Développement de Scripts avec Pig

Lien : https://innov-systems.com/formation/developpement-de-scripts-avec-pig-

DURÉE
3 jours (21h)

RÉFÉRENCE
BSI30

CATÉGORIE
Big Data : développement, langages et NOSQL

OBJECTIFS DE LA FORMATION

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

- Avoir une vue d'ensemble de l'architecture de Apache Pig
- Connaître les fonctionnalités générales et les capacités de Apache Pig
- Développer des requêtes en latin, pour effectuer des transformations sur des données et des analyses de données
- Charger et de manipuler des données de types et formats différents
- Intégrer des données de différents formats

POUR QUI ?

- Chefs de projet
- Data scientists
- Développeurs souhaitant utiliser pig pour l'analyse de données

Innov Systems

Programme détaillé

1 / Introduction

- Le projet Apache Pig, fonctionnalités, versions
- Présentation de Pig dans l'écosystème Hadoop
- Chaîne de fonctionnement
- Comparatif avec l'approche Hive ou Spark

2 / Exploration de l’outil Apache Pig

- Pig : définition, caractéristiques et rayon d’action
- Les cas d’utilisation de Pig
- Rappels sur les commandes HDFS
- Pré-requis techniques
- Configuration de Pig
- Exécuter les différents modes (interactif ou batch)
- Principe de l'exécution de scripts Pig Latin avec Grunt

3 / Le langage Pig Latin

- Présentation du Pig Latin : Utilisation, Fonctionnement
- Modèles de données avec Pig
- Intégration Pig avec MapReduce
- Chargement des données : PigStorage, Schémas, Commande DESCRIBE
- Sauvegarde de données : DUMP, STORE
- Ordres de base : LOAD, FOREACH, FILTER, STORE
- Création d'un ETL de base
- Contrôle d'exécution

4 / Transformation des données via des opérations de filtre avec Pig Latin

Inov Systems

- Groupements, jointures, tris, produits cartésiens
- Transformation de base de la donnée
- Découpages
- Découpages sur filtres

5 / **Analyse de la donnée**

- Échantillonnages
- Filtres
- Rangements avec rank et dense
- Calculs : min/max, sommes, moyennes, ...
- Traitements de chaînes de caractères et traitement de dates

6 / **Intégration**

- Formats d'entrées/sorties
- Interfaçage avro, json
- Cargement de données depuis HDFS vers HBase
- Analyse de données Pig/HBase et restitution Json

7 / **Extensions du Pig Latin**

- Extension du Pig Latin
- Création de fonctions UDF en Java
- Intégration dans les scripts Pig
- Utilisation de Pig Latin depuis des programmes Python
- Exécution de programmes externes, streaming

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

09 au 11 Déc. 2026	Distanciel
27 au 29 Jan. 2027	Présentiel - Casablanca
24 au 26 Mars 2027	Présentiel - Casablanca
19 au 21 Mai 2027	Présentiel - Casablanca
14 au 16 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