Spark SQL : maîtrisez le traitement de données massives et l’analyse Big Data en temps réel

Lien : https://innov-systems.com/formation/spark-sql-maitrisez-le-traitement-de-donnees-massives-et-l analyse-big-data-en-temps-reel

DURÉE
5 jours (35h)

RÉFÉRENCE
BDI62

CATÉGORIE
Big Data : développement, langages et NOSQL

Innov Systems

OBJECTIFS DE LA FORMATION

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

- Comprendre le fonctionnement de Spark et Spark SQL
- Maîtriser le traitement de données massives avec SQL
- Savoir manipuler DataFrames et Datasets
- Optimiser les requêtes et performances Spark
- Intégrer Spark SQL dans des pipelines Big Data
- Exploiter les fonctions analytiques avancées
- Traiter des données structurées à grande échelle
- Développer des solutions data performantes en production

POUR QUI ?

- Data engineers
- Data analysts
- Data scientists
- Développeurs Big Data
- Architectes data
- Ingénieurs logiciels
- Toute personne travaillant sur le traitement de données massives

Innov Systems

Programme détaillé

1 / Introduction à Apache Spark et Spark SQL

- Rôle de Spark dans le Big Data
- Positionnement de Spark SQL dans l'écosystème
- Avantages du traitement distribué

2 / Architecture de Spark

- Driver, Executors et Cluster Manager
- RDD, DataFrame et Dataset
- Fonctionnement du moteur d'exécution

3 / Fondamentaux de Spark SQL

- Introduction au langage SQL dans Spark
- Requêtes structurées sur données massives
- Intégration avec les DataFrames

4 / DataFrames et Datasets

- Création et manipulation des DataFrames
- Typage avec Datasets
- Transformations et actions

5 / Lecture et écriture de données

- Sources de données (CSV, JSON, Parquet)
- Connexion aux bases de données
- Optimisation des formats de stockage

Innov Systems

6 / Optimisation des requêtes SQL

- Catalyst Optimizer
- Plans d’exécution
- Techniques d’optimisation des performances

7 / Fonctions SQL avancées

- Jointures et agrégations
- Fenêtrage (window functions)
- Fonctions analytiques

8 / Spark SQL et Big Data pipelines

- Intégration avec Hadoop et Hive
- Traitement batch et streaming
- Pipelines de données

9 / Performance et tuning

- Partitionnement et cache
- Gestion de la mémoire
- Optimisation des jobs Spark

10 / Cas d’usage et production

- Analyse de données massives en entreprise
- Reporting et BI avec Spark SQL
- Bonnes pratiques en production

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

22 au 26 Fév. 2027	Présentiel - Casablanca
19 au 23 Avr. 2027	Présentiel - Casablanca
14 au 18 Juin 2027	Présentiel - Casablanca
09 au 13 Août 2027	Présentiel - Casablanca
04 au 08 Oct. 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