

## PENTHAO

<b>Durée</b>	<b>3 jours</b>	<b>Référence Formation</b>	<b>4-OS-PENTH</b>
--------------	----------------	----------------------------	-------------------

### Objectifs

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

### Pré-requis

Connaissance du SQL.

### Moyens pédagogiques

Réflexion de groupe et apports théoriques du formateur

Travail d'échange avec les participants sous forme de réunion-discussion

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.

### PROGRAMME

#### Introduction et découverte de l'environnement

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

#### Découverte des flux

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

#### Traitement des données

- 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

- Métadonnées : vue d'ensemble et création.

#### Travailler avec les bases de données

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

#### Développement de transformations avancées

- 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**

- 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**

- 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.