AMBIENT-IT

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

Updated 03/05/2024

Sign up

Dataframe thriller training

3 days (21 hours)

Presentation

Our Polars dataframe training course will help you get the most out of Polars, the open source library for data processing. While data pipelines often run into problems when processing Big Data with Python, Polars paves the way for processing large datasets thanks to its high-performance structures.

Our program will enable you to install Polars on your system and configure it according to your precise needs. You'll learn how to work with CSV and Parquet files, as well as data manipulation techniques in Polars (filtering, selection, sorting, etc.).

Our training also covers the use of Lazy mode and interoperability with conversion to and from Pandas & Numpy. You'll also learn how to read CVS and JSON files.

Like all our training courses, it will run on the latest version of the tool: Polars 0.20.23

Objectives

- Handling large data sets with Polars
- Analyze data sets with expressions
- Using Lazy mode

Target audience

- **Data scientists**
- Data analysts

Prerequisites

- Mastery of Python

OUR DATAFRAME DETECTIVE STORY TRAINING PROGRAM

INTRODUCTION TO POLARS

- What is Polars?
- Differences with Pandas
- Using Polars for data analysis
- Overview of the performance and expressive syntax of Polars
- DataFrames and Series structure in Polars
- Introducing Lazy Evaluation mode

ENVIRONMENT INSTALLATION AND CONFIGURATION

- Installing Polars on different operating systems
- Development environment configuration
- Checking installation and solving common problems
- Preparing the workspace for practical exercises

BASIC DATA STRUCTURES IN POLARS

- Detailed explanation of the Series
- Understanding DataFrames and their benefits
- Basic creation and handling of Series and DataFrames
- Conversion between Polars and Pandas/Numpy for interoperability

DATA MANIPULATION WITH POLARS

- Data import and export with CSV and Parquet files
- Data selection, filtering and sorting techniques
- Adding and transforming columns in a DataFrame
- Handling missing values and data types

LAZY MODE

- Introduction to Lazy mode and its benefits
- Query construction
- Optimizing performance with lazy evaluation
- Practical examples of use

STATISTICS AND AGGREGATIONS

- Using basic statistical functions
- Techniques for counting and summarizing values
- Group By and aggregations, including complex expressions
- Lazy aggregation and related optimizations

OPERATIONS ON GROUPS AND JOINS

- Pivot, melt and other group transformations
- Concatenation and joins (inner, left, fast-track) of DataFrames
- Advanced filtering using conditions and joins

WORKING WITH TEMPORAL DATA

- Introduction to time data types and time zone management
- Temporal data handling and analysis
- Group by dynamics for time series analysis

READING AND WRITING DATA FILES

- Read and write various file formats (CSV, Excel, JSON, Parquet)
- Techniques for processing larger-than-memory data sets
- Reading data from databases

CONCLUSION AND BEST PRACTICES

- Summary of key training concepts
- Discussion of the reasons for choosing Polars over other tools
- Exchange on advanced use cases and integration with other libraries
- Tips for continuing to learn and improve after training

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 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 internal security difficulties within the company (intra-company or virtual classroom) that might be encountered.

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.