

# Course : Apache Hop, orchestrating data flows

Manage your ETL processes visually

*Practical course - 3d - 21h00 - Ref. HOA*

*Price : 2010 € E.T.*

Data is vital to business. Apache Hop is a powerful, customizable open source software package for managing ETL (Extract, Transform, Load) processes. It enables you to automate the collection, transformation and organization of data from heterogeneous sources, then send it to a specific source. Save time by mastering pipelines and workflows through an accessible interface.

## Teaching objectives

At the end of the training, the participant will be able to:

- ✓ Understand and explain the Hop environment and how it works
- ✓ Know how to extract and transform data
- ✓ Distinguish and order different data sources
- ✓ Combine, classify and compare different types of data
- ✓ Apply task automations and analyze errors to respond to them

## Intended audience

Anyone feeding and manipulating data flows in a BI database.

## Prerequisites

Good knowledge of SQL.

## Practical details

### Exercise

Application of theory to concrete cases, group discussions, hands-on practice.

### Teaching methods

Active teaching.

## Course schedule

### PARTICIPANTS

Anyone feeding and manipulating data flows in a BI database.

### PREREQUISITES

Good knowledge of SQL.

### TRAINER QUALIFICATIONS

The experts leading the training are specialists in the covered subjects. They have been approved by our instructional teams for both their professional knowledge and their teaching ability, for each course they teach. They have at least five to ten years of experience in their field and hold (or have held) decision-making positions in companies.

### ASSESSMENT TERMS

The trainer evaluates each participant's academic progress throughout the training using multiple choice, scenarios, hands-on work and more. Participants also complete a placement test before and after the course to measure the skills they've developed.

### TEACHING AIDS AND TECHNICAL RESOURCES

- The main teaching aids and instructional methods used in the training are audiovisual aids, documentation and course material, hands-on application exercises and corrected exercises for practical training courses, case studies and coverage of real cases for training seminars.
- At the end of each course or seminar, ORSYS provides participants with a course evaluation questionnaire that is analysed by our instructional teams.
- A check-in sheet for each half-day of attendance is provided at the end of the training, along with a course completion certificate if the trainee attended the entire session.

## 1 Introducing Apache Hop

- Why Hop?
- History and overview
- Tool installation and configuration

### Hands-on work

Install and configure Apache Hop.

## 2 Generate an initial data extraction

- Workflows and pipelines (scheduling)
- From data extraction to data feeding
- Understanding and managing data flows
- Pipeline and workflow execution

### Hands-on work

Design a pipeline and create a workflow.

## 3 Access source and target data

- The concept of metadata
- Configure access to data sources
- Supported sources/targets
- Links between sources (joins)
- Insert/Update power supply

### Hands-on work

Configure access to data sources by identifying supported targets/sources and joins.

## 4 Handling data

- Sort your flow up or down
- Doubling your flow
- Filter data according to several criteria (to lighten the flow)
- Extract information from a field (character string)
- Replace one data item with another
- Operator management/flow calculation operations
- Using the Cartesian product
- Linking information from heterogeneous data
- Compare data streams

### Hands-on work

Handle flows by sorting, splitting and filtering data. Compare feeds.

## 5 Enrich your data flow

- Log generation
- Create and retrieve variables (dates, numeric, alphanumeric)
- Using the result of a flow
- Properties of a flow and its scheduler

### Hands-on work

Create and retrieve variables, use flow results and enrich your data flow.

### TERMS AND DEADLINES

Registration must be completed 24 hours before the start of the training.

### ACCESSIBILITY FOR PEOPLE WITH DISABILITIES

Do you need special accessibility accommodations? Contact Mrs. Fosse, Disability Manager, at psh-accueil@orsys.fr to review your request and its feasibility.

## 6 Loops

- The issues
- Loops with parameters and the "Copy lines to result" component
- Loops with the components "Copy rows to result" and "Retrieve rows from result".

### Hands-on work

Understand and manipulate loop components.

## 7 Operating

- Managing errors
- Generate logs
- Understand errors and launch alerts (debugging)
- Parallelization (simultaneous execution of multiple data streams)
- Import/export developments
- Task automation/transformation
- Documentation (implementation of standards in the event of errors or rework)

### Hands-on work

Run streams simultaneously, manage errors and automate tasks.

## Dates and locations

### REMOTE CLASS

2026 : 16 Mar., 8 June, 14 Sep., 30 Nov.

### PARIS LA DÉFENSE

2026 : 1 June, 7 Sep., 23 Nov.