Course : Databricks for data engineers

Practical course - 3d - 21h00 - Ref. DKI
Price : 2380 CHF E.T.

NEW

This two-day course gives data engineers the keys to exploiting the full potential of Azure Databricks in a modern cloud context. From notebook creation to job orchestration, via the use of SQL, the Datastore and integration with Power BI, the focus is on performance, automation and data enhancement for industrialized data projects.

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 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.

Practical details

Hands-on work

Discussions, exercises and practical work.

Teaching methods

Active

Course schedule

PARTICIPANTS

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

PREREQUISITES

Good knowledge of SQL and data manipulation.

TRAINER QUALIFICATIONS

The experts leading the training are specialists in the covered subjects. They have been approved by our instructional teams for both their professional knowledge and their teaching ability, for each course they teach. They have at least five to ten years of experience in their field and hold (or have held) decision-making positions in companies.

ASSESSMENT TERMS

The trainer evaluates each participant's academic progress throughout the training using multiple choice, scenarios, hands-on work and more.
Participants also complete a placement test before and after the course to measure the skills they've developed.

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.).

TEACHING AIDS AND TECHNICAL RESOURCES

• The main teaching aids and instructional methods used in the training are audiovisual aids, documentation and course material, hands-on application exercises and corrected exercises for practical training courses, case studies and coverage of real cases for training seminars.
• At the end of each course or seminar, ORSYS provides participants with a course evaluation questionnaire that is analysed by our instructional teams.
• A check-in sheet for each half-day of attendance is provided at the end of the training, along with a course completion certificate if the trainee attended the entire session.

TERMS AND DEADLINES

Registration must be completed 24 hours before the start of the training.

ACCESSIBILITY FOR PEOPLE WITH DISABILITIES

Do you need special accessibility accommodations? Contact Mrs. Fosse, Disability Manager, at psh-accueil@orsys.fr to review your request and its feasibility.

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.