Maîtriser Apache Hop pour l’intégration et la transformation des données

Lien : https://innov-systems.com/formation/matriser-apache-hop-pour-lintegration-et-la-transformation-des-donnees

DURÉE
4 jours (28h)

RÉFÉRENCE
BSI424

CATÉGORIE
Big Data : développement, langages et NOSQL

Innov Systems

OBJECTIFS DE LA FORMATION

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

- Comprendre en profondeur l’architecture et l’écosystème Apache Hop
- Concevoir, automatiser et optimiser des flux ETL performants
- Sécuriser et documenter ses pipelines pour un usage en production

POUR QUI ?

- Développeurs, data engineers et administrateurs amenés à gérer des flux de données inter-applications
- Responsables BI en charge de l’alimentation et de la maintenance de systèmes décisionnels
- Toute personne souhaitant maîtriser un ETL open source moderne

Programme détaillé

1 / Introduction et mise en place de l'environnement

- Comprendre le rôle et la philosophie d'Apache Hop dans l'écosystème ETL
- Identifier les différences clés avec d'autres ETL (Pentaho, Talend, etc.)
- Installation, configuration et prise en main de l'interface Hop GUI et Hop Web

2 / Bases des workflows et pipelines

- Comprendre la structure et les composants essentiels d'un pipeline
- Créer un premier workflow d'extraction et transformation de données
- Exécuter et monitorer un pipeline pas à pas

3 / Connexion aux sources et cibles de données

- Configurer l'accès aux bases relationnelles, fichiers plats, API et autres formats
- Gérer les métadonnées pour centraliser les définitions de connexion
- Réaliser des jointures et liaisons multi-sources

4 / Transformation et nettoyage des données

- Appliquer des filtres avancés pour épurer le flux
- Effectuer des calculs et transformations sur les champs
- Manipuler des chaînes, dates et types numériques avec précision

5 / Enrichissement et gestion des variables

- Générer et exploiter des variables globales ou locales
- Injecter des données calculées dans un flux en cours
- Utiliser des sous-pipelines pour modulariser les traitements

Innov Systems

6 / Automatisation et boucles

- Mettre en place des boucles contrôlées avec paramètres
- Réutiliser les résultats intermédiaires pour optimiser les traitements
- Planifier des exécutions récurrentes via Hop Orchestration et intégration cron

7 / Gestion des erreurs et optimisation des performances

- Implémenter une stratégie de gestion d’erreurs et de logs
- Déboguer et analyser les anomalies rencontrées dans un flux
- Utiliser la parallélisation et les options d’optimisation de performances

8 / Industrialisation et bonnes pratiques

- Organiser et documenter ses pipelines pour la maintenance
- Importer, exporter et versionner ses développements
- Déployer en environnement de production avec supervision et alertes

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

Inov Systems

Prochaines dates programmées

26 au 29 Jan. 2027	Présentiel - Casablanca
23 au 26 Mars 2027	Présentiel - Casablanca
18 au 21 Mai 2027	Présentiel - Casablanca
13 au 16 Juil. 2027	Présentiel - Casablanca
14 au 17 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

Innov Systems