

- Data engineers
- Data scientists
- ML engineers
- Data architects
- Cloud engineers
- DevOps engineers
- BI developers
- Analytics engineers
- Platform engineers
- Consultants data & cloud
- SRE
- Toute personne impliquée dans les architectures data modernes

Innov Systems

Programme détaillé

1 / Introduction à Databricks

- Positionnement dans l’écosystème data moderne
- Concept de Lakehouse (data lake + data warehouse)
- Cas d’usage data engineering, ML et analytics

2 / Architecture de Databricks

- Workspace et clusters
- Data Lakehouse et Delta Lake
- Intégration avec cloud providers (AWS, Azure, GCP)

3 / Apache Spark dans Databricks

- Moteur de traitement distribué
- Optimisation des jobs Spark
- Gestion des ressources et scaling automatique

4 / Delta Lake et gestion des données

- ACID transactions sur data lakes
- Time travel et versioning des données
- Fiabilité et qualité des datasets

5 / Data engineering sur Databricks

- ETL/ELT avec notebooks
- Pipelines de données automatisés
- Orchestration des workflows

Innov Systems

6 / Notebooks collaboratifs

- Développement en Python, SQL, Scala, R
- Collaboration en temps réel
- Versioning et reproductibilité

7 / Machine Learning avec MLflow

- Tracking des expériences ML
- Gestion du cycle de vie des modèles
- Déploiement de modèles en production

8 / Performance et optimisation

- Optimisation des clusters Spark
- Partitionnement et caching
- Réduction des coûts de calcul

9 / Gouvernance et sécurité des données

- Unity Catalog et contrôle d’accès
- Data lineage et audit
- Conformité et sécurité des données

10 / Cas d’usage et industrialisation

- Data pipelines en production
- Analytics à grande échelle
- Architecture data moderne unifiée

Innov 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

30 Nov. au 04 Déc. 2026	Distanciel
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

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