+33 (0)1 48 10 09 46
formation@ambient-it.net
140 Avenue Jean Lolive 93500 Pantin - France

Updated on 18/07/2025

Sign up

Pentaho ETL Fundamentals Training

3 days (21 hours)

Overview

Master Pentaho Data Integration (PDI) in its entirety thanks to this comprehensive, structured and resolutely practical training course. From simple flow modeling to advanced processing automation, you'll learn how to build, orchestrate and optimize your ETL processes to meet data integration needs in a demanding business context.

You will start by discovering the Spoon environment, the notions of transformation, job and step, as well as connections to data sources (files, SQL databases, APIs). Objective
Acquire the basics needed to manipulate your first flows and make them reliable.

You will then learn about processing sequences, flow control, variable passing and scheduling, to create robust, dynamic workflows adapted to your information system.

A module is dedicated to best practices: project structuring, performance optimization, error management and execution auditing. You'll learn how to document, version and secure your developments.

As with all our training courses, this one will be presented with the latest Pentaho ETL Fundamentals updates.

Objectives

- Understand the architecture of Pentaho Data Integration, its role in the BI chain, and its typical use cases in data integration, migration and automation.
- Know how to design, structure and maintain robust Pentaho transformations and jobs, by efficient handling of steps, source connections, conditions, groupings and joins
- master flow control, parameterization, error handling, logging and scheduling mechanisms to build reliable, modular and maintainable ETL pipelines

- Be able to integrate and orchestrate complex processes, while guaranteeing performance, data quality, auditability and traceability of integration processes.
- Apply best practices in structuring, naming, documentation, versioning and security to become rapidly operational in a professional environment.

Target audience

- Data analyst
- Data Base Administrator

Prerequisites

- Basic knowledge of data handling
- Mastering SQL fundamentals

Pentaho ETL Fundamentals training program

Introduction to Pentaho Data Integration

- Global architecture: BI Server, PDI, BA Server
- PDI's place in the BI chain
- Navigation in Spoon
- Creating and opening projects
- Structure of an ETL project in Pentaho
- Transformations vs. Jobs
- Steps and hops
- Variables and parameters

Data manipulation in transformations

- Flat files: CSV, Excel, JSON, XML
- Databases: MySQL, PostgreSQL, SQL Server, Oracle, etc.
- Other sources: Web Services, APIs, MongoDB, etc.
- Table input / output
- Row filter, Calculator, String operations
- Join Rows, Merge, Sort, Group by
- Step error handling
- Dedicated error streams
- Use of "Abort", "Dummy", "Null if" to control flows

Job construction

- Differences between jobs and transformations
- Job-specific steps: Start, Success, Transformation, Job, etc.
- Flow control: if, switch, while, success/failure
- Conditional calling of jobs or transformations
- Event management
- Use of environment variables
- Passing parameters between jobs/transformations
- Scheduling via cron, script, PDI card or external tools

ETL design best practices

- Naming conventions for jobs, transformations and steps
- Project organization: folders, business logic, techniques
- Adding notes and metadata
- Managing project documentation in Spoon
- Use of streaming transformations
- Limiting costly joins
- Use of buffers and partitioning

Advanced querying and data processing

- Advanced SQL queries in Table Input
- Lookup steps
- Data pivoting/unpivoting
- Modified Java Script Value step
- Use of complex conditions
- Calling custom functions
- Format conversion
- Data cleansing: replace, trim, null

Integration with enterprise environment

- Export .kjb jobs and .ktr transformations
- Deployment with PDI card (Pan card) or command line
- Logging and monitoring
- Logging configuration (Step, Job, Transformation)
- Analysis of execution results
- Audit performance and data quality
- Execution from shell / batch scripts
- Integration with Jenkins, Talend or Apache NiFi
- Pentaho Server: scheduled execution and centralized monitoring

Security, versioning and collaboration

- Access to files and databases
- Password encryption in PDI
- Secure use of Pentaho Repository
- Collaborative work on jobs and transformations

- Versioning with Git
- Best practices in version management

Case study and practical workshop

- Building a complete ETL workflow :
- CSV file management
- Database integration
- Cleansing and transformation
- Report generation
- Step-by-step validation
- Use of preview, debug, breakpoint
- Runtime analysis
- Application of performance best practices
- Adding logs and audits

Companies concerned

This course is aimed at both individuals and companies, large or small, wishing to train their teams in a new advanced IT technology, or to acquire specific business knowledge or modern methods.

Positioning on entry to training

Positioning at the start of training complies with Qualiopi quality criteria. As soon as registration is finalized, the learner receives a self-assessment questionnaire which enables us to assess his or her estimated level of proficiency in different types of technology, as well as his or her expectations and personal objectives for the forthcoming course, within the limits imposed by the selected format. This questionnaire also enables us to anticipate any connection or security difficulties within the company (intra-company or virtual classroom) which could be problematic for the follow-up and smooth running of the training session.

Teaching methods

Practical training: 60% hands-on, 40% theory. Training material distributed in digital format to all participants.

Organization

The course alternates theoretical input from the trainer, supported by examples, with brainstorming sessions and group work.

Validation

At the end of the session, a multiple-choice questionnaire verifies the correct acquisition of skills.

Certification

A certificate will be awarded to each trainee who has completed the entire course.