

Talend Open Studio: Implementing Data Integration

Hands-on course of 3 days - 21h

Ref.: TOT - Price 2024: €2 230 (excl. taxes)

Talend Open Studio (TOS) is an open-source ETL application that makes it possible to perform massive data synchronizations from one database to another. This course will teach you how to design, model, and develop TOS jobs in order to handle different challenges.

EDUCATIONAL OBJECTIVES

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

Design and develop jobs in the application ETL Talend

Optimize the jobs developed by using contexts and datasets.

Carrying out more complex transformations by using variables, expressions, and joins

Execute and debug a job, track execution statistics

TEACHING METHODS

Theoretical contributions illustrated with example implementations, then consolidated by creating mini-Talend projects.

HANDS-ON WORK

Series of mini-projects used to design increasing elaborate Talend jobs.

THE PROGRAMME

last updated: 08/2021

1) Presentation

- Data integration ETL solutions.
- TOS: Installation, user preferences. Reference documentation.
- Product philosophy. Designing a job.

Hands-on work : Getting started with the tool. Overview of the specifications of the generic project.

2) Model the need, design a first job

- Business Modeler. Job Designer.
- Main connections. CSV and XML components.
- Simple transformation components.
- Viewing the generated code, executing a job.

Exercise : Developing a job that sorts a CSV source, filtering data, and storing the result in an XML file.

3) Optimize job development

- Configuring reusable connections using metadata.
- Updating the metadata and propagating it to jobs, importing/exporting metadata.
- Configuring jobs with contexts.
- Externalizing context variables in ".properties" and ".ini" files.
- Creating and managing your own variables.
- Generating data sets for tests.

Exercise : Refactoring a job using metadata and contexts. Generating a test data set for this job.

PARTICIPANTS

Developers, project managers, decision-making IT consultants, database administrators.

PREREQUISITES

Good knowledge of RDBMs and SQL. Knowledge of the Java language is a plus.

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.

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.

4) Working with databases

- Supported databases and main components.
- Configuring operations on tables.
- Metadata and context of connecting to a database schema.
- Sharing connections and managing transactions.
- Creating queries using SQLBuilder.

Exercise : Reading and updating a data repository hosted on a MySQL server.

5) Working with multisource data

- Overview of the tMap component.
- Configuring input flows, creating joins.
- Carrying out more transformations using variables, expressions and joins.
- Qualifying data using filters.
- Generating multiple outputs, managing rejects.

Exercise : Consolidating multisource data and generating a warehouse.

6) Add-ons

- Breaking a job down into sub-jobs, using tRunJob. Launching jobs in the command line. Periodic execution.
- Debugging a job, tracking execution statistics
- tJasperOutput reporting.

Exercise : Generating a Jasper Report from a warehouse.

DATES

REMOTE CLASS

2024 : 10 Jul, 18 Nov