

Updated on 20/06/2024

Sign up

DataFold training

3 days (21 hours)

Presentation

Automate your data testing with our DataFold training course. DataFold is the reference tool for [data testing](#) in data engineering.

Our course will teach you how to install and configure the tool, and you'll take it in hand for the first time to run your first tests (validation test, schema test and anomaly detection).

We'll teach you how to write advanced data tests using customized parameters and assertions, and how to validate your pipelines to ensure the quality of your assets.

This course includes an essential QA chapter for analyzing your changes during testing. The program also includes a [CI/CD](#) section for applying continuous integration within your data pipelines.

What's more, you'll learn best practices for managing your data projects by combining collaboration, documentation and versioning on DataFold.

Objectives

- Master the fundamentals of data testing
- Implement advanced data testing techniques
- Set up data validation pipelines
- Integrate DataFold into CI/CD workflows
- Automate data testing and deployment

Target audience

- Project managers
- Data analysts
- Data scientists
- Data engineers
- Business analysts

Prerequisites

Experience in data management.

OUR DATAFOLD TRAINING PROGRAM

INTRODUCTION TO DATAFOLD

- Introducing DataFold
- Key features
- Installation and initial configuration
- Overview of the user interface

THE FUNDAMENTALS OF DATA TESTING

- Introduction to data testing
- Types of data tests
 - Data validation
 - Schematic testing
 - Anomaly detection
- Creating basic tests
- Test case definition
- Setting up data sources
- Test writing and execution
- Interpretation of results

Advanced data testing techniques

- Complex test scenarios
- Writing advanced test cases
- Using custom assertions
- Parameterized tests

DATA VALIDATION

- Data validation techniques
- Automation of validation tests
- Setting up data validation pipelines
- Automate notifications and reports

- Analysis of validation results
- Practical work: Validating data in a pipeline

QA

- Understanding the impact of modifications
- Using DataFold to analyze the impact of changes
- Forecasting the impact on data pipelines
- TP: QA test

CONTINUOUS INTEGRATION AND DEPLOYMENT (CI/CD)

- CI/CD principles for data pipelines
- Integrating DataFold into CI/CD workflows
- Test and deployment automation
- Deployment monitoring and management
- Practical work: Setting up a CI/CD workflow with DataFold

DATA PIPELINES

- Performance optimization
- Solving bottlenecks
- Performance monitoring
- Practical work: Optimizing a data pipeline

DATA SECURITY

- Data security principles
- Securing your data
- Monitoring access and use

DATA PROJECT MANAGEMENT

- Good collaboration practices
- Data project management with DataFold
- Documentation and communication of results
- Version and change management
- Practical work: Collaborating on a data project

Companies concerned

This training course is aimed at both individuals and companies, large or small, wishing to train their teams in a new advanced computer 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 training to come, 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 course: 60% Practical, 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.

Sanction

A certificate will be issued to each trainee who completes the course.