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

Updated 03/31/2025

Sign up

Databricks training

2 days (14 hours)

Presentation

Databricks simplifies your Big Data process. Created by the founders Apache Spark, this platform makes the ETL process more secure.

Databricks uses Apache Spark architecture and adds powerful, reliable pipelines. The tool provides automated management of clusters and IPython-style notebooks.

Data pipelines can be written in a wide range of languages (Scala, SQL, R, Python) and designed in collaborative workspaces.

Databricks is also rigorously secure, thanks to a unified protection model covering all functionalities (identity management, encryption, etc.).

Our Databricks training course will show you how to set up a complete ETL process. We'll start with an overview of the system, then move on to data extraction, loading and transformation, and finally to dashboards and the use of IntelliJ IDE.

As always, we'll be presenting the latest version of the tool, Databricks 15.4.

Objectives

- Get to know Databricks well
- Extracting data with Databricks
- How to transform and load data
- Use dashboards and deploy your process

Target audience

- Developers
- Data Engineer
- Architects
- System administrators
- Data miners
- Data scientists
- Data analysts
- Business intelligence analysts
- Market inteligence analysts

Prerequisites

- Ideally, you should have taken our Spark ML or Spark Tuning Advanced training courses.
- Knowledge of Scala, SQL and ideally Python
- Have a Databricks account

Databricks training program

Introduction

- Tool presentation
- Why use Databricks?
- Databricks vs Apache Spark
- Interface presentation
- Notebooks
- Create a cluster and a table
- Create jobs
- Create a pool

Extracting data

- Import data
- Add schematics
- Managing tables with SQL
- Using Python on Databricks

Transforming your data

- Transforming data with Scala
- Data manipulation with Spark SQL
- Modifying data with Python
- Using the DataFrame API

Loading data

- The nested XML file
- The nested json file
- DELTA tables

Dashboard and process deployment

- Dashboard presentation
- Developing a job to refresh the dashboard
- Creating a project with IntelliJ IDE
- Create your application
- Adding dependencies
- Outsource properties
- Send jobs

Updating Databricks content

- Interface
- New features on the platform
- Best optimization practices

Spark Streaming

- Concepts
- Window functions
- Watermarking
- Real-time aggregations

Delta Live Tables

- Automation and management of data flows

Workflow orchestration

- End-to-end pipeline development

Unity Catalog

- Managing data governance and security in a Lakehouse environment

Companies concerned

This course is aimed at both individuals and companies, large or small, wishing to train its teams in a new advanced IT 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, 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.