DuckDB : La Base Analytique Embarquée pour le Traitement de Données en Local

Lien : https://innov-systems.com/formation/duckdb-la-base-analytique-embarquee-pour-le-traitement-de-donnees-en-local

DURÉE
5 jours (35h)

RÉFÉRENCE
BDI76

CATÉGORIE
Big Data - Analyse des Données et Datavisualisation

Innov Systems

OBJECTIFS DE LA FORMATION

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

- Comprendre l’architecture et les concepts de DuckDB
- Maîtriser le traitement analytique local de données
- Exploiter efficacement les formats Parquet, CSV et JSON
- Réaliser des analyses avancées avec SQL
- Intégrer DuckDB dans des workflows Data Science
- Optimiser les performances des traitements analytiques
- Utiliser DuckDB avec Python, Pandas et Polars
- Construire des pipelines analytiques légers et performants
- Réduire la dépendance aux infrastructures lourdes
- Choisir les cas d’usage adaptés à DuckDB
- Développer des solutions analytiques embarquées
- Étudier des cas concrets de traitement de données

POUR QUI ?

- Data analysts
- Data scientists
- Data engineers
- Business analysts
- Développeurs Python
- Architectes data
- Ingénieurs Big Data
- Chercheurs et universitaires
- Consultants en data analytics
- Responsables BI et reporting
- Toute personne souhaitant analyser efficacement des données en local

Innov Systems

Programme détaillé

1 / Introduction à DuckDB

- Présentation de DuckDB et de son positionnement
- Différences entre OLTP et OLAP
- Cas d’usage de l’analytique embarquée

2 / Architecture et fonctionnement de DuckDB

- Moteur analytique orienté colonnes
- Exécution vectorisée des requêtes
- Gestion optimisée de la mémoire

3 / Installation et prise en main

- Installation sur différents environnements
- Interfaces SQL et API disponibles
- Premières requêtes analytiques

4 / Manipulation et chargement des données

- Import de fichiers CSV, JSON et Parquet
- Lecture directe des données sans ingestion
- Gestion des grands volumes de données

5 / SQL analytique avancé avec DuckDB

- Fonctions d’agrégation et analytiques
- Fenêtres analytiques (Window Functions)
- Jointures et requêtes complexes

Innov Systems

6 / Intégration avec l’écosystème Data

- Utilisation avec Python, R et Pandas
- Intégration avec Jupyter Notebooks
- Compatibilité avec Apache Arrow et Polars

7 / Optimisation des performances

- Compression et stockage en colonnes
- Optimisation des plans d’exécution
- Traitement efficace des jeux de données massifs

8 / Cas d’usage analytiques

- Exploration de données (EDA)
- Reporting local et tableaux de bord
- Analyse de fichiers de logs et données métier

9 / Comparaison avec d’autres solutions analytiques

- DuckDB vs SQLite
- DuckDB vs PostgreSQL analytique
- DuckDB vs Spark et moteurs distribués

10 / Mise en œuvre et bonnes pratiques

- Construction de pipelines analytiques locaux
- Gestion des performances et des ressources
- Études de cas et projets pratiques

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

18 au 22 Jan. 2027	Présentiel - Casablanca
15 au 19 Mars 2027	Présentiel - Casablanca
10 au 14 Mai 2027	Présentiel - Casablanca
05 au 09 Juil. 2027	Présentiel - Casablanca
30 Août au 03 Sep. 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