

Updated 07/27/2023

Sign up

Delta Lake training

2 days (14 hours)

Presentation

Guarantee the reliability and quality of your data lake. Delta Lake is an open storage layer used to enhance the performance of your data lake. Delta Lake is a technology that brings together all your data types (semi-structured, structured and unstructured) in one place. By offering a single source for your data, Delta Lake guarantees its veracity. The tool supports schema validation and ACID transactions, enabling it to offer the reliability lacking in traditional data lakes. Your data will be secure thanks to the Delta Sharing protocol and native Unity Catalog integration. So you can share your data with your colleagues in complete confidence. This Delta Lake training course will teach you the benefits of the program, table manipulation, Delta Time Travel and streaming. An additional one-day module is available if you wish to integrate Delta Lake with Amazon Web Service. As always, we'll teach you the latest version of the tool: [Delta Lake 1.0](#).

Objectives

- Handling tables with Delta Lake
- Using Delta Time Travel
- Use Delta Lake for streaming

Target audience

- Business analysts
- Data analysts
- Data scientists
- Data engineers
- Data miners

Prerequisites

- Knowledge of SQL and database management
- Knowledge of Big Data

Delta Lake Training Program

Introduction to Delta Lake

- The role of data lakes in Big Data architecture
- The veracity of data lakes
- Delta architecture
- How can Delta Lake improve data reliability?
- Why use Delta Lake with Databricks?
- Installation

Getting started with the tool

- Initialize Spark
- Load data
- Overview of table formats
- Creating your first table
- The transaction log

Handling tables

- View and configure table history
- Converting parquet tables to delta tables
- Managing and optimizing competition
- Removing and updating

Delta Time Travel

- Modify your data correctly
- What is time travel?
- Delta directories
- Using time travel
- Vacuum and data retention

Delta streaming

- The data engineer pipeline with Delta Lake
- Best practices for a robust process
- Best practices before integrating Spark Streaming
- Streaming with Delta
- Using Delta as a stream source

Add-on module: Delta Lake with AWS (1 additional day)

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.

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.