

# Course : Talend Open Studio for Data Quality, managing data quality

*Practical course - 2d - 14h00 - Ref. TDQ*

*Price : 1680 € E.T.*

Open Studio for Data Quality, an open source data profiling tool. You'll learn how to use Talend's tool effectively to assess the level of data quality in your information system. You'll implement analyses, measure data compliance with internal or external standards and define strategies for cleaning up erroneous data. This training is exclusively for Talend Open Studio (free, open source, community version). It does not cover the commercial version of Talend Studio, under paid license, integrated into the Qlik-Talend Cloud portal.

## Teaching objectives

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

- ✓ Connect to data sources, produce statistics, identify data to be profiled
- ✓ Select the different types of indicators and analyses best suited to the data to be monitored
- ✓ Implement complex analyses to verify business rules
- ✓ Define strategies for correcting data errors via Talend Data Integration jobs

## Intended audience

Business analysts, data integrators, data managers.

## Prerequisites

Good knowledge of relational databases and SQL. Basic knowledge of Talend Open Studio for Data Integration.

## Practical details

### Teaching methods

70% of the time is devoted to practicing with the tool. Each participant has his or her own workstation.

## Course schedule

### PARTICIPANTS

Business analysts, data integrators, data managers.

### PREREQUISITES

Good knowledge of relational databases and SQL. Basic knowledge of Talend Open Studio for Data Integration.

### 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.

## 1 The problem of data quality

- Assessing the quality of data in an information system.
- Fundamental criteria: data completeness, accuracy and integrity.
- Positioning Talend Open Studio for Data Quality within the Talend suite.

### Hands-on work

Product installation, configuration of preferences.

## 2 Fundamental concepts of TOS for Data Quality

- Metadata: database connections, delimited files and Excel files.
- Overview of the different types of analysis.
- Analysis tools and indicators.
- Data explorer.

### Hands-on work

Perform an initial column analysis on data from a csv file, and evaluate the results.

## 3 Simple analyses

- Duplicate search, interval constraints, date format, email format...
- Table metrics, functional dependencies between columns.
- Identify redundant values.
- Consistency checks between foreign and primary keys.
- Use indicators, templates, rules and source files.

### Hands-on work

Perform an analysis of each type on a set of partially erroneous data.

## 4 Advanced analysis

- Analyze schema and table structure using the Data Explorer.
- Multi-table and multi-column analysis, compliance with business rules.
- Search for and visualize correlations between columns.
- Create your own indicators and source files.
- Manage analyses.

### Hands-on work

Create a complex business rule involving several tables and associate it with a task. Publish the rule in the Talend forge.

## 5 Advanced elements

- Use context variables.
- Create templates based on regular expressions.
- Export/import analyses and analyzed data.
- Correcting data errors with Talend Data Integration.

### Hands-on work

Set up metadata and analyses using context variables, export analyzed data for correction in Talend Data Integration.

### 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](mailto:psh-accueil@orsys.fr) to review your request and its feasibility.

## Times

Courses take place from 09:00 to 12:30 and from 14:00 to 17:30.

Participants may arrive beginning at 08:45. Breaks and lunches are complimentary.

For four- and five-day hands-on courses, sessions end at 16:00 on the last day of the course, regardless of the teaching mode..