Hive, Pig & Impala : maîtrisez l’analyse Big Data sur Hadoop de batch à temps réel

Lien : https://innov-systems.com/formation/hive-pig-impala-maitrisez-lanalyse-big-data-sur-hadoop-de-batch-a-temps-reel

DURÉE
5 jours (35h)

RÉFÉRENCE
BDI39

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’écosystème Hadoop et ses composants analytiques
- Maîtriser Hive pour le requêtage SQL sur Big Data
- Utiliser Pig pour les traitements de données complexes
- Exploiter Impala pour des requêtes rapides en temps réel
- Savoir choisir le bon outil selon le besoin
- Optimiser les performances des traitements Big Data
- Construire des pipelines de données efficaces
- Développer une expertise en analyse de données massives

POUR QUI ?

- Data engineers
- Data analysts
- Développeurs Big Data
- Architectes data
- Administrateurs Hadoop
- Ingénieurs BI
- Toute personne travaillant sur des environnements Hadoop

Innov Systems

Programme détaillé

1 / Introduction à l’écosystème Hadoop

- Rôle de Hadoop dans le Big Data
- Architecture globale (HDFS, MapReduce, YARN)
- Positionnement de Pig, Hive et Impala

2 / Présentation de Apache Hive

- Définition et cas d’usage
- HiveQL et requêtage SQL-like
- Tables internes et externes

3 / Architecture de Hive

- Metastore et stockage des données
- Exécution des requêtes (MapReduce / Tez / Spark)
- Optimisation des performances

4 / Introduction à Apache Pig

- Langage Pig Latin
- Utilisation pour le traitement de données
- Scripts et workflows de transformation

5 / Traitement des données avec Pig

- Chargement et manipulation des données
- Transformations et filtrage
- Agrégation et jointures

Innov Systems

6 / Introduction à Apache Impala

- Requêtes SQL en temps réel sur Hadoop
- Différences avec Hive et Pig
- Cas d’usage analytiques

7 / Requêtage rapide avec Impala

- Exécution de requêtes interactives
- Optimisation des temps de réponse
- Analyse en quasi temps réel

8 / Intégration Hive, Pig et Impala

- Complémentarité des outils
- Choix selon les cas d’usage
- Architecture hybride

9 / Optimisation des performances

- Indexation et partitionnement
- Compression des données
- Gestion des ressources cluster

10 / Bonnes pratiques et cas d’usage

- Pipelines de données Big Data
- Traitement batch vs temps réel
- Bonnes pratiques en production

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

08 au 12 Fév. 2027	Présentiel - Casablanca
05 au 09 Avr. 2027	Présentiel - Casablanca
31 Mai au 04 Juin 2027	Présentiel - Casablanca
26 au 30 Juil. 2027	Présentiel - Casablanca
20 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

Innov Systems — Tous droits réservés