Updated 12/13/2023

Sign up

Data Version Control training: advanced data management

2 days (14 hours)

Presentation

Our Data Version Control training course will enable you to master DVC for optimum management of your data. You'll be able to easily manage versions of your data sets, track changes over time and revert to earlier versions if necessary.

Our DVC training will enable you to share and synchronize your datasets, for efficient project collaboration. You'll be able to efficiently manage large datasets by optimizing storage, using sampling features and saving space.

In this training course, you'll discover how to integrate DVC with other commonly used data science tools, such as TensorFlow, PyTorch and MLflow, for seamless integration into your workflows. You'll also benefit from practical advice, best practices and tips for structuring your project, optimizing performance and solving common problems.

Objectives

- Understand the fundamental concepts of DVC and its key functionalities
- Mastering data set versioning with DVC
- Learn to collaborate effectively with other team members thanks to DVC

Target audience

- Data scientists
- Machine learning engineers
- Data analysts

Prerequisites

- Basic programming skills
- Understanding the basic concepts of version control systems

Our DVC training program

Introduction to DVC

- What is DVC?
- Version control for source code vs. version control for data
- Overview of DVC's main functions

Basic configuration

- Installing DVC and its dependencies
- Initiating a project
- Data storage location configuration

Data set management

- Adding a data set to DVC
- Tracking changes to a data set
- Data set versioning

Collaboration and sharing

- Using remote repositories for collaboration
- Sharing data sets
- Retrieving shared data sets

Experimentation and reproducibility

- Creating an experiment configuration
- Recording results and experimental metrics
- Replication of previous experience

Large data sets

- Managing large data sets

- Strategies for efficient data storage
- Sampling functions

Integration of external tools

- Integration of machine learning frameworks (TensorFlow, PyTorch)
- Experience management platforms (MLflow)

Advanced functions

- Using DVC with data processing pipelines
- Parallel industries and experience
- Tags to mark specific versions of data sets

Security and authorization management

- Data set access control
- Secret management
- Audit and track changes
- Integration with existing security tools

Deployment and production roll-out

- Data sets in production
- Production deployment (Kubernetes, Docker)

Case studies

- Machine learning project
- Data science research project

Additional resources

- Official documentation and online resources
- Community and forums

Companies concerned

This 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 enabling 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 with regard to 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.