Course : Databricks for data engineers

Practical course - 3d - 21h00 - Ref. DKI
Price : 2030 € E.T.

NEW

Over three days, this course equips data engineers with the skills required to fully leverage Azure Databricks within a modern cloud environment. From creating and managing notebooks to orchestrating workflows, and working with Python, SQL, and Datastore, it provides a comprehensive, hands-on approach. The program emphasizes performance optimization, automation, and data value creation to support scalable, production-ready data initiatives.

Teaching objectives

At the end of the training, the participant will be able to:

- Understand and use the various Databricks functions
- Creating and manipulating notebooks
- Setting up jobs
- Using Python and SQL with Databricks
- Understanding and using the Databricks Datastore
- Creating dashboards with SQL Analytics
- Connecting Databricks and Power BI

Intended audience
Data engineers, employees involved in setting up, automating and industrializing data pipelines on modern cloud platforms.

Prerequisites
Good knowledge of SQL and data manipulation. Basic knowledge of Python.

Practical details
Hands-on work
Discussions, exercises and practical work.
Teaching methods
Active

Course schedule

1 Introduction

- Azure: cloud principles, presentation and benefits

2 Databricks: overview

- Introduction to Azure Databricks: concepts (data lakehouse, ETL, data warehouse)
- Data governance: Unity Catalog concepts
- Advantages of Databricks

3 Workspaces

- Azure Databricks resources
- Creating a workspace
- Connecting the workspace to external data sources

Hands-on work
Create a workspace and connect it to external data sources.

4 The different types of resources: clusters and SQL warehouse

- Presentation
- The different types of calculation
- Creating a cluster
- Cluster management

Hands-on work
Identify the different types of calculations, create and manage a cluster.

5 Unity Catalog

- Catalog creation
- Connecting to data from Unity Catalog
- Adding a table to the catalog schema
- Defining authorizations on a catalog table
- Focus: connecting to SQL databases
- Other ways of adding data to the catalog

Hands-on work
Create a catalog and connect data to it. Add a table, define authorizations and master SQL database connection.

6 Databricks: notebooks

- Notebook objectives and benefits
- The cell principle
- Controls
- Magic commands
- PySpark programming, SQL…
- Using cells and visualizing results
- Exporting and sharing notebooks

Hands-on work
Connect Databricks to storage tools (ADLS, SQL database, API), query, transform and store data in Databricks. Understand the different possible formats (Parquet, Delta, external table, etc.).

7 Spark and Delta principle

- Introduction to Spark SQL
- Delta lake concepts
- Table lifecycle management: merge delta, delete/insert

8 Process orchestration

- Calling up notebooks from Azure data factory and configuration
- Using Databricks workflows
- Workflow-based parameterization and configuration of Databricks notebooks
- Introduction to Databricks Delta Live Tables

Hands-on work

Handle Databricks workflows, set up and configure Databricks notebooks from within a workflow.

9 Databricks SQL

- Databricks SQL: administration
- SQL Editor
- Creating and executing a query
- Schedule a request
- Create alerts
- Sharing query results
- Options available on query results: export, display, display filters, result parameterization

Hands-on work

Create a query and share the results.

Dates and locations

PARIS LA DÉFENSE
2026 : 30 Mar., 29 June, 28 Sep., 16 Dec.