



# Penthao

Mise à jour nov. 2023

**Durée** 3 jours (21 heures )

« Délai d'accès maximum 1 mois »

## OBJECTIFS PROFESSIONNELS

- Implémenter une solution ETL avec PDI.
- Découvrir le traitement et l'alimentation des données, la sécurisation et l'optimisation des flux.

## PARTICIPANTS

- Administrateurs systèmes/BDD, Chefs de projets, Développeurs

## PRE-REQUIS

- Connaissance du SQL.

## MOYENS PEDAGOGIQUES

- Réflexion de groupe et apports théoriques du formateur
- Travail d'échange avec les participants sous forme de
- Utilisation de cas concrets issus de l'expérience professionnelle
- Validation des acquis par des questionnaires, des tests d'évaluation, des mises en situation et des jeux pédagogiques.
- Remise d'un support de cours.

## MODALITES D'EVALUATION

- Feuille de présence signée en demi-journée,
- Evaluation des acquis tout au long de la formation,
- Questionnaire de satisfaction,
- Positionnement préalable oral ou écrit,
- Evaluation formative tout au long de la formation,
- Evaluation sommative faite par le formateur ou à l'aide des certifications disponibles,
- Sanction finale : Certificat de réalisation, certification éligible au RS selon l'obtention du résultat par le stagiaire

## MOYENS TECHNIQUES EN PRESENTIEL

- Accueil des stagiaires dans une salle dédiée à la formation, équipée d'ordinateurs, d'un vidéo projecteur d'un tableau blanc et de paperboard. Nous préconisons 8 personnes maximum par action de formation en présentiel

## MOYENS TECHNIQUES DES CLASSES EN CAS DE FORMATION DISTANCIELLE

- A l'aide d'un logiciel comme Teams, Zoom etc... un micro et éventuellement une caméra pour l'apprenant,
- suivez une formation uniquement synchrone en temps réel et entièrement à distance. Lors de la classe en ligne, les apprenants interagissent et communiquent entre eux et avec le formateur.
- Les formations en distanciel sont organisées en Inter-Entreprise comme en Intra-Entreprise.
- L'accès à l'environnement d'apprentissage (support de cours, labs) ainsi qu'aux preuves de suivi et d'assiduité (émargement, évaluation) est assuré. Nous préconisons 4 personnes maximum par action de formation en classe à distance

## ORGANISATION

- Les cours ont lieu de 9h à 12h30 et de 14h à 17h30.

## PROFIL FORMATEUR

- Nos formateurs sont des experts dans leurs domaines d'intervention
- Leur expérience de terrain et leurs qualités pédagogiques constituent un gage de qualité.

## A L'ATTENTION DES PERSONNES EN SITUATION DE HANDICAP

- Les personnes atteintes de handicap souhaitant suivre cette formation sont invitées à nous contacter directement, afin d'étudier ensemble les possibilités de suivre la formation.

## Programme de formation

### Introduction et découverte de l'environnement (02h30)

- Pourquoi Pentaho Data Integration (PDI) ?
- Historique et présentation
- Architecture, vue d'ensemble des composants.
- Installation et Configuration

### Découverte des flux (03h15)

- Les transformations
- Les jobs
- Text file Output et CSV Input
- Exécuter un job.
- Exécuter une transformation.

### Traitement des données (02h30)

- Qualifier les données à traiter : Sort, Unique et Filter Rows...
- Gérer les chaînes de caractères : String cut
- Effectuer des calculs : Calculator
- Joindre des données : Join Rows, Merge Join et Rows...).

### Les métadonnées (00h45)

- Métadonnées : vue d'ensemble et création.

### Travailler avec les bases de données (02h00)

- Bases de données supportées.
- Récupérer une table d'une métadonnée.
- Interaction avec la base de données.

### Développement de transformations avancées (03h15)

- Propriétés d'un job et transformation.
- Utiliser des variables : Set / Get Variables.
- Copy / Get Rows to Result.
- Alimenter des données via des services Web
- Monitoring (" Write to log " et mails).

### Gestion des logs et des erreurs (03h15)

- 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 base...).

### Fonctionnalités avancées, automatisation et

#### documentation (03h15)

- Débogage (prévisualisation et ajout de points d'arrêt).
- Parallélisation des traitements.
- Export des travaux.
- Kitchen : l'exécuteur de jobs (fichiers, repository, tâches planifiées).
- Documenter ses jobs.