Big Data : intégration SQL, Hive, Impala, Phoenix, Spark

Lien : https://innov-systems.com/formation/big-data-integration-sql-hive-impala-phoenix-spark

DURÉE
3 jours (21h)

RÉFÉRENCE
BSI431

CATÉGORIE
Big Data : développement, langages et NOSQL

Innov Systems

OBJECTIFS DE LA FORMATION

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

- Comprendre l’architecture et les interactions entre les bases relationnelles et NoSQL dans un environnement Big Data
- Maîtriser l’utilisation et l’optimisation des outils de requêtage distribués tels que Hive, Impala, Phoenix, Presto et Spark SQL
- Savoir manipuler et analyser de gros volumes de données dans différents formats (Parquet, ORC, CSV…) via des ateliers pratiques
- Acquérir une méthodologie pour comparer et choisir les solutions adaptées selon les besoins et contraintes

POUR QUI ?

- Ingénieurs et experts en bases de données
- Chefs de projet et architectes Big Data
- Développeurs et analystes de données
- Toute personne amenée à interagir avec des plateformes de traitement de données massives

Programme détaillé

1 / Introduction au Big Data et aux architectures hybrides SQL/NoSQL

- Présentation des enjeux du Big Data et des besoins métiers
- Panorama des architectures distribuées et composants clés (HDFS, HBase, Cassandra)
- Formats de stockage : Parquet, ORC, JSON, CSV, clés/valeurs
- Rôle et positionnement des moteurs de requêtes distribuées
- Liens et intégrations entre SQL traditionnel et NoSQL

2 / Hive et optimisation des requêtes

- Fonctionnement interne de Hive et architecture d'exécution
- HiveQL : syntaxe et extensions spécifiques
- Intégration avec le moteur Tez pour l'optimisation
- Création et gestion de tables internes et externes
- Connexion et interaction avec HBase
- Atelier pratique : requêtage optimisé sur un gros volume de données

3 / Impala et Phoenix : requêtes interactives en temps quasi réel

- Présentation d'Impala : usages, architecture, intégration avec Hive Metastore
- Techniques pour améliorer les performances d'Impala
- Présentation de Phoenix : rôle et utilisation avec HBase
- Optimisations des requêtes Phoenix et gestion des index
- Atelier pratique : benchmarks comparatifs Impala vs Phoenix

4 / Presto et Drill : moteurs multi-sources

- Fonctionnement et cas d'usage de Presto
- Connexion à des sources hétérogènes (PostgreSQL, Cassandra, Hive…)

Inov Systems

- Présentation d’Apache Drill et son indépendance vis-à-vis d’Hadoop
- API JDBC/ODBC : intégration avec des outils externes
- Atelier pratique : requêtage fédéré entre plusieurs bases et formats de données

5 / Spark SQL et DataFrame API

- Présentation de Spark SQL et des DataFrames
- Différences entre RDD, DataFrame et Dataset
- Utilisation du Hive Metastore avec Spark
- Optimisation via le Catalyst Optimizer
- Atelier pratique : analyses complexes sur HDFS et HBase avec Spark SQL

6 / Comparaison et bonnes pratiques de choix

- Compatibilité ANSI SQL et limitations spécifiques
- Comparatif des performances et scénarios adaptés
- Coûts, contraintes et évolutivité des solutions
- Stratégie de sélection selon les besoins métiers et techniques
- Etude de cas : choix d’un moteur de requête pour un projet Big Data

7 / Synthèse et perspectives

- Récapitulatif des outils et de leurs forces/faiblesses
- Nouvelles tendances dans les moteurs de requêtes distribuées
- Conseils pour l’industrialisation et la mise 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

03 au 05 Fév. 2027	Présentiel - Casablanca
31 Mars au 02 Avr. 2027	Présentiel - Casablanca
26 au 28 Mai 2027	Présentiel - Casablanca
21 au 23 Juil. 2027	Présentiel - Casablanca
22 au 24 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

Document généré le 25/11/2026 — Réf : BSI431
Innov Systems — Tous droits réservés

Innov Systems