

# Formation : Apache Hop, orchestrer des flux de données

Gérez visuellement vos processus ETL

Formation pratique - 3j - 21h00 - Réf. HOA

Prix : 2010 € H.T.

NEW

Les données sont vitales pour les entreprises. Apache Hop est un logiciel open source puissant et personnalisable pour la gestion des processus ETL (Extract, Transform, Load). Il vous permettra d'automatiser la collecte, la transformation et l'organisation des données provenant de sources hétérogènes puis de les envoyer vers une source précise. Gagnez du temps en maîtrisant les pipelines et les workflows à l'aide d'une interface accessible.

## Objectifs pédagogiques

À l'issue de la formation, le participant sera en mesure de :

- ✓ Comprendre et savoir expliquer l'environnement Hop et son fonctionnement
- ✓ Savoir opérer des extractions et des transformations de données
- ✓ Distinguer et ordonner les différentes sources de données
- ✓ Combiner, classifier et comparer les différents types de données
- ✓ Appliquer des automatisations de tâches et analyser les erreurs pour y répondre

## Public concerné

Toute personne alimentant et manipulant des flux de données d'une base de données décisionnelle.

## Prérequis

Bonne connaissance du langage SQL.

Vérifiez que vous avez les prérequis nécessaires pour profiter pleinement de cette formation en faisant [ce test](#).

## Méthodes et moyens pédagogiques

### Exercice

Application de la théorie sur des cas concrets, échanges en groupe, pratique.

### Méthodes pédagogiques

Pédagogie active.

### PARTICIPANTS

Toute personne alimentant et manipulant des flux de données d'une base de données décisionnelle.

### PRÉREQUIS

Bonne connaissance du langage SQL.

### COMPÉTENCES DU FORMATEUR

Les experts qui animent la formation sont des spécialistes des matières abordées. Ils ont été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours qu'ils enseignent. Ils ont au minimum cinq à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

### MODALITÉS D'ÉVALUATION

Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de QCM, mises en situation, travaux pratiques...

Le participant complète également un test de positionnement en amont et en aval pour valider les compétences acquises.

## Modalités d'évaluation

Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de QCM, mises en situation, travaux pratiques...

Le participant complète également un test de positionnement en amont et en aval pour valider les compétences acquises.

## Programme de la formation

### 1 Présentation de Apache Hop

- Pourquoi Hop ?
- Historique et présentation au sens large
- Installation et configuration de l'outil

#### Travaux pratiques

Installer et configurer Apache Hop.

### 2 Génération d'une première extraction de données

- Les workflows et pipelines (ordonnancement)
- De l'extraction à l'alimentation de données
- Appréhender et gérer les flux de data
- Exécution d'un pipeline et d'un workflow

#### Travaux pratiques

Concevoir un pipeline et élaborer un workflow.

### 3 Accéder aux données sources et cibles

- La notion de métadonnées
- Configurer l'accès aux sources de données
- Les sources/cibles supportées
- Les liens entre les sources (jointures)
- Alimentation en Insert/Update (insertion/mise à jour)

#### Travaux pratiques

Configurer l'accès aux sources de données en identifiant les cibles/sources supportées et les jointures.

### 4 Manipuler les données

- Trier son flux ascendant ou descendant
- Dédoubler son flux
- Filtrer ses données selon plusieurs critères (alléger le flux)
- 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

#### Travaux pratiques

Manipuler les flux en triant, dédoublant, filtrant les données. Comparer les flux.

## MOYENS PÉDAGOGIQUES ET TECHNIQUES

- Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : aides audiovisuelles, documentation et support de cours, exercices pratiques d'application et corrigés des exercices pour les formations pratiques, études de cas ou présentation de cas réels pour les séminaires de formation.
- À l'issue de chaque formation ou séminaire, ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques.
- Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le participant a bien assisté à la totalité de la session.

## MODALITÉS ET DÉLAIS D'ACCÈS

L'inscription doit être finalisée 24 heures avant le début de la formation.

## ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Pour toute question ou besoin relatif à l'accessibilité, vous pouvez joindre notre équipe PSH par e-mail à l'adresse [psh-accueil@orsys.fr](mailto:psh-accueil@orsys.fr).

## 5 Enrichir son flux de données

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

### Travaux pratiques

Créer et récupérer des variables, utiliser le résultat du flux et enrichir son flux de données.

## 6 Les boucles

- Les problématiques
- Les boucles avec paramètres et le composant "Copie lignes vers résultat"
- Les boucles avec les composants "Copie lignes vers résultat" et "Récupération lignes depuis résultat"

### Travaux pratiques

Comprendre et manipuler les composants des boucles.

## 7 Exploitation

- Gérer les erreurs
- Générer des traces (logs)
- Comprendre les erreurs et lancer des alertes (débugage)
- La parallélisation (exécution simultanée de plusieurs flux de données)
- Import/export des développements
- Automatisation des tâches/transformation
- Documentation (mise en place des normes en cas d'erreurs ou de reprises)

### Travaux pratiques

Exécuter des flux simultanément, gérer les erreurs et automatiser des tâches.

## Dates et lieux

### CLASSE À DISTANCE

2026 : 8 juin, 14 sep., 30 nov.

### PARIS LA DÉFENSE

2026 : 1 juin, 7 sep., 23 nov.