Updated on 18/10/2024

Sign up

LanceDB training

2 days (14 hours)

Presentation

LanceDB is an AI database for vector data management and retrieval, ideal for artificial intelligence, machine learning and semantic search applications. Optimized for embeddings, it enables fast, accurate searches on complex data, including images, text and other types of content.

In this program, you'll learn how to install and configure LanceDB, manipulate the user interface and master basic commands. We'll explore the advantages of LanceDB over other databases, and you'll see how to use it in various scenarios such as vector search and data processing.

We'll cover fundamental concepts such as vector and scalar indexing, embedding management for semantic search, and integration with tools like Pandas and DuckDB. You'll also learn how to create tables and build indexes optimized for enhanced performance, including hybrid search.

By the end of this course, you'll be able to migrate data to LanceDB, optimize large-scale information retrieval and solve performance problems. You'll master the integration of machine learning models for smarter searches, equipping you with practical expertise in vector and semantic search.

This training will be presented with the latest technology resources.

Objectives

- Install and configure LanceDB efficiently
- Mastering the user interface and basic commands
- Understanding and applying vector and scalar indexing

- Integrate custom templates and use embeddings
- Optimize performance and manage data migration

Target audience

- Developers
- Data Scientists
- Engineers

Prerequisites

- A good understanding of relational and non-relational databases
- Familiarity with the concept of embeddings and vector search
- Experience in integrating machine learning models and optimizing databases

OUR LanceDB TRAINING PROGRAM

INTRODUCTION TO LanceDB

- Introducing LanceDB
- Installation and initial configuration
- Overview of user interface and basic commands
- LanceDB advantages over other databases
- Typical use scenarios and case studies

FUNDAMENTAL CONCEPTS OF LanceDB

- Introduction to vector research and its importance
- Understanding indexing: methods and best practices
- Storage options and performance implications
- Data management: insertion, update, deletion
- Exploring index types: vector, scalar and hybrid

PRACTICAL WITH TABLES AND INDEXES

- Creating and manipulating tables in LanceDB
- Building and optimizing a vector index
- Advanced vector search and full-text search
- Setting up a scalar index and corresponding search
- Hybrid search for optimized results

EMBEDDING AND INTEGRATION MANAGEMENT

- Understanding embeddings and their key role in semantic searches
- Using embedding functions and integrating custom templates
- Practical examples: multilingual semantic search and CLIP multimodal embeddings
- LanceDB integration with Pandas, PyArrow, Polars, and DuckDB
- Demonstration of advanced integrations such as LangChain and LlamaIndex

MIGRATION AND PERFORMANCE OPTIMIZATION

- Migration guide: strategies for transferring data to LanceDB
- Identification of common migration challenges and solutions
- Advanced techniques for tuning data recovery performance
- Adjust storage parameters to improve performance
- Case studies and live problem-solving

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 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.