Pentaho Data Integration

Lien : https://innov-systems.com/formation/pentaho-data-integration

DURÉE
4 jours (28h)

RÉFÉRENCE
BSI369

CATÉGORIE
Outils Décisionnels : Essbase, Hyperion, Matlab, Jaspersoft, Talend, Qlik, Cognos,...

OBJECTIFS DE LA FORMATION

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

- Comprendre l’environnement Pentaho et son fonctionnement
- Maîtriser l'utilisation d'un ETL (Extract Transform Load)

POUR QUI ?

- Utilisateurs

Innov Systems

Programme détaillé

1 / Introduction à l’outil Pentaho Data Integration

- Pourquoi Pentaho Data Integration (PDI) ?
- Historique et présentation PDI
- Installation et configuration de l'outil

2 / Génération d’une première extraction de données

- Les transformations et tâches (ordonnancement des transformations)
- De l'extraction à l'alimentation de données
- Appréhender et gérer les flux de data
- Exécution d'une transformation d'une tâche

3 / Le référentiel

- Créer un référentiel
- Alimenter un référentiel

4 / Lecture et écriture de fichier

- Données d’entrée et de sortie
- Nettoyage de données
- Paramétrer le chemin d’accès au fichier dans les propriétés de Kettle
- Générer des données de sortie multiples grâce aux paramètres
- Créer un fichier sérialisable à partir de plusieurs fichiers
- Désérialiser un fichier

5 / Interaction avec les bases de données

- Bases de données supportées

Innov Systems

- Se connecter et explorer une base de données source
- La notion de métadonnées
- Récupération d'une table d'une métadonnée
- Epuration de données
- Alimentation en Insert / Update
- Tables input et paramètres timestamp
- L’assistant copie de table (Copy Table Wizard)

6 / Manipuler les données

- Trier son flux en ascendant ou descendant
- Dédoubler son flux
- Filtrer ses données selon plusieurs critères (alléger le flux de données)
- Extraction d'informations à partir d'un champ (chaîne de caractères)
- Remplacer une donnée par une autre
- Gestion des opérateurs / opérations de calculs sur le flux
- Bien utiliser le produit cartésien
- Joindre les informations issues de données hétérogènes
- Comparer des flux de données
- Utiliser les agrégations de valeurs (sommes, moyennes, maximum, etc.)

7 / Enrichir son flux de données

- Génération de logs
- Création / Récupération des variables (dates, numériques, alphanumériques)
- Utilisation du résultat d'un flux
- Les propriétés d’un flux et de son ordonnanceur

8 / Les boucles

- Les problématiques
- Les boucles avec paramètres et le composant « Copie lignes vers résultat »

Innov Systems

- Les boucles avec les composants

9 / Gestion des logs et des erreurs

- Les types d'erreurs
- Gérer les erreurs (dans un job, une transformation)
- Paramètres de gestion d'erreurs
- Le composant "Data Validation"
- Les niveaux de logs (via la console, sauvegardés en BDD...)

10 / Exploitation

- Comprendre les erreurs et lancer des alertes (Débogage)
- La parallélisation (exécution simultanée de plusieurs flux de données)
- Export des travaux
- Utiliser le kitchen pour exécuter des jobs via la ligne de commande
- Documenter ses jobs

11 / Planification

- Paramétrer le planificateur
- Planifier un job
- Pilotage des jobs planifiés

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

23 au 26 Fév. 2027	Présentiel - Casablanca
20 au 23 Avr. 2027	Présentiel - Casablanca
15 au 18 Juin 2027	Présentiel - Casablanca
10 au 13 Août 2027	Présentiel - Casablanca
05 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

Innov Systems